Problems on Large Sparse Graphs

Payam Delgosha

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2020-166
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-166.html

August 14, 2020

Problems on Large Sparse Graphs

by

Payam Delgosha

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering — Electrical Engineering and Computer Sciences

and the Designated Emphasis

in

Communication, Computation and Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Venkat Anantharam, Chair
Professor David Aldous
Professor Jean Walrand

Summer 2020

Problems on Large Sparse Graphs

Abstract

Problems on Large Sparse Graphs

by

Payam Delgosha

Doctor of Philosophy in Engineering — Electrical Engineering and Computer Sciences

and the Designated Emphasis in

Communication, Computation and Statistics

University of California, Berkeley

Professor Venkat Anantharam, Chair

In this thesis, we study two category of problems involving large sparse graphs, namely the problem of compression for graphical data, and load balancing in networks. We achieve this by employing the framework of local weak convergence, or so called the objective method. This framework provides a viewpoint which enables one to make sense of a notion of stationary stochastic processes for sparse graphs.

By employing the local weak convergence framework, we introduce a notion of entropy for probability distributions on rooted graphs. This is a generalization of the notion of entropy introduced by Bordenave and Caputo to graphs which carry marks on their vertices and edges. Such marks can represent information on real-world data. This notion of entropy can be considered as a natural counterpart for the Shannon entropy rate in the world of sparse graphical data. We illustrate this by introducing a universal compression scheme for sparse marked graphs. Furthermore, we study distributed compression of graphical data. In particular, we introduce a version of the Slepian–Wolf theorem for sparse marked graphs.

In addition to studying the problem of compression, we study the problem of load balancing in networks. We do this by modeling the problem as a hypergraph where each hyperedge represents a task carrying one unit of load, and each vertex represents a server. An allocation is a way of distributing this load. we study balanced allocations, which are roughly speaking those allocations in which no demand desires to change its allocation. Employing an extension of the local weak convergence theory to hypergraphs, we study certain asymptotic behaviors of balanced allocations, such as the asymptotic empirical load distribution at a typical server, as well as the asymptotic of the maximum load.

Problems studied in this thesis should be considered as examples showing the wide-range applicability of the local weak convergence theory and the above mentioned notion of entropy. In fact, this framework provides a viewpoint of stationary stochastic processes for sparse marked graphs. The theory of time series is the engine driving an enormous range of applications in areas such as control theory, communications, information theory and signal processing. It is to be expected that a theory of stationary stochastic processes for combinatorial structures, in particular graphs, would eventually have a similarly wide-ranging impact.

To my father, the memory of my mother, Farshid, Behnaz, and Negar.

Contents

List of Figures

Acknowledgments

First and foremost, I would like to express my profound gratitude to my advisor, Prof. Venkat Anantharam, for all his support during my PhD at Berkeley. Venkat has always provided me with the freedom to explore various directions to find the ones that interest me the most, while guiding me in every project to ensure that I make the most out of it. I have always been impressed when Venkat spent many hours reading my manuscripts to provide me with very detailed and precious feedback. Venkat is very knowledgeable, patient, and caring. I have learnt a lot from him, specially from his calmness and patience when we faced obstacles in research.

I am grateful to Tom Courtade for the opportunity to work with him on some research problems, specially during my first year. I appreciate the time he spent, and I really enjoyed his excitement when discussing research problems. I would also like to thank Tom for being a part of my qualifying exam committee. I am also grateful to Kannan Ramchandran for the time he spent to discuss research topics as well as for providing his valuable feedback on my work on several occasions. I would also like to thank Kannan for giving me the opportunity to be a GSI for EE126. I am grateful to David Aldous and Jean Walrand for being a part of my dissertation committee.

I have been very fortunate to be supported financially by the Center for Science of Information (CSoI), a US NSF Science and Technology Center under grant CCF-0939370.

I want to thank the EECS administrative staff, especially Shirley Salanio and Kim Kail, for being extremely reliable and helpful in resolving all my administrative issues.

I would like to thank the great members of WIFO/BLISS who have been very good friends of mine. Lab has been like home to me for the past few years, thanks to my labmates Ramtin Pedarsani, Ashwin Pananjady, Vidya Muthukumar, Orhan Ocal, Soham Phade, Varun Jog, Vasuki Narasimha Swamy, Nihar Shah, Rashmi Vinayak, Kangwook Lee, Giulia Fanti, Ilan Shomorony, Vijay Kamble, Gireeja Ranade, Fanny Yang, Dong Yin, Raaz Dwivedi, Yuting Wei, Naveen Goela, Avishek Ghosh, and Vipul Gupta. I would also like to thank my wonderful friends who have brought joy to my life in the past few years, Reza Abbasi Asl, Maryam Shadmehr, Ehsan Asgari, Nima Anari, Dorsa Sadigh, Mohammad Keshavarzi, Yasaman Yavari, Ali Yadollahi, Ameneh Gholipour, Maryam Moeini, Farzad Pourbabaee, Heydar Davoudi, and Zeinab Vessal.

I would also like to thank my dear friends back in Iran, who are like my family, and have always encouraged me to pursue my passion and interests. Special thanks go to Lida Samadpour, Massoud Roshani, Arash Roshani, Alma Roshani, Saman Abdollahi, and Sassan Harradji. I am also grateful to Prof. Amin Gohari who has guided me through my graduate studies. Amin was a very supportive and kind advisor during my masters studies in Iran, and I have learned a lot from him.

I devote my very special thanks to my family. My dear father Mohammad and the memory of my dear mother Fatemeh for their love, support, and sacrifice. My brother Farshid who has been my role model and always encouraged me to pursue my interests, and his wife Behnaz. I would also like to thank my wife's family, who are like my own family.

My final and special thanks go to my wife Negar Mehr. Negar has brought meaning and happiness to my life. She is extremely kind, and has always supported me and motivated me to pursue my passion. I am thankful to her for all the memorable moments we have spent together. I never imagined that the most difficult times in life can become easy and even joyful with a loved one.

Chapter 1

Introduction

One of the implications of the modern technology is that we deal with a type of information that is best represented in forms of graphs in our daily lives. Whenever we search for a topic on the internet, use social media, or use navigation applications to find the shortest path to our destination, we are taking advantage of algorithms and methods which are executed on graphs. A graph is an abstract combinatorial data structure which is capable of modeling *interactions* between objects, and is used to represent a great variety of modern data. In fact, the class of graphical data is much richer compared to the traditionally studied time series or multidimensional time series. Such *graphical data* arise, for instance, in social networks, molecular and systems biology, web graphs, road networks, and in several other applications. For example, graphical data representing a social network would be a snapshot view of the network at a given time. In this example, the graph may describe whether a pair of individuals has ever had an interaction. Moreover, marks on the vertices represent some characteristics of the individuals currently of interest for the data analysis task, e.g. their preference for coffee versus tea. Furthermore, the marks on the edges represent the characteristics of their interaction, e.g. whether they are friends or not.

Usually, the real–world graphical data are huge in size. Take the graph of the internet as an example, where each vertex represents a web page, and each edge represents a link between two web pages. The resulting graph has several billion vertices. The sheer size of such graphical data makes it challenging to analyze and store them. This argues for the necessity of finding efficient and optimal methods and algorithms for analyzing and storing graphical data. PageRank is an example of such efficient algorithms which is widely used in search engines to address search queries [PBMW99].

In practice, normally the graphical data has some form of sparsity property. Roughly speaking, sparsity means that in a graph with n vertices, the number of edges scales much slower than $\binom{n}{2}$. For instance, in a social network graph, where a vertex represents an individual, and each edge represents a friendship interaction, one person is usually connected to only a subset of other people rather than being connected to a majority of the whole population.

One promising approach to study problems on graphs is through the lens of *random* *graphs*. The field of random graphs has been extensively studied, and there are a variety of random graph models which can be used to model different types of graphical data for a diverse class of applications [Bol98, VDH16]. However, when the size of the graph is very large, it is advantageous to approximate it by an "infinite" object to simplify certain analysis, in particular asymptotic analysis. This viewpoint is not limited to graphs, and is also utilized in the definition of *stochastic processes* for times series. More precisely, a stochastic process in its classical form is an infinite sequence of random variables which is identified by its finite dimensional marginals. Moreover, given a time series, we may compute its empirical distribution with a given window size, say k, and compare it with the finite marginal distribution of the stochastic process with the same dimension k. If they are close, we may think of the time series as "being typical" or "being consistent" with the stochastic process. A conceptually similar approach can be employed for sparse graphs in order to define a notion of stochastic process for them. The framework of *local weak convergence* or so called the *objective method* provides such a viewpoint [BS01, AS04, AL07]. The sparsity regime of interest for the local weak convergence framework is, roughly speaking, when the number of edges in the graph proportional to the number of vertices in the graph.[1] We use this framework as a counterpart of stochastic processes in the world of sparse graphical data to study some problems on sparse graphs. We will discuss these problems in the following.

We first give a very rough explanation of the main ideas behind the local weak convergence framework here, without being mathematically rigorous. To simplify the discussion, here we focus on unmarked graphs. However, the same ideas are applicable to marked graphs, as we will discuss the details in Chapter 2. Given a finite graph G on the vertex set V, we define the empirical distribution of G as follows. We pick a vertex v uniformly at random from V, and look at the structure of the graph from the point of view of this vertex v. This yields a random rooted graph, which we denote by $U(G)$. See Figure 1.1 for an example. Note that in this process, we look at the structure of the graph by removing the vertex labels. Here, the term *label* refers to the integer index $1, 2, \dots$ associated to a vertex, and is distinct from a vertex mark. Therefore, the resulting object $U(G)$ will be a probability distribution on the space of *unlabeled* rooted graphs. In other words, $U(G)$ keeps the frequency of different *local patterns* existing in the graph G. When the graph G is large, we may be able to *approximate* $U(G)$ with a probability distribution on rooted graphs with *possibly infinite depth*. We denote this set of unlabeled rooted graphs by $\mathcal{G}_*$. Motivated by this, we may treat a probability distribution μ on $\mathcal{G}_*$ as a *graph stochastic process*. With this, we can think of a graph G as being typical if $U(G)$ is *close* to μ in a certain sense.

This framework naturally yields a notion of *convergence* for sparse graphs. In other words, we may say that a sequence of finite graphs G_n converges in the local weak sense to a probability distribution μ on $\mathcal{G}_*$ if $U(G_n)$ *converges* to μ in a certain sense. As an example, assume that G_n is a realization of the sparse Erdős–Rényi random graph model $\mathcal{G}(n, \alpha/n)$, where $\alpha > 0$ is fixed, and each edge in G_n is independently present with probability α/n. If n is large, the degree of a vertex v in the graph chosen uniformly at random has approximately a

[1]this regime is sometimes referred to as the *very sparse* regime in the literature [BCCZ19]

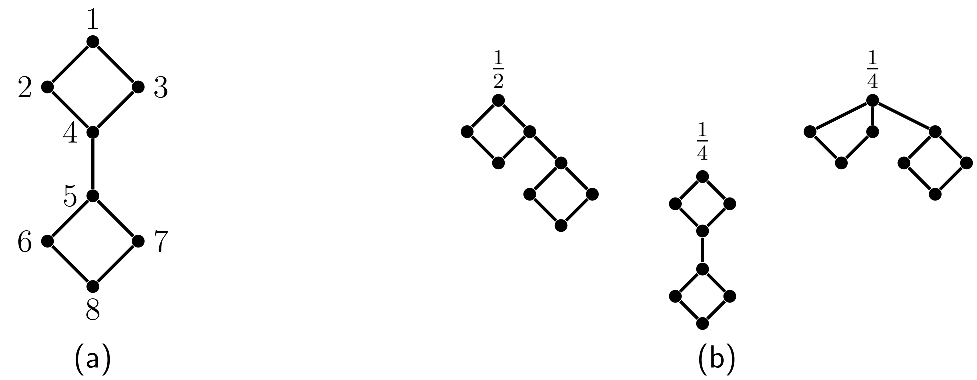

Figure 1.1: With G being the graph in (a), (b) depicts $U(G)$, which is a probability distribution the space of unlabeled rooted graphs. For the sake of simplicity, we have assumed that G is unmarked in this figure. Note that if we pick a vertex v uniformly at random in G, with probably $4/8 = 1/2$, v belongs to the set $\{2, 3, 6, 7\}$. Furthermore, the structure of the graph rooted at any of these four vertices is the same, and is the left–most object illustrated in (b). Note that the term *label* refers to the integer index $1, 2, \dots, 8$ associated to a vertex, and is distinct from a vertex mark.

Poisson distribution with mean α. Moreover, the same argument holds for any of the vertices adjacent to v. Also, it can be shown that the probability of having cycles in a neighborhood of v is vanishing. Therefore, $U(G)$ is *approximately* a Poisson Galton-Watson tree with mean degree α. A Poisson Galton-Watson tree is a branching process where the root has a Poisson number of children, each child has a Poisson number of children and so on. Therefore, we say that the above sequence of sparse Erdős–Rényi graphs converges in the local weak sense to a Poisson Galton-Watson tree. Due to the existence of a notion of convergence, this framework has been useful in studying certain asymptotic problems on large spare graphs. For instance, Aldous employed this framework to study the asymptotic behavior of the random assignment problem [Ald01]. Lyons used this framework for asymptotic enumeration of spanning trees in large graphs [Lyo05]. Other applications include, but not limited to, spectral graph theory [BL10], planar triangulation [Ang03, AS03], and combinatorial optimization [AS04, Ste02, GNS06, Gam04]. The concept of looking at a discrete process from the point of view of an individual already exists in other frameworks, such as the Palm theory, which employs a similar concept for point processes [BB12]. Baccelli et al. have recently introduced a framework on finitely bounded discrete metric spaces which simultaneously generalizes the local weak convergence theory and the Palm theory [BHMK20].

1.1 Thesis Approach

In this thesis, we study two classes of problems on large sparse graphs. Namely, compression and load balancing. We employ the framework of local weak convergence as a counterpart of the notion of stochastic processes for sparse graphical data. In particular, we develop a notion of entropy for this framework, and we show that it is indeed the information theoretical limit of compression for sparse graphs. Below, we explain the structure of this thesis in more details.

1.1.1 Part I: Local weak convergence and the marked BC entropy

First, in Chapter 2, we review in detail the local weak convergence framework that we highlighted above. Then, in Chapter 3, we introduce a notion of entropy for this framework, which we call the *marked BC entropy*. This is a generalization of the notion of entropy introduced by Bordenave and Caputo in [BC15] to the regime where vertices and edges in the graph can carry *marks* on top of the connectivity structure of the graph.

Roughly speaking, the entropy associated to a probability distribution μ on the space $\mathcal{G}_*$, which we denote by $\Sigma(\mu)$, is defined through studying the asymptotic behavior of the size of the set of *typical* graphs. Bordenave and Caputo observed in [BC15] that, roughly speaking, if $\mathcal{G}_{n,m_n}(\mu, \epsilon)$ denotes the set of graphs G on the vertex set $\{1, \dots, n\}$ having m_n edges such that $U(G)$ is ϵ–close to μ (with respect to a metric which we will discuss in detail), then we have

$$\log |\mathcal{G}_{n,m_n}(\mu, \epsilon)| \approx m_n \log n + n\Sigma(\mu).$$

In other words, the log of the size of the set of typical graphs has a leading term which is $m_n \log n$. Furthermore, the entropy, which is denoted by $\Sigma(\mu)$ here, is the coefficient of the second order term which scales linearly in the number of vertices in the graph. As we will see in Chapter 3, in the marked regime where vertices and edges in the graph can have marks coming from a finite set, a similar scaling exists which leads to our definition of the marked BC entropy. Recall that we assume that the graphs are sparse, and m_n scales linearly with n. In fact, $m_n \approx \frac{d}{2}n$ where d is the expected degree at the root in μ. Therefore, the leading term $m_n \log n$ scales as $\frac{d}{2}n \log n$. In fact, putting aside the leading term which is of order $n \log n$, this notion of entropy captures the *per vertex* growth rate of the size of the typical graphs.

As we will see in Part II, this notion of entropy is indeed the information theoretic limit of compression for sparse graphs. However, it is important to mention that this notion of entropy, together with the local weak convergence as a counterpart of stochastic processes for sparse graphical data, can have a wide range of applications in addressing many problems for sparse graphs, not necessarily limited to compression.

1.1.2 Part II: Compression of Graphical Data

As we discussed above, in practical application, usually the size of the graph underlying the data is large. Designing efficient compression schemes to store and analyze the data is therefore of significant importance. In Chapter 4, we introduce a universal lossless compression for sparse marked graphical data. Our notion of universality is similar in nature to that of time series. For time series, a universal compression scheme is capable of achieving the optimal compression rate without knowing the stochastic model from which the time series is generated. We have a similar notion of universality for graphical data, where the stochastic process is now replaced with the notion of local weak convergence, and the optimal compression rate is governed by the marked BC entropy. The results in Chapter 4 were previously published in [DA17b] and [DA20].

For time series, the problem of universal compression has been extensively studied [CT12], and efficient algorithms such as Lempel-Ziv [ZL77, ZL78, Wel84] are proposed which are capable of achieving the optimal information theoretic limit of compression. For sparse marked graphs, we propose a lossless compression scheme in Chapter 4. Moreover, this scheme is universal in the sense that for a sequence of marked graphs G_n converging to a limit object μ in the local weak sense, the normalized codeword length associated to G_n does not asymptotically exceed the marked BC entropy of μ. To address universality, we assume that the encoder does not know the limit object μ a priori.

The literature on compression and evaluating the information content of graphical data can be divided into two categories based on whether there is a stochastic model for the graphical data. Works that do not consider a stochastic model usually aim to compress specific types of graphical data, such as web graphs [BBH+98, SMHM99, BKM+00, BV04], social networks [CKL+09, MP10, BRSV11, Mas12], or biological networks [DWvW12, ADK12, KK14, SSA+16, HPP16]. These works often take advantage of some properties specific to a data source, where such properties are usually inferred through observing real-world data samples. For example, Boldi and Vigna proposed the WebGraph framework to encode web graphs, where each node represents a URL, and two nodes are connected if there is link between them [BV04]. Later, Boldi et al. proposed a method called layered label propagation as a compression scheme for social networks [BRSV11].

Among models making stochastic assumptions, Choi and Szpankowski studied structural compression of the Erdős–Rényi ensemble $\mathcal{G}(n, p)$ [CS12]. There has been a recent series of works addressing the universal compression of binary trees, see for instance [KYS09], [ZYK13], [MTS18], [GHLB19]. Aldous and Ross studied models of sparse random graphs with vertex labels [AR14]. They considered several models on sparse random graphs, and studied the asymptotic behavior of the entropy of such models. They observed that the leading term in these models scales as $n \log n$, where n is the number of vertices in the graph. Abbe studied the asymptotic behavior of the entropy of stochastic block models, and discussed the optimal compression rate for such models up to the first order term [Abb16].

The key property distinguishing our approach is universality, as we discussed above. There has been some attempt to address universality for compressing graphs, but they are usually restricted to special contexts. For instance, Zhang et al. have addressed universal compression of binary trees [ZYK13], and Basu and Varshney have addressed the problem of source coding for deep neural networks, assuming that the network structure is known, but the weights come from distributions with unknown parameters [BV17]. In contrast to such approaches, by employing the local weak convergence framework, we introduce a general and *non-parametric* approach. In addition to universality, the notion of entropy that we employ captures the per vertex growth rate of typical graphs after appropriately separating out the leading term. However, the existing literature usually consider the random graph ensemble entropy up to only the leading term. Finally, we consider marked graphs rather than simple graphs, where the marks, as we discussed above, can model certain types of information on top of the connectivity structure of the graph.

We further go beyond source coding for a single graphical source by studying distributed compression of graphical data in Chapter 5. As the data is not always available in one location, it is also important to consider distributed compression of graphical data. Traditionally, when dealing with time series, distributed lossless compression is modeled using two (or more) possibly dependent jointly stationary and ergodic processes representing the components of the data at the individual locations. In this case, the rate region, which characterizes how efficiently the data can be compressed, is given by the Slepian–Wolf Theorem [CT12]. We adopt an analogous framework, namely that two jointly defined marked random graphs on the same vertex set are presented to two encoders, one to each encoder. Each encoder is then required to individually compress its data such that a third party, having access to the two compressed representations, can recover both marked graph realizations with a vanishing probability of error in the asymptotic limit of the size of the data. We characterize the compression rate region for two scenarios, namely, a sequence of marked sparse Erdős–Rényi ensembles and a sequence of marked configuration model ensembles. Moreover, we generalize this two graphical source result to the case where there are more than two graphical sources. Part of the results in Chapter 5 was previously published in [DA18a].

1.1.3 Part III: Load Balancing

The problem of load balancing is ubiquitous in networks. As examples, consider the problem of routing traffic through a communication network or of assigning tasks among the servers in a cloud computing framework. What is common in these scenarios is a number of servers and a number of tasks whose load should be distributed among the servers. Examples of servers are paths through the network from a given source to a given destination in the routing scenario, or processors in a cloud computing framework. Examples of tasks are the amount of traffic to be routed from the source to the destination or the computational work to be done at the servers, respectively. There are typically restrictions as to which resources are available to a given task. Performance considerations require that the allocation of the load of a task among the resources available to it should be done in a way that optimizes a measure of performance, such as delay or queue length. When the problem size is large, it may be expensive to compute the detailed characteristics of an optimal or sufficiently good allocation of the load. Instead, it is interesting to focus on the statistical characteristics of the allocation, such as the empirical distribution of the load faced by a typical resource in the network. Chapter 6 is concerned with developing such a viewpoint in the context of a specific kind of load balancing problem which has broad applicability. The results in Chapter 6 were previously published in [DA17a] and [DA18b].

We build upon the notion of load balancing which was studied by Hajek [Haj90] who, in particular, formulated the notion of a *balanced allocation*. It is natural to expect that a task would be happier to use servers that are currently handling less load, if available, as opposed to those handling more load. Hajek modeled the load balancing problem as a graph, where each vertex represents a server, and each edge represents a task which has a unit load, and this unit load can be arbitrarily divided between the two servers corresponding to the two endpoints of that edge. An allocation is therefore a way of distributing the load on each edge among the two servers corresponding to that edge. Hence, every server (or vertex) in the graph receives a total amount of load, which is the aggregation of individual loads coming from the tasks (or the edges) connected to that vertex. Roughly speaking, a balanced allocation is defined to be an allocation in which no demand desires to change its allocation. Hajek then conjectured that if the underlying graph is randomly generated according to a sparse Erdős–Rényi ensemble, the total load corresponding to a balanced allocation at a vertex in the graph chosen uniformly at random converges to a limit distribution, as the number of vertices in the graph goes to infinity. Moreover, he conjectured that this limit distribution could be identified through a certain distributional fixed point equation.

Anantharam and Salez settled this conjecture using the framework of local weak convergence [AS16]. In fact, as we discussed above, the local weak limit of the sparse Erdős–Rényi random graphs is a Poisson Galton-Watson tree, and the recursive nature of this limit allows one to identify the limiting load distribution at the root via a distributional fixed point equation. In Chapter 6, we extend this analysis to the setting where each task can have accesses to more than two servers. We model this using a hypergraph, where each hyperedge represents a task, and the endpoints of this hyperedge represent the servers to which the task has accesses to, i.e. the servers among which the unit load of the task can be distributed. We study the problem of the convergence of the distribution of the total load in a balanced allocation by employing the local weak convergence framework. To achieve this, we develop a counterpart of the local weak convergence framework for hypergraphs in Chapter 6. We believe that this generalized framework could be of independent interest in a variety of problems in which the underlying model is best expressed in terms of hypergraphs rather than graphs.

Part I

Local Weak Convergence and the Marked BC Entropy

Chapter 2

The Framework of Local Weak Convergence

In this chapter, we introduce the local weak convergence framework. The reader is referred to [BS01, AS04, AL07] for background and further details. We begin with introducing some notation in Section 2.1 which will be used throughout this thesis. In Section 2.2, we review the notion of weak convergence in probability spaces. In Section 2.3, we set up our notation for marked graphs, which is going to be used in this chapter as well as in Chapters 3 through 5. In Section 2.4, we define the local topology on the space of rooted marked graphs, use it to define convergence in the local weak sense, and give the definition of unimodularity. In Section 2.5 we present some examples to illustrate the concept of local weak convergence. Section 2.6 extends local weak convergence to multigraphs. In Section 2.7, we introduce marked unimodular Galton–Waton trees, which form an important class of unimodular probability distributions on the space of rooted marked trees. Finally, we conclude the chapter in Section 2.8.

2.1 Notation

$\mathbb{N}$ denotes the set of natural numbers, $\mathbb{Z}_+$ the set of nonnegative integers, $\mathbb{Z}$ the set of integers, and $\mathbb{R}$ the set of real numbers. For $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, \dots, n\}$. All logarithms in this document are to the natural base unless otherwise stated. We therefore use nats instead of bits as the unit of information. For two sequences $(a_n, n \geq 1)$ and $(b_n, n \geq 1)$ of positive real numbers, we write $a_n = O(b_n)$ if $\sup_n a_n/b_n < \infty$, and we write $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$. We write $\{0,1\}^* - \emptyset$ for the set of sequences of zeros and ones of finite length, excluding the empty sequence. For $x \in \{0,1\}^* - \emptyset$, we denote the length of the sequence x by $\mathsf{nats}(x)$, which is obtained by multiplying the length of x in bits by $\log 2$. Equality by definition is denoted by $:=$ and $=:$. For a positive integer N and a sequence of nonnegative integers $\{a_i\}_{1\leq i\leq k}$ such that $\sum_{i=1}^k a_i \leq N$, we define

$$\binom{N}{\{a_i\}_{1\leq i\leq k}} := \frac{N!}{a_1!\dots a_k!(N-a_1-\dots-a_k)!}.$$

We denote by $\mathbb{1}[A]$ the indicator of the event A. For a probability distribution P , $X \sim P$ denotes that the random variable X has law P. A finite sequence of nonnegative integers $(d(1), \dots, d(n))$ is said to be *graphic* if there is a simple graph on n vertices with vertex i having degree $d(i)$ for $1 \leq i \leq n$. A simple characterization of graphic sequences is provided by the well known theorem of Erdös and Gallai [Cho86, EG60]. For a probability distribution $Q = (q_1, \dots, q_n)$ defined on a finite set, $H(Q)$ denotes the Shannon entropy of Q, which is defined as

$$H(Q) := \sum_{i=1}^n -q_i \log q_i.$$

Throughout this thesis, we identify $0 \log 0$ with 0. Also, for a random variable X taking values in a finite set, we denote by $H(X)$ its Shannon entropy. Other notation used in this document is introduced at its first appearance.

2.2 The Topology of Weak Convergence

A Polish space is a complete and separable metric space. For a Polish space X, let $\mathcal{P}(X)$ and $\mathcal{M}(X)$ respectively denote the set of probability measures and nonnegative finite measures on X, with respect to the Borel σ–field of X. When referring to a Polish space, we always employ its Borel σ–field. We use the abbreviations "a.s." and "a.e." for the phrases "almost surely" and "almost everywhere", respectively. For a Polish space X, we say that a sequence of probability measures μ_n converges weakly to a probability measure $\mu \in \mathcal{P}(X)$, and write $\mu_n \Rightarrow \mu$, if for any bounded continuous function $f : X \to \mathbb{R}$, we have

$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu.$$

See [Bil71] and [Bil13] for more details on weak convergence of probability measures. The following result, called the *portmanteau* theorem, gives useful conditions equivalent to weak convergence.

Theorem 2.1 (Theorem 2.1 in [Bil13])**.** *Given a Polish space X, a sequence of probability measures μ_n on X, and $\mu \in \mathcal{P}(X)$, the following conditions are equivalent:*

1. $\mu_n \Rightarrow \mu$.
2. $\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$ *for all bounded, uniformly continuous f.*
3. $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for all closed F.*

4. *$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ for all open G.*

5. *$\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all Borel set A whose boundary has measure zero under μ (such a set is called a μ–continuity set).*

Now, we define a metric on $\mathcal{P}(X)$ for a Polish space X. For a Borel subset A of X and $\epsilon > 0$, the ϵ–extension of A, which we denote by A^ϵ, is defined to be the union of ϵ–balls around the elements in A. Given two probability measures μ and ν in $\mathcal{P}(X)$, the Lévy–Prokhorov distance between μ and ν, which is denoted by $d_{\text{LP}}(\mu, \nu)$, is defined to be the infimum over $\epsilon > 0$ such that for all Borel set A, we have

$$\mu(A) \leq \nu(A^\epsilon) + \epsilon, \qquad \text{and} \qquad \nu(A) \leq \mu(A^\epsilon) + \epsilon.$$

It can be shown that d_{LP} introduces a metric on $\mathcal{P}(X)$ which is equivalent to the topology of weak convergence that we discussed above [Bil13].

Given two measurable spaces $(X_1, \mathcal{F}_1)$ and $(X_2, \mathcal{F}_2)$, a measurable mapping $f : X_1 \to X_2$, and a nonnegative measure μ_1 on $\mathcal{F}_1$, the pushforward measure $f_*(\mu_1)$ on $\mathcal{F}_2$ is defined by

$$f_*(\mu_1)(A) = \mu_1(f^{-1}(A)),$$

for $A \in \mathcal{F}_2$.

2.3 Marked Graphs

All graphs in this document are defined on a finite or countably infinite vertex set, and are assumed to be locally finite, i.e. the degree of each vertex is finite. Given a graph G, we denote its vertex set by $V(G)$. A simple graph is a graph without self-loops or multiple edges between pairs of vertices. A simple marked graph is a simple graph where each edge carries two marks coming from a finite edge mark set, one towards each of its endpoints, and each vertex carries a mark from a finite vertex mark set. We denote the edge and vertex mark sets by Ξ and Θ respectively. For an edge between vertices $v, w \in V(G)$, we denote its mark towards the vertex v by $\xi_G(w, v)$, and its mark towards the vertex w by $\xi_G(v, w)$. Also, $\tau_G(v)$ denotes the mark of a vertex $v \in V(G)$. Let $\mathcal{G}_n$ denote the set of graphs and $\bar{\mathcal{G}}_n$ the set of marked graphs on the vertex set $[n]$. See Figure 2.1 for an example. A *marked tree* is a marked graph T where the underlying graph is a tree.

All graphs and marked graphs appearing in this document are also assumed to be simple, unless otherwise stated. Therefore we will use the terms "graph" and "marked graph" as synonymous with "simple locally finite graph" and "simple locally finite marked graph" respectively. Further, since a graph can be considered to be a marked graph with the edge and vertex mark sets being of cardinality 1, all definitions that are made for marked graphs will be considered to have been simultaneously made for graphs.

Let G be a finite marked graph. We define the *edge mark count vector* of G by $\vec{m}_G := (m_G(x, x') : x, x' \in \Xi)$ where $m_G(x, x')$ is the number of edges (v, w) in G where $\xi_G(v, w) = x$

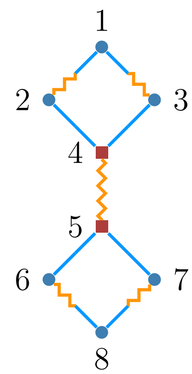

Figure 2.1: A marked graph G on the vertex set $\{1,\dots,8\}$ where edges carry marks from $\Xi = \{$Blue (solid), Orange (wavy)$\}$ (e.g. $\xi_G(1,2) =$ Orange while $\xi_G(2,1) =$ Blue; also, $\xi_G(2,4) = \xi_G(4,2) =$ Blue) and vertices carry marks from $\Theta = \{\bullet, \blacksquare\}$ (e.g. $\tau_G(3) = \bullet$).

and $\xi_G(w,v) = x'$, or $\xi_G(v,w) = x'$ and $\xi_G(w,v) = x$. Likewise, we define the *vertex mark count vector* of G by $\vec{u}_G := (u_G(\theta) : \theta \in \Theta)$ where $u_G(\theta)$ is the number of vertices $v \in V(G)$ with $\tau_G(v) = \theta$.

For a marked graph G and vertices $v, w \in V(G)$, we write $v \sim_G w$ to denote that v and w are adjacent in G. Moreover, for a vertex $o \in V(G)$, $\deg_G^{x,x'}(o)$ denotes the number of vertices v connected to o in G such that $\xi_G(v,o) = x$ and $\xi_G(o,v) = x'$, and $\deg_G(o)$ denotes the degree of o, i.e. the total number of vertices connected to o in G, which is precisely $\sum_{x,x' \in \Xi} \deg_G^{x,x'}(o)$.

A *path* between two vertices v and w in the marked graph G, is a sequence of distinct vertices $v_0, v_1, \dots, v_k$, such that $v_0 = v$, $v_k = w$ and, for all $1 \le i \le k$, we have $v_{i-1} \sim_G v_i$. The length of such a path is defined to be k. Additionally, for vertices $v, w \in V(G)$, $\mathrm{dist}_G(v,w)$ denotes the distance between v and w, which is the length of the shortest path connecting v to w. If there is no such path, the distance is defined to be ∞.

A marked forest is a marked graph with no cycles. A marked tree is a connected marked forest.

2.4 Local Weak Convergence and Unimodularity

Given a connected marked graph G on a finite or countably infinite vertex set and a vertex $o \in V(G)$, we call the pair (G, o) a *rooted connected marked graph.* We extend this notation to a marked graph G that is not necessarily connected and a vertex $o \in V(G)$ by defining (G, o) to be $(G(o), o)$, where $G(o)$ denotes the connected component of o in G. In general, we call (G, o) a *rooted marked graph.*

Definition 2.1. *Let G and G' be marked graphs. Let $o \in V(G)$ and $o' \in V(G')$. We say that (G,o) and (G',o') are isomorphic, and write $(G,o) \equiv (G',o')$, if there exists a bijection between the sets of vertices of $G(o)$ and $G'(o')$ which maps o to o' while preserving vertex marks, the adjacency structure of these connected components, and the edge marks.*

Isomorphism defines an equivalence relation on rooted connected marked graphs. The isomorphism class of a rooted marked graph (G,o) is denoted by $[G,o]$, and is determined by $(G(o),o)$. The set comprised of the isomorphism classes $[G,o]$ of all rooted marked graphs on any finite or countably infinite vertex set, where the edge and vertex marks come from the sets Ξ and Θ respectively, is denoted by $\bar{\mathcal{G}}_*(\Xi,\Theta)$. When the mark sets are clear from the context, we use $\bar{\mathcal{G}}_*$ as a shorthand for $\bar{\mathcal{G}}_*(\Xi,\Theta)$. Likewise, let $\bar{\mathcal{T}}_*(\Xi,\Theta)$ denote the subset of $\bar{\mathcal{G}}_*(\Xi,\Theta)$ consisting of all isomorphism classes $[T,o]$ where (T,o) is a rooted marked forest. As for general graphs, the isomorphism class of (T,o) is determined by the marked tree $(T(o),o)$, where $T(o)$ is the connected component of the vertex $o \in T$. When the mark sets are clear from the context, we use $\bar{\mathcal{T}}_*$ as a shorthand for $\bar{\mathcal{T}}_*(\Xi,\Theta)$.

For an integer $h \geq 0$, we denote by $(G,o)_h$ the h–neighborhood of the vertex $o \in V(G)$, rooted at o. This is defined by considering the subgraph of G consisting of all the vertices $v \in V(G)$ such that $\mathrm{dist}_G(o,v) \leq h$ and then making this subgraph rooted at o. The isomorphism class of the h–neighborhood $(G,o)_h$ is denoted by $[G,o]_h$. It is straightforward to check that $[G,o]_h$ is determined by $[G,o]$.

For $[G,o],[G',o'] \in \bar{\mathcal{G}}_*$, we define $d_*([G,o],[G',o'])$ to be $1/(1+h_*)$, where h_* is the maximum over integers $h \geq 0$ such that $(G,o)_h \equiv (G',o')_h$. If $(G,o)_h \equiv (G',o')_h$ for all $h \geq 0$, it is easy to see that $(G,o) \equiv (G',o')$, i.e. $[G,o] = [G',o']$. In this case, $d_*([G,o],[G',o'])$ is defined to be zero. It can be easily checked that $\bar{\mathcal{G}}_*$, equipped with d_*, is a metric space. In particular, it satisfies the triangle equality. In fact, it can be shown, for any finite sets Ξ and Θ, that $\bar{\mathcal{G}}_*(\Xi,\Theta)$ and $\bar{\mathcal{T}}_*(\Xi,\Theta)$ are complete and separable metric spaces, i.e. Polish spaces [AL07].[1]

For an integer $h \geq 0$, let $\bar{\mathcal{G}}_*^h \subset \bar{\mathcal{G}}_*$ consist of isomorphism classes of rooted marked graphs where all the vertices of the connected component of the root are at distance at most h from the root. For instance, for $[G,o] \in \bar{\mathcal{G}}_*$, we have $[G,o]_h \in \bar{\mathcal{G}}_*^h$. We define $\bar{\mathcal{T}}_*^h \subset \bar{\mathcal{T}}_*$ similarly. Note that, by definition, we have $\bar{\mathcal{G}}_*^0 \subset \bar{\mathcal{G}}_*^1 \subset \cdots \subset \bar{\mathcal{G}}_*$. Consequently, for $[G,o] \in \bar{\mathcal{G}}_*^h$ and $0 \leq k \leq h$, we have $[G,o]_k \in \bar{\mathcal{G}}_*^k$.

For a marked graph G on a finite vertex set, we define $U(G) \in \mathcal{P}(\bar{\mathcal{G}}_*)$ as

$$U(G) := \frac{1}{|V(G)|} \sum_{v \in V(G)} \delta_{[G,v]}, \tag{2.1}$$

where $[G,v]$ denotes the isomorphism class of the connected component of v in G rooted at v. In words, $U(G)$ is the neighborhood structure of the graph G from the point of view of a

[1] In fact, a more general statement without requiring that Ξ and Θ be finite sets holds, but we refer the reader to [AL07] for more details about this, as we do not need that more general statement here.

vertex chosen uniformly at random. Moreover, for $h \geq 0$, let

$$U(G)_h := \frac{1}{|V(G)|} \sum_{v \in V(G)} \delta_{[G,v]_h}, \tag{2.2}$$

be the depth h neighborhood structure of a vertex in G chosen uniformly at random. Note that $U(G)_h \in \mathcal{P}(\bar{\mathcal{G}}_*^h)$.

Given a sequence $(G_n : n \in \mathbb{N})$ of marked graphs, if $U(G_n) \Rightarrow \mu$ for some $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$, then we say that the sequence G_n converges *in the local weak sense* to μ, and say that μ is the *local weak limit* of the sequence. A Borel probability measure $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is called *sofic* if it is the local weak limit of a sequence of finite marked graphs. Not all Borel probability measures on $\bar{\mathcal{G}}_*$ are sofic. A necessary condition for a measure to be sofic exists, called *unimodularity* [AL07]. To define this, let $\bar{\mathcal{G}}_{**}$ be the set of isomorphism classes $[G, o, v]$ of marked connected graphs with two distinguished vertices $o, v \in V(G)$ (which are ordered, but need not be distinct). Here, isomorphism is naturally defined as a bijection preserving marks and adjacency structure which maps the two distinguished vertices of one object to the respective ones of the other. A measure $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is called unimodular if, for all measurable non–negative functions $f : \bar{\mathcal{G}}_{**} \to \mathbb{R}_+$, we have

$$\int \sum_{v \in V(G)} f([G, o, v]) d\mu([G, o]) = \int \sum_{v \in V(G)} f([G, v, o]) d\mu([G, o]), \tag{2.3}$$

where in each expression the summation is over $v \in V(G)$ that are in the same connected component of G as o, since otherwise the expression $[G, o, v]$ is not defined. It can be seen that, in order to check unimodularity, it suffices to check the above condition for functions f such that $f([G, o, v]) = 0$ unless v is adjacent to o. This is called *involution invariance* [AL07]. We denote the set of unimodular probability measures on $\bar{\mathcal{G}}_*$ by $\mathcal{P}_u(\bar{\mathcal{G}}_*)$. Similarly, as $\bar{\mathcal{T}}_* \subset \bar{\mathcal{G}}_*$, we can define the set of unimodular probability measures on $\bar{\mathcal{T}}_*$, which we denote by $\mathcal{P}_u(\bar{\mathcal{T}}_*)$.

For $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$, and $\theta \in \Theta$, we denote by $\Pi_\theta(\mu)$ the probability under μ of the root having mark θ, i.e. $\mathbb{P}(\tau_G(o) = \theta)$ where $[G, o]$ has law μ.[2] With this, let $\vec{\Pi}(\mu) := (\Pi_\theta(\mu) : \theta \in \Theta)$ be the *probability vector of the root mark.* Also, for $x, x' \in \Xi$, we define $\deg_{x,x'}(\mu) := \mathbb{E}\left[\deg_G^{x,x'}(o)\right]$ where $[G, o]$ has law μ.[3] In fact, $\deg_{x,x'}(\mu)$ denotes the expected number of edges connected to the root with mark x towards the root and mark x' towards the other endpoint. Moreover, let $\deg(\mu)$ be the expected degree at the root. Note that, by definition, we have $\deg(\mu) = \sum_{x,x' \in \Xi} \deg_{x,x'}(\mu)$. Furthermore, let $\vec{\deg}(\mu) := (\deg_{x,x'}(\mu) : x, x' \in \Xi)$.

All the preceding definitions and concepts have the obvious parallels in the case of unmarked graphs. These can be arrived at by simply walking through the definitions while

[2] Here we observe that $\tau_G(o)$ is the same for all (G, o) in the equivalence class $[G, o]$, so we can unambiguously write $\tau_G(o)$ given only the equivalence class $[G, o]$.

[3] Here we observe that $\deg_G^{x,x'}(o)$ is the same for all (G, o) in the equivalence class $[G, o]$.

restricting the mark sets Θ and Ξ to be of cardinality 1. It is convenient, however, to sometimes use the special notation for the unmarked case that matches the one currently in use in the literature. We will therefore write $\mathcal{G}_*$ for the set of rooted isomorphism classes of unmarked graphs. This is just the set $\bar{\mathcal{G}}_*$ in the case where both Ξ and Θ are sets of cardinality 1. We will also denote the metric on $\mathcal{G}_*$ by d_*, which is just $\bar{d}_*$ when both Ξ and Θ are sets of cardinality 1.

Every $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ that appears in this document will be assumed to satisfy $\deg(\mu) < \infty$. However, for clarity, we will explicitly repeat this condition wherever necessary.

The following lemma gives a useful tool for establishing when local weak convergence holds. This lemma is proved in Appendix A.1.

Lemma 2.1. *Let $\{\mu_n\}_{n\geq 1}$ and μ be Borel probability measures on $\bar{\mathcal{G}}_*$ such that the support of μ is a subset of $\bar{\mathcal{T}}_*$. Then $\mu_n \Rightarrow \mu$ iff the following condition is satisfied: For all $h \geq 0$ and for all rooted marked trees (T, i) with depth at most h, if*

$$A^h_{(T,i)} := \{[G, o] \in \bar{\mathcal{G}}_* : (G, o)_h \equiv (T, i)\}, \tag{2.4}$$

then $\mu_n(A^h_{(T,i)}) \to \mu(A^h_{(T,i)})$.

An important consequence of unimodularity is that, roughly speaking, every vertex has a positive probability to be the root. The following is a rephrasing of Lemma 2.3 in [AL07].

Lemma 2.2 (Everything Shows at the Root)**.** *Let $\mu \in \mathcal{P}_u(\bar{\mathcal{G}}_*)$ be unimodular. If for a subset $\Theta_0 \subset \Theta$ the mark at the root is in Θ_0 almost surely (with $[G, o]$ distributed as μ), then the mark at every vertex is in Θ_0 almost surely. Furthermore, if for a subset $A \subset \Xi \times \Xi$ it holds that for every vertex v adjacent to the root o the pair of edge marks $(\xi_G(v, o), \xi_G(o, v))$ on the edge connecting o to v is in A almost surely (with $[G, o]$ distributed as μ), then for every edge (u, w) the pair of edge marks $(\xi_G(u, w), \xi_G(w, u))$ is in A almost surely.*

2.5 Some Examples

We next present some examples to illustrate the concepts defined so far.

1. Let G_n be the finite lattice $\{-n, \dots n\} \times \{-n, \dots, n\}$ in $\mathbb{Z}^2$. As n goes to infinity, the local weak limit of this sequence is the distribution that gives probability one to the lattice $\mathbb{Z}^2$ rooted at the origin. The reason is that, if we fix a depth $h \geq 0$, then for n large almost all of the vertices in G_n cannot see the borders of the lattice when they look at the graph around them up to depth h, so these vertices cannot locally distinguish the graph on which they live from the infinite lattice $\mathbb{Z}^2$.

2. Suppose G_n is a cycle of length n. The local weak limit of this sequence of graphs gives probability one to an infinite 2–regular tree rooted at one of its vertices. The intuitive explanation for this is essentially identical to that for the preceding example.

3. Let G_n be a realization of the sparse Erdős–Rényi graph $\mathcal{G}(n, \alpha/n)$ where $\alpha > 0$, i.e. G_n has n vertices and each edge is independently present with probability α/n. One can show that if all the G_n are defined on a common probability space then, almost surely, the local weak limit of the sequence is the Poisson Galton–Watson tree with mean α, rooted at the initial vertex. To justify why this should be true without going through the details, note that the degree of a vertex in G_n is the sum of $n-1$ independent Bernoulli random variables, each with parameter α/n. For n large, this approximately has a Poisson distribution with mean α. This argument could be repeated for any of the vertices to which the chosen vertex is connected, which play the role of the offspring of the initial vertex in the limit. The essential point is that the probability of having loops in the neighborhood of a typical vertex up to a depth h is negligible whenever h is fixed and n goes to infinity.

4. Let G_n be a marked bipartite graph on the $2n$ vertices $\{1, \dots, 2n\}$, the edge mark set having cardinality 1 and the vertex mark set being $\Theta = \{R, B\}$. Suppose $\{1, \dots, n\}$ is the set of left vertices, all of them having the mark R, and $\{n+1, \dots, 2n\}$ is the set of right vertices, all of them having the mark B. There are $3n$ edges in the graph, comprised of the edges $(i, n + [\![i]\!])$, $(i, n + [\![i+1]\!])$, and $(i, n + [\![i+2]\!])$ for $1 \le i \le n$, where for an integer k, $[\![k]\!]$ is defined to be n if $k \mod n = 0$, and $k \mod n$ otherwise, so that $1 \le [\![k]\!] \le n$. See Figure 2.2a for an example. The local weak limit of this sequence of graphs gives probability $\frac{1}{2}$ to the equivalence class of each of the two rooted marked infinite graphs described below. The underlying rooted unmarked infinite graph equivalence class for each of these two rooted marked equivalence classes is the same and can be described as follows: There is a single vertex at level 0, which is the root, three vertices at level 1, and four vertices at each of the levels m for $m \ge 2$. For the purpose of describing the limit (there is no such numbering in the limit), one can number the vertex at level zero as 0, the three vertices at level 1 as $(1,1)$, $(1,2)$ and $(1,3)$, and the four vertices at level m, for each $m \ge 2$, as $(m,1)$, $(m,2)$, $(m,3)$ and $(m,4)$ such that the edges are the following: Vertex 0 is connected to each of the vertices $(1,1)$, $(1,2)$ and $(1,3)$. Vertex $(1,1)$ is connected to $(2,1)$ and $(2,2)$, vertex $(1,2)$ is connected to $(2,2)$ and $(2,3)$, and vertex $(1,3)$ is connected to $(2,3)$ and $(2,4)$. The edges between the vertices at level k and those at level $k+1$, for $k \ge 2$, are given by the pattern $((k,1),(k+1,1))$, $((k,1),(k+1,2))$, $((k,2),(k+1,2))$, $((k,3),(k+1,3))$, $((k,4),(k+1,3))$, $((k,4),(k+1,4))$. There are no other edges. As for the distinction between the two rooted marked equivalence classes which each get probability $\frac{1}{2}$ in the limit, this corresponds to the distinction between choosing the mark R for the root and then alternating between marks B and R as one moves from level to level, or choosing the mark B for the root and then alternating between marks R and B as one moves from level to level. See Figure 2.2b for an example.

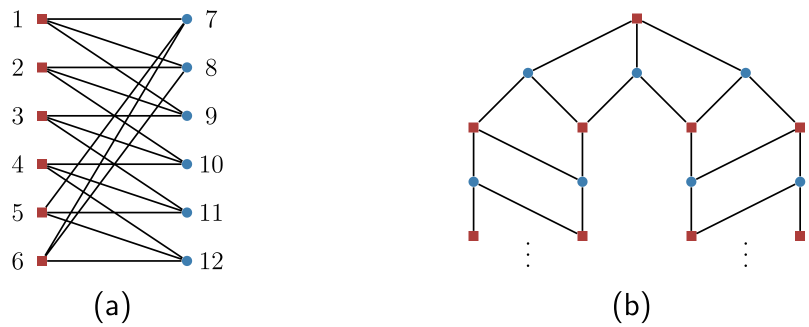

Figure 2.2: The graph in Example 4, (a) illustrates the graph G_6 which has 12 vertices and 18 edges. The vertex mark set is $\Theta = \{B(\bullet), R(\blacksquare)\}$, and the edge mark set Ξ has cardinality 1. The local weak limit of G_n is a random rooted graph which gives probability 1/2 to the rooted marked infinite graph illustrated in (b), and gives probability 1/2 to a similar rooted marked graph which has a structure identical to (b), but the mark of each vertex is switched from R to B and vice versa.

2.6 Local Weak Convergence for Multigraphs

The framework above, which was defined for (locally finite, simple) graphs, can be extended to multigraphs, as defined in [BC15, Section 2]. Here we give a brief introduction, and refer the reader to [BC15], and also to [AS04], [AL07], for further reading.

A multigraph on a finite or countably infinite vertex set V is a pair $G = (V, \omega)$ where $\omega : V^2 \to \mathbb{Z}_+$ is such that, for $u, v \in V$, $\omega(u,u)$ is even and $\omega(u,v) = \omega(v,u)$. We interpret $\omega(u,u)/2$ as the number of self-loops at vertex u, and $\omega(u,v)$ as the number of edges between vertices u and v. The degree of a vertex u is defined to be $\deg(u) := \sum_{v \in V} \omega(u,v)$. The notions of path, distance and connectivity are naturally defined for multigraphs. A multigraph G is called locally finite if $\deg(v) < \infty$ for all $v \in V$.

All multigraphs encountered in this document will be locally finite, so the term "multigraph" will be considered synonymous with "locally finite multigraph". Further, we assume that all multigraphs are unmarked.

It can be checked that a multigraph is a graph (i.e. a locally finite multigraph is a simple locally finite graph) precisely when $\omega(u,v) \in \{0,1\}$ for all pairs of vertices u and v (in particular, $\omega(u,u) = 0$ for all vertices u).

A rooted multigraph (G, o) is a multigraph on a finite or countably infinite vertex set V together with a distinguished vertex $o \in V$.

Definition 2.2. *Two rooted multigraphs $(G_1, o_1) = ((V_1, \omega_1), o_1)$ and $(G_2, o_2) = ((V_2, \omega_2), o_2)$ are said to be isomorphic if there is a bijection σ between the sets of vertices of the respective connected components of the roots which preserves the roots and connectivity. Namely,*

$\sigma(o_1) = o_2$ and we have $\omega_2(\sigma(v), \sigma(u)) = \omega_1(v, u)$ for all u and v in the connected component of o_1. We denote this by writing $(G_1, o_1) \equiv (G_2, o_2)$.

This notion of isomorphism defines an equivalence relation on rooted connected multigraphs, where the equivalence class to which a rooted multigraph belongs is determined by the connected component of the root. Let $\widehat{\mathcal{G}}_*$ be the set of all equivalence classes $[G, o]$ of rooted multigraphs corresponding to this isomorphism relation.

For $h \geq 0$, let $(G, o)_h$ denote the induced multigraph defined by the vertices in G with distance no more than h from o, rooted at o. Let $[G_1, o_1], [G_2, o_2] \in \widehat{\mathcal{G}}_*$ and (G_1, o_1) and (G_2, o_2) be arbitrary members of $[G_1, o_1]$ and $[G_2, o_2]$, respectively. The distance between $[G_1, o_1], [G_2, o_2] \in \widehat{\mathcal{G}}_*$ is defined to be $1/(1 + h_*)$, where h_* is the maximum h such that $(G_1, o_1)_h \equiv (G_2, o_2)_h$. If $(G_1, o_1)_h \equiv (G_2, o_2)_h$ for all $h \geq 0$, then we define the distance to be zero, because this occurs precisely when $(G_1, o_1) \equiv (G_2, o_2)$. It can be checked that this distance defined on $\widehat{\mathcal{G}}_*$ is indeed a metric, and $\widehat{\mathcal{G}}_*$ equipped with this metric is a Polish space [AL07].

2.7 Marked Unimodular Galton–Watson Trees

In this section, we introduce an important class of unimodular probability distributions on $\bar{\mathcal{T}}_*$, called *marked unimodular Galton–Watson trees*. These probability distributions can be thought of as the counterpart of finite memory Markov processes in the local weak convergence language. The construction here is a generalization of the one in Section 1.2 of [BC15]. Before giving the definition, we need to set up some notation.

Given $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$, let $\mu_h \in \mathcal{P}(\bar{\mathcal{G}}_*^h)$ denote the law of $[G, o]_h$, where $[G, o]$ has law μ. We similarly define $\mu_h \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ for $\mu \in \mathcal{P}(\bar{\mathcal{T}}_*)$, recalling that $\bar{\mathcal{T}}_* \subset \bar{\mathcal{G}}_*$. For a marked graph G, on a finite or countably infinite vertex set, and adjacent vertices u and v in G, we define $G(u, v)$ to be the pair $(\xi_G(u, v), (G', v))$ where G' is the connected component of v in the graph obtained from G by removing the edge between u and v. Similarly, for $h \geq 0$, $G(u, v)_h$ is defined as $(\xi_G(u, v), (G', v)_h)$. See Figure 2.3 for an example. Let $G[u, v]$ denote the pair $(\xi_G(u, v), [G', v])$, so $G[u, v] \in \Xi \times \bar{\mathcal{G}}_*$. Likewise, for $h \geq 0$, let $G[u, v]_h$ denote $(\xi_G(u, v), [G', v]_h)$, so $G[u, v]_h \in \Xi \times \bar{\mathcal{G}}_*^h$.

For $g \in \Xi \times \bar{\mathcal{G}}_*$, we call the Ξ component of g its *mark component* and denote it by $g[m]$. Moreover, we call the $\bar{\mathcal{G}}_*$ component of g its *subgraph component* and denote it by $g[s]$. Given a marked graph G and adjacent vertices u and v in G, and for $g \in \Xi \times \bar{\mathcal{G}}_*$, we write $G(u, v) \equiv g$ to denote that $\xi_G(u, v) = g[m]$ and also (G', v) falls in the isomorphism class $g[s]$. We define the expression $G(u, v)_h \equiv g$ for $g \in \Xi \times \bar{\mathcal{G}}_*^h$ in a similar fashion. For $g \in \Xi \times \bar{\mathcal{G}}_*^h$ and an integer $k \geq 0$, we define $g_k \in \Xi \times \bar{\mathcal{G}}_*^{\min\{h,k\}}$ to have the same mark component as g, i.e. $g_k[m] := g[m]$, and subgraph component the truncation of the subgraph component of g up to depth k, i.e. $g_k[s] := (g[s])_k$. For a marked graph G, two adjacent vertices u, v in G, and $h \geq 1$, we define the *depth h type* of the edge (u, v) as

$$\varphi_G^h(u, v) := (G[v, u]_{h-1}, G[u, v]_{h-1}) \in (\Xi \times \bar{\mathcal{G}}_*^{h-1}) \times (\Xi \times \bar{\mathcal{G}}_*^{h-1}). \tag{2.5}$$

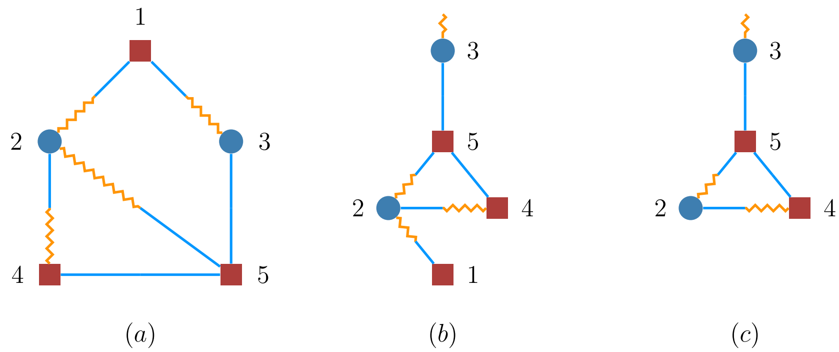

Figure 2.3: (a) A marked graph G on the vertex set $\{1,\dots,5\}$ with vertex mark set $\Theta = \{●, ■\}$ and edge mark set $\Xi = \{$Blue (solid), Orange (wavy)$\}$. In (b), $G(1,3)$ is illustrated where the first component $\xi_G(1,3)$ is depicted as a half edge with the corresponding mark going towards the root 3, and (c) illustrates $G(1,3)_2$. Note that $G(1,3)$ can be interpreted as cutting the edge between 1 and 3 and leaving the half edge connected to 3 in place. Moreover, note that, in constructing $G(u,v)$, although we are removing the edge between u and v, it might be the case that u is still reachable from v through another path, as is the case in the above example.

Note that we have employed the convention that the first component on the right hand side (i.e. $G[v,u]_{h-1}$) is the neighborhood of the first vertex appearing on the left hand side (i.e. u). See Figure 2.4 for an example.

For a rooted marked graph (G,o), integer $h \geq 1$, and $g, g' \in \Xi \times \bar{\mathcal{G}}_*^{h-1}$, we define

$$E_h(g,g')(G,o) := |\{v \sim_G o : \varphi_G^h(o,v) = (g,g')\}|. \tag{2.6}$$

Also, for $[G,o] \in \bar{\mathcal{G}}_*$, we can write $E_h(g,g')([G,o])$ for $E_h(g,g')(G,o)$, where (G,o) is an arbitrary member of $[G,o]$. This notation is well-defined, since $E_h(g,g')(G,o)$, thought of as a function of (G,o) for fixed integer $h \geq 1$ and $g,g' \in \Xi \times \bar{\mathcal{G}}_*^{h-1}$, is invariant under rooted isomorphism.

For $h \geq 1$, $P \in \mathcal{P}(\bar{\mathcal{G}}_*^h)$, and $g, g' \in \Xi \times \bar{\mathcal{G}}_*^{h-1}$, define

$$e_P(g,g') := \mathbb{E}_P\left[E_h(g,g')(G,o)\right].$$

Here, (G,o) is a member of the isomorphism class $[G,o]$ that has law P. This notation is well-defined for the same reason as above.

Definition 2.3. *Let $h \geq 1$. A probability distribution $P \in \mathcal{P}(\bar{\mathcal{G}}_*^h)$ is called* admissible *if $\mathbb{E}_P\left[\deg_G(o)\right] < \infty$ and $e_P(g,g') = e_P(g',g)$ for all $g,g' \in \Xi \times \bar{\mathcal{G}}_*^{h-1}$.*

The following simple lemma indicates the importance of the concept of admissibility.

Figure 2.4: $\varphi^3_G(1,3)$ for the graph in Figure 2.3, with the first component on the left and the second component on the right. Note that the order in setting the notation $\varphi^h_G(u,v)$ is chosen so that the first component ($G[3,1]_2$ here) is the neighborhood of the first vertex mentioned in the notation (1 in this example), and the second component is the neighborhood of the vertex mentioned second (3 in this example). Also note that the subgraph part of each of the two components of $\varphi^3_G(1,3)$ in this example is an equivalence class, which is the reason why there are no vertex labels.

Lemma 2.3. *Let $h \geq 1$, and let $\mu \in \mathcal{P}_u(\bar{\mathcal{G}}_*)$ be a unimodular probability measure with* $\deg(\mu) < \infty$. *Let $P := \mu_h$. Then P is admissible.*

Proof. Using the definition of unimodularity, for $g, g' \in \Xi \times \bar{\mathcal{G}}^{h-1}_*$, we have

$$
\begin{aligned}
e_P(g,g') &= \mathbb{E}_\mu\left[\sum_{v \sim_G o} \mathbb{1}\left[\varphi^h_G(o,v) = (g,g')\right]\right] \\
&= \mathbb{E}_\mu\left[\sum_{v \sim_G o} \mathbb{1}\left[\varphi^h_G(v,o) = (g,g')\right]\right] = \mathbb{E}_\mu\left[\sum_{v \sim_G o} \mathbb{1}\left[\varphi^h_G(o,v) = (g',g)\right]\right] = e_P(g',g).
\end{aligned}
$$

□

For the case of rooted marked trees, all the above notation can be defined similarly by substituting for $\bar{\mathcal{G}}_*$ with $\bar{\mathcal{T}}_*$, since $\bar{\mathcal{T}}_* \subset \bar{\mathcal{G}}_*$. While we have defined the notion of an admissible probability distribution P for $P \in \mathcal{P}(\bar{\mathcal{G}}^h_*)$, $h \geq 1$, we will soon see that it suffices to be focused on the case $P \in \mathcal{P}(\bar{\mathcal{T}}^h_*)$, $h \geq 1$.

For $t, t' \in \Xi \times \bar{\mathcal{T}}_*$, define $t \oplus t' \in \bar{\mathcal{T}}_*$ as the isomorphism class of the rooted tree (T, o) where o has a subtree isomorphic to $t[s]$, and o has an extra offspring v where the subtree rooted at v is isomorphic to $t'[s]$. Furthermore, $\xi_T(v,o) = t[m]$ and $\xi_T(o,v) = t'[m]$. See Figure 2.5 for an example. Note that, in general, $t \oplus t'$ is different from $t' \oplus t$. Also, note that if $t \in \Xi \times \bar{\mathcal{T}}^k_*$ and $t' \in \Xi \times \bar{\mathcal{T}}^l_*$, then we have $t \oplus t' \in \bar{\mathcal{T}}^{\max\{k, l+1\}}_*$.

The operation $\oplus$ described above helps to elucidate the structure of marked rooted trees of fixed depth, i.e. members of $\bar{\mathcal{T}}^h_*$, $h \geq 0$. Some of their properties are gathered in Appendix A.2.

Now, for $h \geq 1$, given an admissible $P \in \mathcal{P}(\bar{\mathcal{T}}^h_*)$, we define a Borel probability measure $\mathsf{UGWT}_h(P) \in \mathcal{P}(\bar{\mathcal{T}}_*)$, which is called the *marked unimodular Galton–Watson tree with depth*

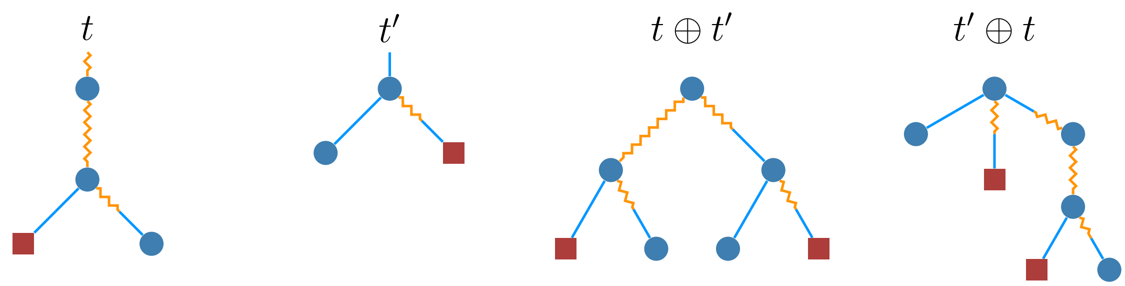

Figure 2.5: $t \oplus t'$ and $t' \oplus t$ for $t, t' \in \Xi \times \bar{\mathcal{T}}_*$ depicted on the left. We have employed our general convention in drawing objects in $\Xi \times \bar{\mathcal{T}}_*$, which is to draw the mark component as a half edge towards the root. In the figures for $t \oplus t'$ and $t' \oplus t$ the root is vertex at the top of the figure. Note that in this example $t \oplus t'$ is different from $t' \oplus t$.

h neighborhood distribution P, as follows. For $t, t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$ such that $e_P(t, t') > 0$, define $\widehat{P}_{t,t'} \in \mathcal{P}(\Xi \times \bar{\mathcal{T}}_*^h)$ via:

$$\widehat{P}_{t,t'}(\tilde{t}) := \mathbb{1}\left[\tilde{t}_{h-1} = t\right] \frac{P(\tilde{t} \oplus t') E_h(t, t')(\tilde{t} \oplus t')}{e_P(t, t')}, \qquad \text{for } \tilde{t} \in \Xi \times \bar{\mathcal{T}}_*^h. \tag{2.7}$$

Moreover, in case $e_P(t, t') = 0$, we define $\widehat{P}_{t,t'}(\tilde{t}) = \mathbb{1}\left[\tilde{t} = t\right]$.

We first check that $\widehat{P}_{t,t'}(\tilde{t})$ defines a probability distribution over $\tilde{t}$. This is clear when $e_P(t, t') = 0$, so assume that $e_P(t, t') > 0$. By definition, we have

$$e_P(t, t') = \sum_{t'' \in \bar{\mathcal{T}}_*^h} P(t'') E_h(t, t')(t'').$$

Note that $E_h(t, t')(t'') > 0$ iff for some $\tilde{t} \in \Xi \times \bar{\mathcal{T}}_*^h$ with $\tilde{t}_{h-1} = t$, we have $t'' = \tilde{t} \oplus t'$. Also, it is easy to see that two different $\tilde{t}^{(1)}$ and $\tilde{t}^{(2)}$ in $\Xi \times \bar{\mathcal{T}}_*^h$ with $\tilde{t}^{(1)}_{h-1} = \tilde{t}^{(2)}_{h-1} = t$ give rise to different objects $\tilde{t}^{(1)} \oplus t'$ and $\tilde{t}^{(2)} \oplus t'$. This readily implies that summing $\widehat{P}_{t,t'}(\tilde{t})$ over all $\tilde{t} \in \Xi \times \bar{\mathcal{T}}_*^h$ such that $\tilde{t}_{h-1} = t$ gives 1, and hence $\hat{P}_{t,t'}(\tilde{t})$ is a probability distribution over $\tilde{t}$.

With this, we define $\mathsf{UGWT}_h(P)$ to be the law of $[T, o]$ where (T, o) is the random rooted marked tree constructed as follows. First, we sample the h neighborhood of the root, $(T, o)_h$, according to P. Then, for each offspring $v \sim_T o$ of the root, we sample $\tilde{t} \in \Xi \times \bar{\mathcal{T}}_*^h$ according to the law $\widehat{P}_{t,t'}(.)$ where $t = T[o, v]_{h-1}$ and $t' = T[v, o]_{h-1}$. Note that, by definition, we have $\tilde{t}_{h-1} = t$. This means that the subtree component of $\tilde{t}$ agrees with the subtree component of t up to depth $h-1$. This allows us to add at most one layer to $T(o, v)_{h-1}$ so that $T(o, v)_h \equiv \tilde{t}$. We carry out the same procedure independently for each offspring of the root. At this step, the rooted tree has depth at most $h+1$. Subsequently, we follow the same procedure for vertices at depth 2, 3, and so on inductively to construct (T, o). More specifically, for a vertex v at depth k of (T, o) with parent w, we sample $\tilde{t}$ from $\widehat{P}_{t,t'}(.)$ with $t = T[w, v]_{h-1}$

and $t' = T[v,w]_{h-1}$. Since by definition, we have $\tilde{t}_{h-1} = t$, we can add at most one layer to $T(w,v)_{h-1}$ so that $T(w,v)_h \equiv \tilde{t}$. We do this independently for all vertices at depth k. If, at the time we do the above procedure for vertices at depth k, there is no vertex at that depth, we stop the procedure. Finally, we define $\mathsf{UGWT}_h(P)$ to be the law of $[T,o]$.

As shown in Corollary A.1 in Appendix A.3, if $[T,o]$ is outside a measure zero set with respect to $\mathsf{UGWT}_h(P)$, for all vertices $v \in V(T) \setminus \{o\}$ we have $e_P(t,t') > 0$ where $t = T[w,v]_{h-1}$ and $t' = T[v,w]_{h-1}$, with w being the parent of v. This means that the need to refer to the definition of $\widehat{P}_{t,t'}$ when $e_P(t,t') = 0$ will not arise, with probability 1.

For each integer $h \geq 1$, the probability distribution $\mathsf{UGWT}_h(P) \in \mathcal{P}(\bar{\mathcal{T}}_*)$ satisfies a useful continuity property in its defining admissible probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$. This is stated in the following Lemma 2.4, whose proof is in Appendix A.4.

Lemma 2.4. *Let $h \geq 1$. Assume that an admissible probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ together with a sequence of admissible probability distributions $P^{(n)} \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ are given such that $P^{(n)} \Rightarrow P$ and, for all $t, t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$, we have $e_{P^{(n)}}(t,t') \to e_P(t,t')$. Then we have $\mathsf{UGWT}_h(P^{(n)}) \Rightarrow \mathsf{UGWT}_h(P)$.*

The following Lemma 2.5 justifies the terminology used for the probability distribution $\mathsf{UGWT}_h(P) \in \mathcal{P}(\bar{\mathcal{T}}_*)$ constructed from an admissible probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$, by establishing that $\mathsf{UGWT}_h(P)$ is unimodular. The proof is given in Appendix A.5.

Lemma 2.5. *Let $h \geq 1$. For an admissible probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$, let $\mathsf{UGWT}_h(P) \in \mathcal{P}(\bar{\mathcal{T}}_*)$ denote the marked unimodular Galton–Watson tree with depth h neighborhood distribution P. Then $\mathsf{UGWT}_h(P)$ is a unimodular distribution.*

The following proposition states a key property of the probability distribution $\mathsf{UGWT}_h(P)$, which should be reminiscent of a finite order Markov property. This is an important result for understanding the structure of $\mathsf{UGWT}_h(P)$. The proof, which is provided in Appendix A.6, is very similar to the proof of the second part of Proposition 1.1 in [BC15].

Proposition 2.1. *Let $h \geq 1$ and let $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$, i.e. P is an admissible probability distribution. Then, for all $k \geq h$, we have*

$$\mathsf{UGWT}_k((\mathsf{UGWT}_h(P))_k) = \mathsf{UGWT}_h(P). \tag{2.8}$$

The following proposition is not used in any way in the subsequent discussion. The proof depends on several results to be developed during the course of this document, and is provided in Appendix A.8.

Proposition 2.2. *Given an integer $h \geq 1$ and an admissible probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$, the probability distribution $\mathsf{UGWT}_h(P)$ is sofic.*

For $h \geq 1$ and admissible $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ such that $d := \mathbb{E}_P[\deg_T(o)] > 0$, let π_P denote the probability distribution on $(\Xi \times \bar{\mathcal{T}}_*^{h-1}) \times (\Xi \times \bar{\mathcal{T}}_*^{h-1})$ defined as

$$\pi_P(t,t') := \frac{e_P(t,t')}{d}.$$

Since for each $[T,o] \in \bar{\mathcal{T}}_*$ we have

$$\deg_T(o) = \sum_{t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}} E_h(t,t')(T,o),$$

we have $d = \sum_{t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}} e_P(t,t')$. Consequently, π_P is indeed a probability distribution.

For $h \geq 1$ and admissible $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ with $H(P) < \infty$ and $\mathbb{E}_P[\deg_T(o)] > 0$, define

$$J_h(P) := -s(d) + H(P) - \frac{d}{2} H(\pi_P) - \sum_{t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}} \mathbb{E}_P[\log E_h(t,t')!]\,, \tag{2.9}$$

where $d := \mathbb{E}_P[\deg_T(o)]$ is the average degree at the root and $s(d) = \frac{d}{2} - \frac{d}{2}\log d$. Note that $s(d)$ is finite, since $d < \infty$. Also, $H(P) < \infty$, $H(\pi_P) \geq 0$, and for each $t, t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$, $\mathbb{E}_P[\log E_h(t,t')!] \geq 0$. Thereby, $J_h(P)$ is well-defined and is in the range $[-\infty, \infty)$.

Definition 2.4. *For integer $h \geq 1$, we say that a probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ is* strongly admissible *if P is admissible, $H(P) < \infty$, and $\mathbb{E}_P[\deg_T(o) \log \deg_T(o)] < \infty$. Let $\mathcal{P}_h$ denote the set of strongly admissible probability distributions $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$.*

In part 2 of Corollary 2.1 of Lemma 2.6 below, we show that, for $h \geq 1$ and $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$, the admissibility of P, together with the condition $\mathbb{E}_P[\deg_T(o) \log \deg_T(o)] < \infty$ is necessary and sufficient for P to be strongly admissible, i.e. $P \in \mathcal{P}_h$. Namely, the requirement that $H(P) < \infty$ in the definition of strong admissibility of P is automatic given the other requirements, and need not be explicitly imposed.

In particular, this means that, for a unimodular $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, if $\mathbb{E}_\mu[\deg_T(o) \log \deg_T(o)] < \infty$, then for all $h \geq 1$ we have $\mu_h \in \mathcal{P}_h$. This is because μ being unimodular with $\deg(\mu) < \infty$ implies that μ_h is admissible for all $h \geq 1$, as we show in Lemma 2.3.

The proof of Lemma 2.6 below is given in Appendix A.7.

Lemma 2.6. *Given a unimodular $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ and an integer $h \geq 1$, assume that with $P := \mu_h$, we have P is strongly admissible, i.e. $P \in \mathcal{P}_h$. Then, with $\widetilde{P} := \mu_{h+1}$, we have $\widetilde{P} \in \mathcal{P}_{h+1}$.*

Corollary 2.1. *The following hold:*

1. *Assume that for a unimodular measure $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, we have $\mathbb{E}_\mu[\deg_T(o) \log \deg_T(o)] < \infty$. Then, for all integers $h \geq 1$, we have $\mu_h \in \mathcal{P}_h$, i.e. μ_h is strongly admissible.*

2. *Let $h \geq 1$ and $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$. Then $P \in \mathcal{P}_h$, i.e. P being strongly admissible, is equivalent to P admissible and $\mathbb{E}_P[\deg_T(o) \log \deg_T(o)] < \infty$.*

Proof. To prove part 1, let $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ with $\mathbb{E}_\mu[\deg_T(o) \log \deg_T(o)] < \infty$. By Lemma 2.6, to show that $\mu_h \in \mathcal{P}_h$ for all $h \geq 1$, it suffices to show that $\mu_1 \in \mathcal{P}_1$. Let $P := \mu_1$. From $\mathbb{E}_\mu[\deg_T(o) \log \deg_T(o)] < \infty$ we have $\mathbb{E}_\mu[\deg_T(o)] < \infty$. Since $\deg(\mu) = \mathbb{E}_\mu[\deg_T(o)]$, from Lemma 2.3 we see that P is admissible. We also have $\mathbb{E}_P\left[\deg_T(o)\log\deg_T(o)\right] = \mathbb{E}_\mu\left[\deg_T(o)\log\deg_T(o)\right] < \infty$. By the definition of strong admissibility in Definition 2.4, all that remains to show is that $H(P) < \infty$. For this, observe that a rooted tree $[T, o] \in \bar{\mathcal{T}}_*^1$ is uniquely determined by knowing the integers

$$N_{x,x'}^{\theta,\theta'}(T,o) := |\{v \sim_T o : \xi_T(v,o) = x, \tau_T(o) = \theta, \xi_T(o,v) = x', \tau_T(v) = \theta'\}|,$$

for all $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$. On the other hand, for $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$, $\mathbb{E}_P\left[N_{x,x'}^{\theta,\theta'}(T,o)\right] \le \mathbb{E}_P\left[\deg_T(o)\right] < \infty$. Consequently, when $[T, o] \sim P$, the entropy of the random variable $N_{x,x'}^{\theta,\theta'}(T,o)$ is finite. To see this, for $k \ge 0$, let p_k denote the probability under P that $N_{x,x'}^{\theta,\theta'}(T,o) = k$. Furthermore, let $q_k := \frac{1}{2^{k+1}}$. Then we have

$$H(N_{x,x'}^{\theta,\theta'}(T,o)) = \sum_{k=0}^{\infty} p_k \log\frac{1}{p_k} \overset{(a)}{\le} \sum_{k=0}^{\infty} p_k \log\frac{1}{q_k} = \left(1 + \mathbb{E}_P\left[N_{x,x'}^{\theta,\theta'}(T,o)\right]\right)\log 2 < \infty,$$

where step (a) comes from Gibbs' inequality, i.e. the nonnegativity of relative entropy, $\sum_{k=0}^{\infty} p_k \log\frac{p_k}{q_k} \ge 0$. Since Ξ and Θ are finite sets, we have $H(P) < \infty$, which completes the proof of part 1.

To see part 2, first note that, by definition, if $P \in \mathcal{P}_h$ then P is admissible and $\mathbb{E}_P\left[\deg_T(o)\log\deg_T(o)\right] < \infty$. To show the other direction, define $\mu := \mathsf{UGWT}_h(P)$. By Lemma 2.5 we have $\mu \in \mathcal{P}_u(\mathcal{T}_*)$. Further, we have

$$\mathbb{E}_\mu\left[\deg_T(o)\log\deg_T(o)\right] = \mathbb{E}_P\left[\deg_T(o)\log\deg_T(o)\right] < \infty.$$

Consequently, the first part of this corollary implies that $P = \mu_h \in \mathcal{P}_h$, and this completes the proof. □

2.8 Conclusion

In this chapter, we reviewed the framework of local weak convergence. We saw that this framework introduces a notion of convergence for sparse marked graphs by studying the local neighborhood structure of a typical vertex. Therefore, the limit object is a probability distribution on the space of unlabeled marked rooted graphs. We also discussed unimodularity, which can be considered as a certain stationarity condition for a probability distribution on the space of unlabeled marked rooted graphs, and is a necessary condition for such an object to appear as the local weak limit of a sequence of finite graphs. Furthermore, we introduced the marked unimodular Galton–Watson trees, which form an important class of unimodular distributions.

Chapter 3

The Marked BC Entropy

In this chapter, we introduce a generalization of the notion of entropy defined in [BC15] for the marked regime discussed in Chapter 2. Our entropy function is going to be defined for probability distributions $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ with $0 < \deg(\mu) < \infty$. We call this notion of entropy the *marked BC entropy* after Bordenave and Caputo. In Section 3.1, we make the initial steps towards defining our notion of entropy. Then, in Section 3.2, we give the formal definition of the entropy in the marked regime and state its properties. Section 3.3 introduces a generalization of the classical configuration model which is going to be crucial in proving some properties of the marked BC entropy. In Section 3.4, we prove the main properties of the marked BC entropy. Finally, we conclude the chapter in Section 3.5.

3.1 Towards the Definition of the Marked BC Entropy

In this section, we make the initial steps towards defining our notion of entropy. Let the finite edge and vertex mark sets Ξ and Θ respectively be given. An edge mark count vector is defined to be a vector of nonnegative integers $\vec{m} := (m(x, x') : x, x' \in \Xi)$ such that $m(x, x') = m(x', x)$ for all $x, x' \in \Xi$. A vertex mark count vector is defined to be a vector of nonnegative integers $\vec{u} := (u(\theta) : \theta \in \Theta)$. Since Ξ is finite, we may assume it is an ordered set. We define $\|\vec{m}\|_1 := \sum_{x \le x' \in \Xi} m(x, x')$ and $\|\vec{u}\|_1 := \sum_{\theta \in \Theta} u(\theta)$.

For an integer $n \in \mathbb{N}$ and edge mark and vertex mark count vectors $\vec{m}$ and $\vec{u}$, define $\mathcal{G}^{(n)}_{\vec{m},\vec{u}}$ to be the set of marked graphs on the vertex set $[n]$ such that $\vec{m}_G = \vec{m}$ and $\vec{u}_G = \vec{u}$. Note that $\mathcal{G}^{(n)}_{\vec{m},\vec{u}}$ is empty unless $\|\vec{m}\|_1 \le \binom{n}{2}$ and $\|\vec{u}\|_1 = n$. Furthermore, if these two conditions are satisfied, it is easy to see that

$$|\mathcal{G}^{(n)}_{\vec{m},\vec{u}}| = \frac{n!}{\prod_{\theta \in \Theta} u(\theta)!} \times \frac{\frac{n(n-1)}{2}!}{\prod_{x \le x' \in \Xi} m(x, x')! \times \left(\frac{n(n-1)}{2} - \|\vec{m}\|_1\right)!} \times 2^{\sum_{x < x' \in \Xi} m(x,x')}. \tag{3.1}$$

An *average degree vector* is defined to be a vector of nonnegative reals $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ such that for all $x, x' \in \Xi$, we have $d_{x,x'} = d_{x',x}$. Moreover, we require that $\sum_{x,x' \in \Xi} d_{x,x'} > 0$.

Definition 3.1. *Given an average degree vector $\vec{d}$ and a probability distribution $Q = (q_\theta : \theta \in \Theta)$, we say that a sequence $(\vec{m}^{(n)}, \vec{u}^{(n)})$ of edge mark count vectors and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ is adapted to $(\vec{d}, Q)$, if the following conditions hold:*

1. *For each n, we have $\|\vec{m}^{(n)}\|_1 \leq \binom{n}{2}$ and $\|\vec{u}^{(n)}\|_1 = n$.*
2. *For $x \in \Xi$, we have $m^{(n)}(x, x)/n \to d_{x,x}/2$.*
3. *For $x \neq x' \in \Xi$, we have $m^{(n)}(x, x')/n \to d_{x,x'} = d_{x',x}$.*
4. *For $\theta \in \Theta$, we have $u^{(n)}(\theta)/n \to q_\theta$.*
5. *For $x, x' \in \Xi$, $d_{x,x'} = 0$ implies $m^{(n)}(x, x') = 0$ for all n.*
6. *For $\theta \in \Theta$, $q_\theta = 0$ implies $u^{(n)}(\theta) = 0$ for all n.*

If $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are sequences such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$ then, using Stirling's approximation, we have

$$\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}| = \|\vec{m}^{(n)}\|_1 \log n + nH(Q) + n \sum_{x,x' \in \Xi} s(d_{x,x'}) + o(n), \tag{3.2}$$

where

$$s(d) := \begin{cases} \frac{d}{2} - \frac{d}{2} \log d & d > 0, \\ 0 & d = 0. \end{cases}$$

See Appendix B.1 for the details on how to derive (3.2). To simplify the notation, we may write $s(\vec{d})$ for $\sum_{x,x' \in \Xi} s(d_{x,x'})$.

To lead up to the definition of the BC entropy in Definition 3.3, we now give the definitions of upper and lower BC entropy.

Definition 3.2. *Assume $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is given, with $0 < \deg(\mu) < \infty$. For $\epsilon > 0$, and edge and vertex mark count vectors $\vec{m}$ and $\vec{u}$, define*

$$\mathcal{G}^{(n)}_{\vec{m}, \vec{u}}(\mu, \epsilon) := \{G \in \mathcal{G}^{(n)}_{\vec{m}, \vec{u}} : d_{LP}(U(G), \mu) < \epsilon\}.$$

Fix an average degree vector $\vec{d}$ and a probability distribution $Q = (q_\theta : \theta \in \Theta)$, and also fix sequences of edge and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$. With these, define

$$\overline{\Sigma}_{\vec{d},Q}(\mu, \epsilon)|_{(\vec{m}^{(n)}, \vec{u}^{(n)})} := \limsup_{n \to \infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon)| - \|\vec{m}^{(n)}\|_1 \log n}{n},$$

which we call the ϵ–upper BC entropy. Since this is increasing in ϵ, we can define the upper BC entropy *as*

$$\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} := \lim_{\epsilon \downarrow 0} \overline{\Sigma}_{\vec{d},Q}(\mu,\epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}.$$

We may define the ϵ–lower BC entropy $\underline{\Sigma}_{\vec{d},Q}(\mu,\epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ similarly as

$$\underline{\Sigma}_{\vec{d},Q}(\mu,\epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} := \liminf_{n\to\infty} \frac{\log|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)| - \|\vec{m}^{(n)}\|_1 \log n}{n}.$$

Since this is increasing in ϵ, we can define the lower BC entropy $\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ *as*

$$\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} := \lim_{\epsilon \downarrow 0} \underline{\Sigma}_{\vec{d},Q}(\mu,\epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}.$$

To close this section, we prove an upper semicontinuity result that will be superseded later by the upper semicontinuity result of Theorem 3.4.

Lemma 3.1. *Assume that a sequence $\mu_k \in \mathcal{P}(\bar{\mathcal{G}}_*)$ together with $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ are given such that $\mu_k \Rightarrow \mu$. Let $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ be an average degree vector and $Q = (q_\theta : \theta \in \Theta)$ a probability distribution. Let $\vec{m}^{(n)}, \vec{u}^{(n)}$ be sequences such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$. Then, we have*

$$\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq \limsup_{k\to\infty} \underline{\Sigma}_{\vec{d},Q}(\mu_k)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}.$$

Proof. For $\epsilon > 0$, let $B(\mu,\epsilon)$ denote the ball around μ of radius ϵ with respect to the Lévy–Prokhorov distance. Since $\bar{\mathcal{G}}_*$ is Polish, weak convergence in $\mathcal{P}(\bar{\mathcal{G}}_*)$ is equivalent to convergence with respect to the Lévy–Prokhorov metric. Hence, for $\epsilon > 0$, $\mu_k \Rightarrow \mu$ implies that for k large enough, we have $B(\mu,\epsilon) \supseteq B(\mu_k,\epsilon/2)$. Therefore, we have $|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu_k,\epsilon/2)| \leq |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)|$. Consequently,

$$\underline{\Sigma}_{\vec{d},Q}(\mu,\epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq \underline{\Sigma}_{\vec{d},Q}(\mu_k,\epsilon/2)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq \underline{\Sigma}_{\vec{d},Q}(\mu_k)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}.$$

Taking the limsup on the right hand side and then sending ϵ to zero on the left hand side, we get the desired result. □

3.2 Definition of the Marked BC Entropy and Main Results

In this section, we state the main theorems proved in this document. These theorems establish properties of the upper and lower marked BC entropy, which enable us to define the marked BC entropy and establish some of its properties. The main propositions that are used to prove these theorems are also stated in this section and we give the proofs of these theorems, assuming that the propositions are proved. The proofs of the propositions themselves are given later in the document.

The following Theorem 3.1 shows that certain conditions must be met for the marked BC entropy to be of interest.

Theorem 3.1. *Let an average degree vector* $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ *and a probability distribution* $Q = (q_\theta : \theta \in \Theta)$ *be given. Suppose* $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ *with* $0 < \deg(\mu) < \infty$ *satisfies any one of the following conditions:*

1. μ *is not unimodular.*
2. μ *is not supported on* $\bar{\mathcal{T}}_*$.
3. $\deg_{x,x'}(\mu) \neq d_{x,x'}$ *for some* $x, x' \in \Xi$*, or* $\Pi_\theta(\mu) \neq q_\theta$ *for some* $\theta \in \Theta$.

Then, for any choice of the sequences $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *such that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{d}, Q)$*, we have* $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$.

Theorem 3.1 is proved by means of Propositions 3.1 and 3.2 below.

Proposition 3.1. *Assume that* $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ *with* $0 < \deg(\mu) < \infty$ *is given. Also, assume that a degree vector* $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ *and a probability distribution* $Q = (q_\theta : \theta \in \Theta)$ *are given. Let* $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *be sequences such that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{d}, Q)$*. If* μ *is not unimodular, or* $\vec{d} \neq \vec{\deg}(\mu)$*, or* $Q \neq \vec{\Pi}(\mu)$*, we have* $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$.

Proposition 3.2. *Assume* $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ *with* $0 < \deg(\mu) < \infty$ *is given such that* $\mu(\bar{\mathcal{T}}_*) < 1$*. Then, if* $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *are any sequences such that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$*, we have* $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$.

The proofs of these statements are given in Section 3.4.1 and it is immediate to see that they prove Theorem 3.1. A consequence of Theorem 3.1 is that the only case of interest in the discussion of marked BC entropy is when $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, $\vec{d} = \vec{\deg}(\mu)$, $Q = \vec{\Pi}(\mu)$, and the sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. Namely, the only upper and lower marked BC entropies of interest are $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ and $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ respectively.

The following Theorem 3.2 establishes that the upper and lower marked BC entropies do not depend on the choice of the defining pair of sequences $(\vec{m}^{(n)}, \vec{u}^{(n)})$. Further, this theorem establishes that the upper marked BC entropy is always equal to the lower marked BC entropy,

Theorem 3.2. *Assume that an average degree vector* $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ *together with a probability distribution* $Q = (q_\theta : \theta \in \Theta)$ *are given. For any* $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ *such that* $0 < \deg(\mu) < \infty$*, we have*

1. *The values of* $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ *and* $\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ *are invariant under the specific choice of the sequences* $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *such that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{d}, Q)$*. With this, we may simplify the notation and unambiguously write* $\overline{\Sigma}_{\vec{d},Q}(\mu)$ *and* $\underline{\Sigma}_{\vec{d},Q}(\mu)$.

2. $\overline{\Sigma}_{\vec{d},Q}(\mu) = \underline{\Sigma}_{\vec{d},Q}(\mu)$. *We may therefore unambiguously write* $\Sigma_{\vec{d},Q}(\mu)$ *for this common value, and call it the* marked BC entropy *of* $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ *for the average degree vector* $\vec{d}$ *and a probability distribution* $Q = (q_\theta : \theta \in \Theta)$. *Moreover,* $\Sigma_{\vec{d},Q}(\mu) \in [-\infty, s(\vec{d}) + H(Q)]$.

From Theorem 3.1 we conclude that unless $\vec{d} = \vec{\deg}(\mu)$, $Q = \vec{\Pi}(\mu)$, and μ is a unimodular measure on $\bar{\mathcal{T}}_*$, we have $\Sigma_{\vec{d},Q}(\mu) = -\infty$. In view of this, for $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ with $\deg(\mu) < \infty$, we write $\Sigma(\mu)$ for $\Sigma_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)$. Likewise, we may write $\underline{\Sigma}(\mu)$ and $\overline{\Sigma}(\mu)$ for $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)$ and $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)$, respectively. Note that, unless $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, we have $\overline{\Sigma}(\mu) = \underline{\Sigma}(\mu) = \Sigma(\mu) = -\infty$.

We are now in a position to define the marked BC entropy.

Definition 3.3. *For* $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ *with* $0 < \deg(\mu) < \infty$, *the marked BC entropy of* μ *is defined to be* $\Sigma(\mu)$.

Next, we give a recipe to compute the marked BC entropy for the marked unimodular Galton–Watson trees defined in Section 2.7. We also characterize the marked BC entropy of any $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ in terms of the marked BC entropies of the marked unimodular Galton–Watson trees with neighborhood distribution given by the truncation of μ up to any depth.

Theorem 3.3. *Let* $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ *be a unimodular probability measure with* $0 < \deg(\mu) < \infty$. *Then,*

1. *If* $\mathbb{E}_\mu [\deg_T(o) \log \deg_T(o)] = \infty$, *then* $\underline{\Sigma}(\mu) = \overline{\Sigma}(\mu) = \Sigma(\mu) = -\infty$.
2. *If* $\mathbb{E}_\mu [\deg_T(o) \log \deg_T(o)] < \infty$, *then, for each* $h \geq 1$, *the probability measure* μ_h *is admissible, and* $H(\mu_h) < \infty$. *Furthermore, the sequence* $(J_h(\mu_h) : h \geq 1)$ *is nonincreasing, and*
$$\underline{\Sigma}(\mu) = \overline{\Sigma}(\mu) = \Sigma(\mu) = \lim_{h\to\infty} J_h(\mu_h).$$

Now, we proceed to state the propositions needed to prove Theorems 3.2 and 3.3 and explain how they prove the two theorems. We give the proofs of these propositions in Section 3.4. Our proof techniques are similar to those given in [BC15].

In view of Propositions 3.1 and 3.2, in order to address parts 1 and 2 of Theorem 3.2, we may assume that $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, $\vec{d} = \vec{\deg}(\mu)$, and $Q = \vec{\Pi}(\mu)$, since otherwise $\underline{\Sigma}_{\vec{d},Q}(\mu) = \overline{\Sigma}_{\vec{d},Q}(\mu) = -\infty$. To prove part 1 of Theorem 3.2, the strategy is to find a lower bound for $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ and an upper bound for $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$, and then to show that they match. We first prove a lower bound for $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ when μ is of the form $\mathsf{UGWT}_h(P)$ for $P \in \mathcal{P}_h$ being strongly admissible.

Proposition 3.3. *Let* $h \geq 1$. *Let* $P \in \mathcal{P}_h$, *i.e.* P *is strongly admissible. Assume that with* $\mu := \mathsf{UGWT}_h(P)$ *we have* $0 < \deg(\mu) < \infty$. *Then, if* $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *are any sequences such* *that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$*, we have*

$$\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq J_h(P).$$

The proof of Proposition 3.3 is given in Section 3.4.2. Now, for a unimodular probability measure $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ such that $0 < \deg(\mu) < \infty$ and $\mathbb{E}_\mu [\deg_T(o) \log \deg_T(o)] < \infty$, Corollary 2.1 implies that, for all $h \geq 1$, μ_h is strongly admissible, i.e. $\mu_h \in \mathcal{P}_h$. In particular, $H(\mu_h) < \infty$ and $J_h(\mu_h)$ is well defined. With this observation in mind, we next give, for each $h \geq 1$, an upper bound for $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$, for $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ such that $0 < \deg(\mu) < \infty$ and $H(\mu_h) < \infty$.

Proposition 3.4. *Let* $h \geq 1$*. Let* $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ *be a unimodular probability measure, with* $0 < \deg(\mu) < \infty$ *and* $H(\mu_h) < \infty$*. Then, if* $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *are sequences such that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$*, we have*

$$\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \leq J_h(\mu_h). \tag{3.3}$$

The proof of Proposition 3.4 is given in Section 3.4.3. Now, we consider the case $\mathbb{E}_\mu [\deg_T(o) \log \deg_T(o)] = \infty$ and show that the marked BC entropy is $-\infty$ in this case.

Proposition 3.5. *Let* $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ *be a unimodular probability measure such that* $0 < \deg(\mu) < \infty$ *and* $\mathbb{E}_\mu [\deg_T(o) \log \deg_T(o)] = \infty$*. Then, if* $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *are sequences such that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$*, we have*

$$\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty. \tag{3.4}$$

Proposition 3.5 is proved in Section 3.4.4.

We now demonstrate how the propositions in this section can be used to prove Theorems 3.2 and 3.3. We have already observed that Propositions 3.1 and 3.2 imply that, in order to address parts 1 and 2 of Theorem 3.2, we may assume that $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$, $\vec{d} = \vec{\deg}(\mu)$, and $Q = \vec{\Pi}(\mu)$. Proposition 3.5 then immediately implies parts 1 and 2 of Theorem 3.2 and part 1 of Theorem 3.3, for every $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ for which $0 < \deg(\mu) < \infty$ and $\mathbb{E}_\mu [\deg_G(o) \log \deg_G(o)] = \infty$.

Thus it remains to consider the case of unimodular $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ with $0 < \deg(\mu) < \infty$ and $\mathbb{E}_\mu [\deg_T(o) \log \deg_T(o)] < \infty$. We have already observed that Corollary 2.1 implies that for such μ, for all $h \geq 1$, μ_h is strongly admissible, i.e. $\mu_h \in \mathcal{P}_h$ and that this implies, in particular, that $H(\mu_h) < \infty$ and $J_h(\mu_h)$ is well defined.

We first show that, in this case, the sequence $J_h(\mu_h)$ is nonincreasing in h. For $h \geq 1$, let $\nu^{(h)} := \mathsf{UGWT}_h(\mu_h)$. Observe that $\vec{\deg}(\mu) = \vec{\deg}(\nu^{(h)})$ and $\vec{\Pi}(\mu) = \vec{\Pi}(\nu^{(h)})$. From Propositions 3.3 and 3.4, we have

$$\begin{aligned} J_{h+1}(\mu_{h+1}) &\leq \underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\nu^{(h+1)})|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \\ &\leq \overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\nu^{(h+1)})|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \end{aligned}$$

$$\leq J_h((\nu^{(h+1)})_h)$$
$$= J_h(\mu_h),$$

where the last equality uses the fact that $(\nu^{(h+1)})_h = (\mathsf{UGWT}_{h+1}(\mu_{h+1}))_h = \mu_h$, which is proved in Proposition 2.1. Hence, $J_\infty(\mu) := \lim_{h\to\infty} J_h(\mu_h)$ exists. Further, since Proposition 3.4 proves that $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \leq J_h(\mu_h)$ holds for all $h \geq 1$, we get

$$\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \leq J_\infty(\mu).$$

Now, we show that $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq J_\infty(\mu)$. Note that, since $\mu_h \in \mathcal{P}_h$ is strongly admissible, Proposition 3.3 implies that $\underline{\Sigma}_{\vec{\deg}(\nu^{(h)}),\vec{\Pi}(\nu^{(h)})}(\nu^{(h)})|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq J_h(\mu_h) \geq J_\infty(\mu)$, where we have noted that, since $\vec{\deg}(\mu) = \vec{\deg}(\nu^{(h)})$ and $\vec{\Pi}(\mu) = \vec{\Pi}(\nu^{(h)})$, the pair of sequences $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\nu^{(h)}), \vec{\Pi}(\nu^{(h)}))$. On the other hand, $\nu^{(h)} \Rightarrow \mu$. Therefore, using Lemma 3.1, we have

$$\begin{aligned}\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{\vec{m}^{(n)},\vec{u}^{(n)}} &\geq \limsup_{h\to\infty} \underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\nu^{(h)})|_{\vec{m}^{(n)},\vec{u}^{(n)}} \\ &= \limsup_{h\to\infty} \underline{\Sigma}_{\vec{\deg}(\nu^{(h)}),\vec{\Pi}(\nu^{(h)})}(\nu^{(h)})|_{\vec{m}^{(n)},\vec{u}^{(n)}} \\ &\geq J_\infty(\mu)\end{aligned}$$

We have established that $J_\infty(\mu) \leq \underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \leq \overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \leq J_\infty(\mu)$. This, in particular, implies that $\underline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = \overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$. To complete the proof of part 1 of Theorem 3.2 and the proof of part 2 of Theorem 3.3 note that $J_\infty(\mu)$ does not depend on the choice of the sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$.

To complete the proof of part 2 of Theorem 3.2, note that the inequality $\Sigma_{\vec{d},Q}(\mu) \leq s(\vec{d}) + H(Q)$ is a direct consequence of (3.2).

The proof of the propositions stated in this section rely on a generalization of the classical graph configuration model called a *colored configuration model*, which was introduced in [BC15]. In Section 3.3 below, we review this framework and generalize its properties to the marked regime. Using the tools developed in Section 3.3, we give the proof of these propositions in Section 3.4.

To close this section, assuming the truth of all the preceding propositions (which are proved in the subsequent sections), we prove an upper semicontinuity result of marked BC entropy, which supersedes the result of Lemma 3.1.

Theorem 3.4. *Let an average degree vector $\vec{d} = (d_{x,x'} : x, x' \in \Xi)$ and a probability distribution $Q = (q_\theta : \theta \in \Theta)$ be given. For any $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ with $0 < \deg(\mu) < \infty$, the BC entropy $\Sigma_{\vec{d},Q}(.)$ is upper semicontinuous at μ, i.e. if μ_k is a sequence in $\mathcal{P}(\bar{\mathcal{G}}_*)$ converging weakly to $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ such that $0 < \deg(\mu_k) < \infty$ for all k, then we have $\Sigma_{\vec{d},Q}(\mu) \geq \limsup_{k\to\infty} \Sigma_{\vec{d},Q}(\mu_k)$.*

Proof. Let $\vec{m}^{(n)}, \vec{u}^{(n)}$ be sequences such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$. Then, as established in part 1 of Theorem 3.2, $\Sigma_{\vec{d},Q}(\mu)$ equals $\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ and $\Sigma_{\vec{d},Q}(\mu_k)$ equals $\underline{\Sigma}_{\vec{d},Q}(\mu_k)|_{(\vec{m}^{(n)},\vec{u}^{(n)})}$ for all k. The claim is therefore an immediate consequence of Lemma 3.1. □

3.3 Colored Configuration Model

In this section, we review and generalize results from [BC15, Section 4]. First, in Section 3.3.1, we review the notion of *directed colored multigraphs* from [BC15, Section 4.1]. Then, in Section 3.3.2, we review the *colored configuration model* from [BC15, Section 4.2]. In Sections 3.3.3 we review the notion of *colored unimodular Galton–Watson trees* and a local weak convergence result related to such trees, from [BC15, Sections 4.4, 4.5]. In Sections 3.3.4 and 3.3.5, we draw a connection between directed colored multigraphs and marked graphs, generalizing the results in [BC15, Sections 4.6]. We also discuss the colored configuration model arising from the colored degree sequences associated to the directed colored graphs arising from a marked graph. This discussion is used in Section 3.3.6 to prove a weak convergence result for any admissible probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ with finite support, for any $h \geq 1$. Finally, in Section 3.3.7, we use the tools developed in this section to prove a local weak convergence result for marked graphs obtained from a colored configuration model, which will be useful in our analysis in Section 3.4. Note that the terms *color* (defined in this section) and *mark* (defined in Section 2.3) refer to two different concepts and should not be confused with each other.

3.3.1 Directed Colored Multigraphs

Let $L \geq 1$ be a fixed integer, and define $\mathcal{C} := \{(i,j) : 1 \leq i,j \leq L\}$. Each element $(i,j) \in \mathcal{C}$ is interpreted as a *color*. Note that the terms *color* and *mark* refer to different concepts and should not be confused with each other. Let $\mathcal{C}_= := \{(i,i) \in \mathcal{C}\}$, $\mathcal{C}_< := \{(i,j) \in \mathcal{C} : i < j\}$ and $\mathcal{C}_{\neq} := \{(i,j) \in \mathcal{C} : i \neq j\}$. We define $\mathcal{C}_\leq$, $\mathcal{C}_>$, and $\mathcal{C}_\geq$ similarly. For $c := (i,j) \in \mathcal{C}$, we use the notation $\bar{c} := (j,i)$.

We now define a set $\widehat{\mathcal{G}}(\mathcal{C})$ of *directed colored multigraphs* with colors in $\mathcal{C}$, comprised of multigraphs (as defined in Section 2.6) where the edges are colored with elements in $\mathcal{C}$ in a *directionally consistent way*. More precisely, each $G \in \widehat{\mathcal{G}}(\mathcal{C})$ is of the form $G = (V, \omega)$ where V is a finite or a countable vertex set, and $\omega = (\omega_c : c \in \mathcal{C})$ where for each $c \in \mathcal{C}$, $\omega_c : V^2 \to \mathbb{Z}_+$ with the following properties:

1. For $c \in \mathcal{C}_=$, $\omega_c(v,v)$ is even for all $v \in V$, and $\omega_c(u,v) = \omega_c(v,u)$ for all $u,v \in V$.
2. For $c \in \mathcal{C}_{\neq}$, we have $\omega_c(u,v) = \omega_{\bar{c}}(v,u)$ for all $u,v \in V$.
3. For all $u \in V$ and $c \in \mathcal{C}$, $\sum_{v \in V} \omega_c(u,v) < \infty$.

See Figure 3 in [BC15] for an example of an element of $\widehat{\mathcal{G}}(\mathcal{C})$.

For a directed colored multigraph $G = (V, \omega) \in \widehat{\mathcal{G}}(\mathcal{C})$ the associated *colorblind multigraph* is the multigraph $\mathsf{CB}(G) := (V, \bar{\omega})$ on the same vertex set V, where $\bar{\omega} : V^2 \to \mathbb{Z}_+$ is defined via

$$\bar{\omega}(u, v) := \sum_{c \in \mathcal{C}} \omega_c(u, v).$$

It can be checked that $\mathsf{CB}(G)$ is a multigraph, as defined in Section 2.6. Distinct directed colored multigraphs can give rise to the same multigraph as their associated colorblind multigraph, and we can think of each of them as arising from this multigraph by coloring it in a directionally consistent way as expressed in properties 1 and 2.

Given $G \in \widehat{\mathcal{G}}(\mathcal{C})$, if $\mathsf{CB}(G)$ has no multiple edges and no self–loops, i.e. it is a graph, then we call G a *directed colored graph*. We let $\mathcal{G}(\mathcal{C})$ denote the subset of $\widehat{\mathcal{G}}(\mathcal{C})$ comprised of directed colored graphs.

We introduce the notation $\mathcal{M}_L$ for the set of $L \times L$ matrices with nonnegative integer valued entries.

Let $G = (V, \omega) \in \widehat{\mathcal{G}}(\mathcal{C})$, where V is a finite set. For $u \in V$ and $c \in \mathcal{C}$, define

$$D_c^G(u) := \sum_{v \in V} \omega_c(u, v).$$

$D_c^G(u)$ is the number of color c edges going out of the vertex u. Let $D^G(v) := (D_c^G(v) : c \in \mathcal{C})$. Note that $D^G(v) \in \mathcal{M}_L$. $D^G(v)$ is called the *colored degree matrix* of the vertex v. Let $\vec{D}^G := (D^G(v) : v \in V)$. We call $\vec{D}^G$ the *colored degree sequence* corresponding to G.

3.3.2 Colored Configuration Model

Fix an integer $L \geq 1$, and let $\mathcal{C} := \{(i, j) : 1 \leq i, j \leq L\}$ be the associated set of colors. For $n \in \mathbb{N}$, let $\mathcal{D}_n$ be the set of vectors $(D(1), \dots, D(n))$ where, for each $1 \leq i \leq n$, we have $D(i) = (D_c(i) : c \in \mathcal{C}) \in \mathcal{M}_L$ and, further, $S := \sum_{i=1}^n D(i)$ is a symmetric matrix with even coefficients on the diagonal. Note that for $G \in \widehat{\mathcal{G}}(\mathcal{C})$ we have $\vec{D}^G \in \mathcal{D}_n$. Given $\vec{D} = (D(1), \dots, D(n)) \in \mathcal{D}_n$, define $\widehat{\mathcal{G}}(\vec{D})$ to be the set of directed colored multigraphs $G \in \widehat{\mathcal{G}}(\mathcal{C})$ with the vertex set $V = [n]$ such that, for all $i \in [n]$, we have $D^G(i) = D(i)$. Further, given $n \in \mathbb{N}$, $\vec{D} \in \mathcal{D}_n$, and $h \geq 1$, let $\mathcal{G}(\vec{D}, h)$ be the set of directed colored multigraphs $G \in \widehat{\mathcal{G}}(\vec{D})$ such that $\mathsf{CB}(G)$ has no cycles of length $l \leq h$. Note that $\mathcal{G}(\vec{D}, h+1) \subseteq \mathcal{G}(\vec{D}, h)$ for all $h \geq 1$, and that $\mathcal{G}(\vec{D}, 2) \subset \mathcal{G}(\mathcal{C})$.

Now, given $\vec{D} = (D(1), \dots, D(n)) \in \mathcal{D}_n$, we give a recipe to generate a random directed colored multigraph $G \in \widehat{\mathcal{G}}(\mathcal{C})$ such that $\vec{D}^G = \vec{D}$, i.e. a random directed colored multigraph in $\widehat{\mathcal{G}}(\vec{D})$. The procedure is similar to that in the classical configuration model. For each $c \in \mathcal{C}$, let $W_c := \cup_{i=1}^n W_c(i)$ be a set of distinct *half edges* of color c where $|W_c(i)| = D_c(i)$. We think of the half edges in $W_c(i)$ as attached to the vertex i. We require a half edge with color c to get connected to another half edge with color $\bar{c}$. For this, for $c \in \mathcal{C}_<$, let Σ_c be the set of bijections $\sigma_c : W_c \to W_{\bar{c}}$. Since $\vec{D} \in \mathcal{D}_n$, $|W_c| = |W_{\bar{c}}|$ and such bijections exist. Likewise, for $c \in \mathcal{C}_=$, let Σ_c be the set of perfect matchings on the set W_c. Since $\vec{D} \in \mathcal{D}_n$, $|W_c|$ is even and such matchings exist.

Given a choice of $\sigma_c \in \Sigma_c$ for each $c \in \mathcal{C}_\leq$, we write σ for $(\sigma_c : c \in \mathcal{C}_\leq)$. Let Σ denote the product of Σ_c for $c \in \mathcal{C}_\leq$. Given $\sigma \in \Sigma$, we construct a directed colored multigraph, denoted $\Gamma(\sigma)$, as follows. For $c \in \mathcal{C}_<$, if σ_c maps a half edge of color c at vertex u to another half edge of color $\bar{c}$ at vertex v, then we place an edge directed from u towards v having color c and an edge directed from v towards u, having color $\bar{c}$. Here it is allowed that $u = v$. For $c \in \mathcal{C}_=$, if σ_c matches a half edge of color c at vertex u to another half edge of the same color at vertex v, then we place two directed edges, one directed from u towards v, and one directed from v towards u, both with color c. Here also it is allowed that $u = v$.

Note that, for $\sigma \in \Sigma$, the construction above gives $\Gamma(\sigma) \in \widehat{\mathcal{G}}(\vec{D})$. For $\vec{D} \in \mathcal{D}_n$, let $\mathsf{CM}(\vec{D})$ be the law of $\Gamma(\sigma)$ where σ is chosen uniformly at random in Σ.

Theorem 3.5 below is from [BC15], and states a key property of the configuration model defined above. To state that theorem, given a positive integer δ, let $\mathcal{M}_L^{(\delta)}$ denote the set of $L \times L$ matrices with nonnegative integer entries bounded by δ. Assume that $R \in \mathcal{P}(\mathcal{M}_L^{(\delta)})$ is given. Let $\vec{D}^{(n)} = (D^{(n)}(1), \dots, D^{(n)}(n)) \in \mathcal{D}_n$ be a sequence satisfying the following two conditions:

$$D^{(n)}(i) \in \mathcal{M}_L^{(\delta)} \qquad \forall i \in [n], \tag{3.5a}$$

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{D^{(n)}(i)} \Rightarrow R. \tag{3.5b}$$

Theorem 3.5 states that, given the above conditions, for every $h \geq 1$, a positive fraction of random directed colored multigraphs generated from the above configuration model do not have and cycles of length h or less.

Theorem 3.5 (Theorem 4.5 in [BC15]). *Fix $\delta \in \mathbb{N}$, $R \in \mathcal{P}(\mathcal{M}_L^{(\delta)})$, and a sequence $\vec{D}^{(n)}$ satisfying (3.5a) and (3.5b). Let G_n have distribution $\mathsf{CM}(\vec{D}^{(n)})$ on $\widehat{\mathcal{G}}(\vec{D}^{(n)})$. Then, for every $h \geq 1$, there exists $\alpha_h > 0$ such that*

$$\lim_{n \to \infty} \mathbb{P}\left(G_n \in \mathcal{G}(\vec{D}^{(n)}, h)\right) = \alpha_h.$$

To close this section, we give an asymptotic counting for the set $\mathcal{G}(\vec{D}^{(n)}, h)$. This calculation is also from [BC15]. For two sequences a_n and b_n we write $a_n \sim b_n$ if $a_n / b_n \to 1$ as $n \to \infty$. Moreover, for an even integer N, $(N-1)!!$ is defined as $\frac{N!}{(N/2)! 2^{N/2}}$, or equivalently $(N-1) \times (N-3) \times \dots 3 \times 1$. Note that $(N-1)!!$ is the number of perfect matchings on a set of size N.

Corollary 3.1 (Corollary 4.6 in [BC15]). *In the setting of Theorem 3.5, write $S_c^{(n)} := \sum_{i \in [n]} D_c^{(n)}(i)$, which, we recall, form the entries of a symmetric matrix. For all $h \geq 2$ we* *have*

$$|\mathcal{G}(\vec{D}^{(n)}, h)| \sim \alpha_h \frac{\prod_{c \in \mathcal{C}_<} S_c^{(n)}! \prod_{c \in \mathcal{C}_=} (S_c^{(n)} - 1)!!}{\prod_{c \in \mathcal{C}} \prod_{i=1}^n D_c^{(n)}(i)!}.$$

We give a brief sketch of how this counting statement results from Theorem 3.5 and refer the reader to [BC15] for the proof. By construction, $|\Sigma|$, which is the total number of configurations, is equal to $\prod_{c \in \mathcal{C}_<} S_c^{(n)}! \prod_{c \in \mathcal{C}_=} (S_c^{(n)} - 1)!!$. Each directed colored multigraph can be constructed via different configurations. However, every $G \in \mathcal{G}(\vec{D}, h)$ for $h \geq 2$ is a colored graph, i.e. is in $\mathcal{G}(\mathcal{C})$. It is easy to see that, for such G, there are precisely $\prod_{c \in \mathcal{C}} \prod_{i=1}^n D_c^{(n)}(i)!$ many configurations $\sigma \in \Sigma$ for which $\Gamma(\sigma) = G$. Also, from Theorem 3.5, the asymptotic probability of $\Gamma(\sigma)$ being in $\mathcal{G}(\vec{D}^{(n)}, h)$ is α_h. This provides a rough explanation of where Corollary 3.1 comes from.

3.3.3 Colored Unimodular Galton–Watson trees

In this section we review the definition of colored unimodular Galton–Watson trees from [BC15, Section 4.4]. This should not be confused with the notion of marked unimodular Galton–Watson trees defined in Section 2.7. Later, in Section 3.3.7, we explain the connection between the two notions. To reduce the chance of confusion, we employ the notation CUGWT to denote the object constructed here, which is slightly different from the notation used in [BC15].

Given $L \in \mathbb{N}$ and the set of colors $\mathcal{C} := \{(i, j) : 1 \leq i, j \leq L\}$, we first define a set of equivalence classes of rooted directed colored multigraphs, denoted by $\widehat{\mathcal{G}}_*(\mathcal{C})$. Each member of $\widehat{\mathcal{G}}_*(\mathcal{C})$ is of the form $[G, o]$ where $G \in \widehat{\mathcal{G}}(\mathcal{C})$ is connected and o is a distinguished vertex in G. $[G, o]$ denotes the equivalence class corresponding to (G, o) where the equivalence relation is defined through relabeling of the vertices, while preserving the root and the edge structure together with the directed colors. As is discussed in [BC15], the framework of local weak convergence introduced in Section 2.6 for multigraphs can be naturally extended to directed colored multigraphs. An element $[G, o] \in \widehat{\mathcal{G}}_*(\mathcal{C})$ is called a rooted directed colored tree if its associated colorblind multigraph $\mathsf{CB}(G)$ has no cycles.

Recall that $\mathcal{M}_L$ denotes the set of $L \times L$ matrices with nonnegative integer valued entries. Let $P \in \mathcal{P}(\mathcal{M}_L)$ be a probability distribution such that for all $c \in \mathcal{C}$, we have $\mathbb{E}[D_c] = \mathbb{E}[D_{\bar{c}}]$, where $D \in \mathcal{M}_L$ has law P. For $c \in \mathcal{C}$ such that $\mathbb{E}[D_c] > 0$, define $\widehat{P}^c \in \mathcal{P}(\mathcal{M}_L)$ as follows:

$$\widehat{P}^c(M) := \frac{(M_{\bar{c}} + 1) P(M + E^{\bar{c}})}{\mathbb{E}[D_c]}, \tag{3.6}$$

where $E^{\bar{c}} \in \mathcal{M}_L$ denotes the matrix with the entry at coordinate $\bar{c}$ being 1 and all the other entries being zero. If $\mathbb{E}[D_c] = 0$, we set $\widehat{P}^c(M) = 1$ if $M = 0$ and zero otherwise. It is straightforward to check that $\sum_{M \in \mathcal{M}_L} \widehat{P}^c(M) = 1$ for all $c \in \mathcal{C}$.

With this setup, we define the *colored unimodular Galton–Watson tree* $\mathsf{CUGWT}(P) \in \mathcal{P}(\widehat{\mathcal{G}}_*(\mathcal{C}))$ to be the law of $[T, o]$ where (T, o) is a rooted directed colored multigraph defined as follows. We start from the root o and generate $D(o)$ with law P. Then, for each $c \in \mathcal{C}$, we attach $D_c(o)$ many vertex offspring of type c to the root. For each offspring v of type c, we add a directed edge from o to v with color c and another directed edge from v to o with color $\bar{c}$. Subsequently, for an offspring v of type c, we generate $D(v)$ with law $\widehat{P}^c$, independent from all other offspring. Then, we continue the process. Namely, for each $c \in \mathcal{C}$, we add $D_c(v)$ many vertex offspring of type c to v where, for each offspring w of type c, there is an edge directed from v towards w with color c and another edge directed from w towards v with color $\bar{c}$. This process is continued inductively to define $\mathsf{CUGWT}(P)$.

Let $\mathcal{G}_*(\mathcal{C})$ denote the subset of $\widehat{\mathcal{G}}_*(\mathcal{C})$ consisting of equivalence classes of rooted directed colored graphs, i.e. for which the associated colorblind multigraph $\mathsf{CB}(G)$ is a graph, see the end of Section 3.3.1. Note that $\mathsf{CUGWT}(P)$ is supported on $\mathcal{G}_*(\mathcal{C})$. The following result from [BC15] will be useful for our future analysis.

Theorem 3.6 (Theorem 4.8 in [BC15])**.** *Let $R \in \mathcal{P}(\mathcal{M}_L^{(\delta)})$ be given. Let $\vec{D}^{(n)} \in \mathcal{D}_n$ be a sequence satisfying (3.5a) and (3.5b). Moreover, assume that $G_n \in \widehat{\mathcal{G}}(\vec{D}^{(n)})$ has law $\mathsf{CM}(\vec{D}^{(n)})$, and that G_n are jointly defined to be independent on a single probability space. Then, with probability one, $U(G_n) \Rightarrow \mathsf{CUGWT}(R)$. Also, the same result holds when G_n is uniformly sampled from $\mathcal{G}(\vec{D}^{(n)}, h)$, for any $h \geq 2$.*

3.3.4 From a Marked Graph to a Directed Colored Graph and Back

In this section we first associate, for any fixed $h \geq 1$, a specific directed colored graph to a given marked graph, by treating the types of edges, as defined in (2.5), as colors. We also discuss a procedure that, starting with a directed colored graph whose colors can be interpreted in terms of the types for a given $h \geq 1$, returns a marked graph.

For a marked graph G on the vertex set $[n]$ and an integer $h \geq 1$, we define a directed colored graph denoted by $\mathsf{C}(G)$. Let $\mathcal{F} \subset \Xi \times \bar{\mathcal{G}}_*^{h-1}$ be the set of all distinct $G[u,v]_{h-1}$ for adjacent vertices u and v in G. Since G is finite, $\mathcal{F}$ is a finite subset of $\Xi \times \bar{\mathcal{G}}_*^{h-1}$. Therefore, with $L := |\mathcal{F}|$, we can enumerate the elements in $\mathcal{F}$ in some order, with integers $1, \dots, L$. Recall from (2.5) that $\varphi_G^h(u,v) = (G[v,u]_{h-1}, G[u,v]_{h-1})$ is the depth h type of the edge (u,v). Now, we define $\mathsf{C}(G)$ to be a directed colored graph with colors in $\mathcal{C} = \mathcal{F} \times \mathcal{F}$ on the vertex set $[n]$ as follows. For two adjacent vertices u and v in G, in $\mathsf{C}(G)$ we put an edge directed from u towards v with color $\varphi_G^h(u,v)$ and another directed edge from v towards u with color $\varphi_G^h(v,u)$. Since G is simple, $\mathsf{C}(G)$ is a directed colored graph, i.e. $\mathsf{C}(G) \in \mathcal{G}(\mathcal{C})$. In fact, $\mathsf{CB}(\mathsf{C}(G))$ is just the graph which results from G by erasing its marks.

We can also go in the other direction. Fix $h \geq 1$. Let $\mathcal{F} \subset \Xi \times \bar{\mathcal{G}}_*^{h-1}$ be a finite set with cardinality L, whose elements are identified with the integers $1, \dots, L$. Let $\mathcal{C} := \mathcal{F} \times \mathcal{F}$. Given a directed colored graph $H \in \mathcal{G}(\mathcal{C})$, defined on a finite or countable vertex set V, and a sequence $\vec{\beta} = (\beta(v) : v \in V)$ with elements in Θ, we define a marked graph on V, called the *marked color blind* version of $(\vec{\beta}, H)$, denoted by $\mathsf{MCB}_{\vec{\beta}}(H)$, as follows. For any

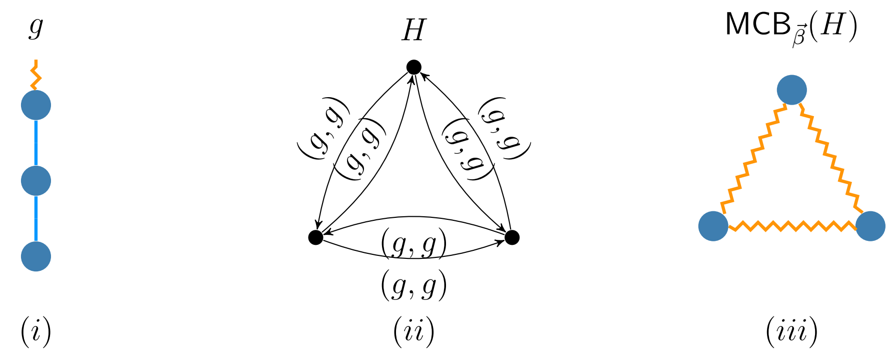

Figure 3.1: (i): $g \in \Xi \times \bar{\mathcal{G}}_*^2$ where $\Theta = \{$●, ■$\}$ and $\Xi = \{$Blue (solid), Orange (wavy)$\}$. (ii): a simple directed colored graph $H \in \mathcal{G}(\mathcal{C})$ where $\mathcal{C} = \{(g, g)\}$, (iii): $G = \mathsf{MCB}_{\vec{\beta}}(H)$ where $\vec{\beta} = \{$●, ●, ●$\}$. Note that none of $\varphi_G^2(1, 2)$, $\varphi_G^2(1, 3)$ and $\varphi_G^2(2, 3)$ is equal to (g, g).

pair of adjacent vertices u and v in H where the color of the edge directed from u to v is (g, g') (and hence the color of the edge directed from v to u is (g', g)), we put an edge in $\mathsf{MCB}_{\vec{\beta}}(H)$ between u and v with the mark towards u and v being $g[m]$ and $g'[m]$, respectively. Moreover, the mark of a vertex $v \in V$ in $\mathsf{MCB}_{\vec{\beta}}(H)$ is defined to be $\beta(v)$.

Note that it is not necessarily the case that the colors of H are consistent with those in the directed colored graph $\mathsf{C}(\mathsf{MCB}_{\vec{\beta}}(H))$. Namely, $\varphi^h_{\mathsf{MCB}_{\vec{\beta}}}(u, v)$ for adjacent vertices u, v can be different from the color of the edge between u and v in H. See Figure 3.1 for an example. Proposition 3.6 below gives conditions under which this consistency holds. To be able to state this result, we first need some definitions and tools, which are gathered in the next section.

3.3.5 Consistency in going from a directed colored graph to a marked graph and back

In this section we first give conditions under which the edge colors of a directed colored graph are related to the edge colors of the directed colored graph derived from its marked colorblind version. This is done in Proposition 3.6. Next, building on this result, we study the configuration model given by the colored degree sequence of the directed colored graph associated to a given marked graph, and relate the marked color blind versions of the directed colored graphs arising as realizations from this configuration model to the original marked graph we started with.

Definition 3.4. *Fix $h \in \mathbb{N}$ and assume $\mathcal{F} \subset \Xi \times \bar{\mathcal{T}}_*^{h-1}$ is a finite set with cardinality L. Define $\mathcal{C} := \mathcal{F} \times \mathcal{F}$. Given a matrix $D = (D_{t,t'} : t, t' \in \mathcal{F}) \in \mathcal{M}_L$ and $\theta \in \Theta$, we say that the pair (θ, D) is "graphical" if there exists $[T, o] \in \bar{\mathcal{T}}_*^h$ such that $\tau_T(o) = \theta$ and, for all $t, t' \in \mathcal{F}$,*

Figure 3.2: (ii) depicts $(\theta, x) \otimes t$ for $t \in \Xi \times \bar{\mathcal{T}}_*^2$ as shown in (i), $\theta = $ ■ and $x =$ Blue (solid). We have used our convention of Figure 2.3 for showing t, i.e. the half edge towards the root is the mark component.

we have $E_h(t,t')(T,o) = D_{t,t'}$. *Moreover, for* $\tilde{t}, \tilde{t}' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$ *such that either* $\tilde{t} \notin \mathcal{F}$ *or* $\tilde{t}' \notin \mathcal{F}$, *we require* $E_h(\tilde{t}, \tilde{t}')(T,o)$ *to be zero.*

From Lemma A.4 in Appendix A.2, $[T,o]$ in the above definition, if it exists, is unique.

Fix an integer $h \geq 1$. For $t \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$, $x \in \Xi$, and $\theta \in \Theta$, define $(\theta, x) \otimes t$ to be the element in $\bar{\mathcal{T}}_*^h$ where the root o has mark θ, and attached to it is one offspring v, where the subtree of v is isomorphic to $t[s]$ and the edge connecting o to v has mark x towards o and $t[m]$ towards v. See Figure 3.2 for an example. For $s \in \bar{\mathcal{T}}_*$ and $x \in \Xi$, let $x \times s$ be $t \in \Xi \times \bar{\mathcal{T}}_*$ where $t[m] = x$ and $t[s] = s$. For two rooted trees $s, s' \in \bar{\mathcal{T}}_*$ which have the same vertex mark at the root, define $s \odot s'$ to be the element in $\bar{\mathcal{T}}_*$ obtained by joining s and s' at a common root, see Figure 3.3 for an example. Note that $\odot$ is commutative and associative. Therefore, we may write $\bigodot_{i=1}^k s_k$ for a collection $s_i, 1 \leq i \leq k$, of elements in $\bar{\mathcal{T}}_*$, which all have the same mark at the root.

Let G be a locally finite marked graph on a finite or countable vertex set V. Let v and w be adjacent vertices in G such that $\deg_G(v) \geq 2$. For $h \geq 1$, if $(G,v)_h$ is a rooted tree, then it is easy to see that we have

$$G[w,v]_h = \xi_G(w,v) \times \left[\bigodot_{\substack{w' \sim_G v \\ w' \neq w}} \left((\tau_G(v), \xi_G(w',v)) \otimes G[v,w']_{h-1} \right) \right]. \tag{3.7}$$

Also, if v is a vertex in G with $\deg_G(v) \geq 1$, it is easy to see that if $(G,v)_h$ is a rooted tree, we have

$$[G,v]_h = \bigodot_{w \sim_G v} \left((\tau_G(v), \xi_G(w,v)) \otimes G[v,w]_{h-1} \right). \tag{3.8}$$

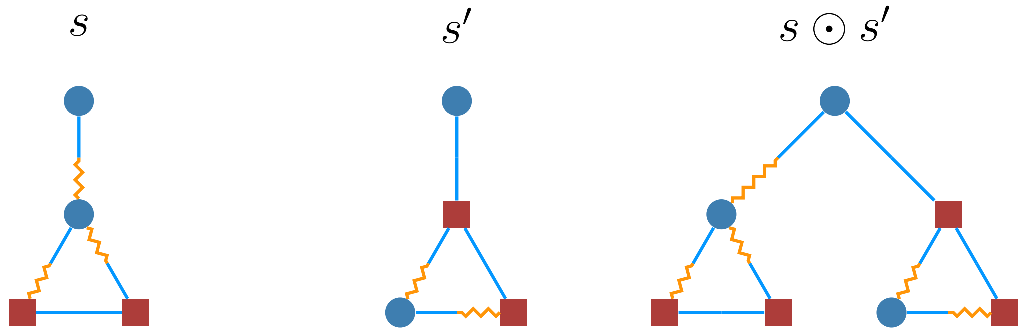

Figure 3.3: $s \odot s'$ for two rooted marked trees $s, s' \in \bar{\mathcal{T}}_*$ that have the same vertex mark at the root.

With this, we are ready to state conditions under which the edge colors of a directed colored graph are related to the edge colors of the directed colored graph derived from its marked colorblind version. The following proposition can be considered to be a generalization of Lemma 4.9 in [BC15].

Proposition 3.6. *Fix an integer $h \geq 1$. Let $\mathcal{F} \subset \Xi \times \bar{\mathcal{T}}_*^{h-1}$ be a finite set with cardinality L and set $\mathcal{C} = \mathcal{F} \times \mathcal{F}$. Let $H \in \mathcal{G}(\mathcal{C})$ be a simple directed colored graph on a finite or countable vertex set V, and let $\vec{\beta} = (\beta(v) : v \in V)$ have elements in Θ. Define A_h to be the set of vertices $v \in V$ such that the h–neighborhood of v in $\mathsf{CB}(H)$ is a rooted tree and also, for all vertices w with distance no more than h from v in $\mathsf{CB}(H)$, $(\beta(w), D^H(w))$ is graphical. Then, if $G = \mathsf{MCB}_{\vec{\beta}}(H)$, it holds that*

1. *For each vertex $v \in A_h$, we have $(G, v)_h \equiv [T_v, o_v]_h$ where $[T_v, o_v]$ is the rooted tree corresponding to the graphical pair $(\beta(v), D^H(v))$.*

2. *If $v, w \in A_h$ are adjacent vertices in H and the edge directed from v towards w has color (t, t') in H, we have $\varphi_G^h(v, w) = (t, t')$, i.e. $G(w, v)_{h-1} \equiv t$ and $G(v, w)_{h-1} \equiv t'$.*

Proof. For adjacent vertices u and v in H (which are, by definition, also adjacent in G), let $c(u, v) \in \mathcal{F}$ be the first component of the color of the edge directed from u towards v. Note that H is simple, meaning that there is only one edge directed from u towards v, so $c(u, v)$ is well-defined. Also, recall from the definition of $\mathcal{G}(\mathcal{C})$ that the color of the edge directed from v towards u is $\bar{c}$, with c being the color of the edge directed from u towards v. Therefore, the color of the edge directed from u towards v is $(c(u, v), c(v, u))$. Define A_0 to be the set of vertices $v \in V$ such that $(\beta(v), D^H(v))$ is graphical. Moreover, for $1 \leq l \leq h$, define A_l to be the set of vertices $v \in V$ such that $(G, v)_l$ is a rooted tree and, for all $w \in V$ with distance at most l from v in G, $(\beta(w), D^H(w))$ is graphical. Note that we have $A_0 \supseteq A_1 \supseteq \cdots \supseteq A_h$. On the other hand, note that removing the marks in G yields $\mathsf{CB}(H)$, hence A_h defined above coincides with that in the statement of Proposition 3.6. For each $v \in A_0$, let $[T_v, o_v] \in \bar{\mathcal{T}}_*^h$

be the rooted tree corresponding to the graphical pair $(\beta(v), D^H(v))$, and let (T_v, o_v) be an arbitrary member of the isomorphism class $[T_v, o_v]$. Observe that, for each vertex $v \in A_0$ with $\deg_G(v) \geq 1$, there exists a bijection f_v that maps the set of vertices adjacent to v in G to the set of vertices adjacent to o_v in T_v such that for all $w \sim_G v$, we have

$$\begin{aligned} c(v,w) &= T_v[f_v(w), o_v]_{h-1}, \\ c(w,v) &= T_v[o_v, f_v(w)]_{h-1}. \end{aligned} \tag{3.9}$$

This is because applying to $(\beta(v), D^H(v))$ the definition of what it means to be a graphical pair implies that, for each $t, t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$, we have that $E_h(t,t')(T_v, o_v)$, which is the number of vertices $\tilde{w} \sim_{T_v} o_v$ such that $T_v(\tilde{w}, o_v) \equiv t$ and $T_v(o_v, \tilde{w}) \equiv t'$, is equal to the number of vertices $w \sim_G v$ such that $c(v,w) = t$ and $c(w,v) = t'$.

Now, for each pair of adjacent vertices (v,w) in G, and $0 \leq r \leq h-1$, we inductively define $M_r(v,w) \in \Xi \times \bar{\mathcal{T}}_*^r$ as follows, We first define $M_0(v,w) \in \Xi \times \bar{\mathcal{T}}_*^0$ to have its mark component equal to $\xi_G(w,v) = c(v,w)[m]$ and its subtree component a single vertex with mark $\beta(v)$. In fact, $M_0(v,w) = G[w,v]_0$. For $v \sim_G w$ and $1 \leq r \leq h-1$, if $\deg_G(v) = 1$, i.e. w is the only vertex adjacent to v, we define $M_r(v,w)$ to be equal to $M_0(v,w)$. Otherwise, we define

$$M_r(v,w) := c(v,w)[m] \times \left[\bigodot_{\substack{w' \sim_G v \\ w' \neq w}} (\beta(v), c(v,w')[m]) \otimes M_{r-1}(w',v) \right]. \tag{3.10}$$

See Remark 3.1 below for a message passing interpretation for $M_r(v,w)$ motivated by (3.7). By induction on r, we show the following

$$v \in A_r, w \sim_G v \quad \Rightarrow \quad M_r(v,w) = c(v,w)_r, \qquad \forall 0 \leq r \leq h-1, \tag{3.11a}$$

$$v \in A_r, w \sim_G v \quad \Rightarrow \quad M_r(v,w) = G[w,v]_r, \qquad \forall 0 \leq r \leq h-1. \tag{3.11b}$$

Recall that $c(v,w)_r = (x, t_r)$, where x and t are the mark and the subgraph components of $c(v,w) \in \mathcal{F}$, respectively. Then, we use (3.11a) and (3.11b) to show that

$$v \in A_r \quad \Rightarrow \quad (G,v)_r \equiv (T_v, o_v)_r, \qquad \forall 0 \leq r \leq h. \tag{3.12}$$

Combining (3.11a) and (3.11b), we realize that for adjacent vertices $v, w \in A_h$, we have $G(v,w)_{h-1} \equiv c(w,v)_{h-1} = c(w,v)$ and $G(w,v)_{h-1} \equiv c(v,w)_{h-1} = c(v,w)$, which is the second part of the statement in Proposition 3.6. The first part is a result of (3.12) for $r = h$. Therefore, it suffices to show (3.11a), (3.11b) and (3.12) to complete the proof.

To start the proof, note that, for $r = 0$, $v \in A_0$, and $w \sim_G v$, the mark component of $M_0(v,w)$ is $\xi_G(w,v)$ and its subtree component is a single root with mark $\beta(v)$. On the other hand, the mark component of $c(v,w)_0$ is $c(v,w)[m] = \xi_G(w,v)$ and its subtree component, using (3.9), is the subtree component of $T_v[f_v(w), o_v]_0$. But since the pair $(\beta(v), D^H(v))$ is graphical, $T_v[f_v(w), o_v]_0$ is a single root with mark $\beta(v)$. This establishes (3.11a) for $r = 0$. Moreover, (3.11b) follows from the facts that, by the definition of $G = \mathsf{MCB}_{\vec{\beta}}(H)$, the mark component of $G[w, v]_0$ is $\xi_G(w, v) = c(v, w)[m]$ and its subtree component is a single root with mark $\beta(v)$.

Now, we use induction to show (3.11a) and (3.11b). First, we directly show (3.11a) and (3.11b) for a vertex v with $\deg_G(v) = 1$. If w is the only vertex adjacent to such v, we have $M_r(v, w) = M_0(v, w) \in \Xi \times \bar{\mathcal{T}}_*^0$, by definition. Recall that the mark component of $M_0(v, w)$ is $\xi_G(w, v) = c(v, w)[m]$ and its subtree component is a single root with mark $\beta(v)$. But this is precisely $G[w, v]_0$, which shows (3.11b). To show (3.11a), from (3.9), we have $c(v, w) = T_v[f_v(w), o_v]_{h-1}$. But since f_v is a bijection, and the pair $(\beta(v), D^H(v))$ is graphical, we have $\deg_{T_v}(o_v) = 1$ and hence the subtree component of $T_v[f_v(w), o_v]_{h-1}$ is a single root with mark $\beta(v)$, which is precisely the subtree component of $M_0(v, w)$. The mark components of $M_r(v, w) = M_0(v, w)$ and $c(v, w)$ are both equal to $\xi_G(w, v)$. This establishes (3.11a) in case $\deg_G(v) = 1$.

Now, we show (3.11a) and (3.11b) for $v \in A_r$ such that $\deg_G(v) \geq 2$. If $v \in A_r$ then all the vertices adjacent to v are in A_{r-1}. Therefore, using the induction hypothesis (3.11a) for $r - 1$ on the right hand side of (3.10), we realize that for such v and $w \sim_G v$ we have

$$M_r(v, w) = c(v, w)[m] \times \left[\bigodot_{\substack{w' \sim_G v \\ w' \neq w}} (\beta(v), c(v, w')[m]) \otimes c(w', v)_{r-1} \right].$$

Using (3.9) and the fact that $\beta(v) = \tau_{T_v}(o_v)$, we get

$$M_r(v, w) = \xi_{T_v}(f_v(w), o_v) \times \left[\bigodot_{\substack{w' \sim_G v \\ w' \neq w}} (\tau_{T_v}(o_v), \xi_{T_v}(f_v(w'), o_v)) \otimes T_v[o_v, f_v(w')]_{r-1} \right].$$

Observe that f_v is a bijection, hence the set of vertices w' in G such with $w' \sim_G v$ and $w' \neq w$ is mapped by f_v to the set of vertices $\tilde{w}$ in T_v such that $\tilde{w} \sim_{T_v} o_v$ and $\tilde{w} \neq f_v(w)$. With this, we can rewrite the above relation as

$$M_r(v, w) = \xi_{T_v}(f_v(w), o_v) \times \left[\bigodot_{\substack{\tilde{w} \sim_{T_v} v \\ \tilde{w} \neq f_v(w)}} (\tau_{T_v}(o_v), \xi_{T_v}(\tilde{w}, o_v)) \otimes T_v[o_v, \tilde{w}]_{r-1} \right].$$

Using (3.7), since (T_v, o_v) is a rooted tree, the right hand side is precisely $T_v[f_v(w), v]_r$. Another usage of (3.9) implies (3.11a).

To show (3.11b) for $v \in A_r$ with $\deg_G(v) \geq 2$ and $w \sim_G v$, again using the fact that $w' \in A_{r-1}$ for all $w' \sim_G v$, we realize that by first using (3.11b) for $r - 1$ and substituting in the right hand side of (3.10), then using $c(v,w')[m] = \xi_G(w',v)$ for all $w' \sim_G v$, and finally using $\beta(v) = \tau_G(v)$, we get

$$M_r(v,w) = \xi_G(w,v) \times \left[\bigodot_{\substack{w' \sim_G v \\ w' \neq w}} (\tau_G(v), \xi_G(w',v)) \otimes G[v,w']_{r-1} \right].$$

Since $v \in A_r$, $(G,v)_r$ is a rooted tree. Thereby, (3.7) implies that the right hand side of the preceding equation is precisely $G[w,v]_r$ which completes the proof of (3.11b).

Now, it remains to show (3.12). We first do this for $v \in A_r$ such that $\deg_G(v) \geq 1$. Observe that, since $v \in A_r$, $(G,v)_r$ is a rooted tree. Consequently, using (3.8), we have

$$[G,v]_r = \bigodot_{w \sim_G v} (\tau_G(v), \xi_G(w,v)) \otimes G[v,w]_{r-1}.$$

Since $w \in A_{r-1}$ for all $w \sim_G v$, using (3.11a) and (3.11b) for $r-1$, we realize that, for each $w \sim_G v$ on the right hand side, we have $G[v,w]_{r-1} = c(w,v)_{r-1}$. Moreover, we have $\tau_G(v) = \beta(v) = \tau_{T_v}(o_v)$ for $w \sim_G v$. Furthermore, by (3.9), $\xi_G(w,v) = c(v,w)[m] = \xi_{T_v}(f_v(w), o_v)$ for all $w \sim_G v$. Substituting these into the above relation and using (3.9), we get

$$[G,v]_r = \bigodot_{w \sim_G v} (\tau_{T_v}(o_v), \xi_{T_v}(f_v(w), o_v)) \otimes T_v[o_v, f_v(w)]_{r-1}.$$

Since f_v induces a one to one correspondence between the neighbors w of v in G and the neighbors $\tilde{w}$ of o_v in T_v, we may rewrite the above as

$$[G,v]_r = \bigodot_{\tilde{w} \sim_{T_v} o_v} (\tau_{T_v}(o_v), \xi_{T_v}(\tilde{w}, o_v)) \otimes T_v[o_v, \tilde{w}]_{r-1}.$$

Since (T_v, o_v) is a rooted tree, (3.8) implies that the right hand side is precisely $[T_v, o_v]_r$. This means that $[G,v]_r = [T_v, o_v]_r$ or equivalently $(G,v)_r \equiv (T_v, o_v)_r$, which is precisely (3.12).

To show (3.12) for $v \in A_r$ such that $\deg_G(v) = 0$, note that, for such v, $(G,v)_r$ is a single root with mark $\beta(v)$. Moreover, since $\deg_G(v) = \sum_{t,t' \in \mathcal{F}} D^H_{t,t'}(v)$, we must have $D^H_{t,t'}(v) = 0$ for all $t,t' \in \mathcal{F}$. Therefore, it must be the case that, for all $t,t' \in \Xi \times \bar{\mathcal{T}}^*_{h-1}$, $E_h(t,t')(T_v, o_v) = 0$. This means that $\deg_{T_v}(o_v) = 0$, and hence (T_v, o_v) is a single root with mark $\beta(v)$. Therefore, $(G,v)_r \equiv (T_v, o_v)_r$ and the proof is complete. □

Remark 3.1. *Motivated by the definition of $M_r(v,w)$ in (3.10), we can interpret $M_r(v,w)$ as the message the vertex v sends to the vertex w at time r, which is obtained by aggregating the messages sent by the neighbors of v, except for w, at time $r-1$. The proof of Proposition 3.6 above implies that, if $v \in A_r$, $M_r(v,w)$ is in fact the local r–neighborhood of v in G after removing the edge between v and w, i.e. $G[w,v]_r$. In fact, motivated by (3.7), the message $M_r(v,w)$ is inductively constructed in a way so that this holds.*

In the second part of Section 3.3.4, we started with a directed colored graph $H \in \mathcal{G}(\mathcal{C})$, defined on a finite or countable vertex set V, and a sequence $\vec{\beta} = (\beta(v) : v \in V)$ with elements in Θ, and studied the corresponding marked color blind version, denoted by $\mathsf{MCB}_{\vec{\beta}}(H)$. We will now start with a marked graph, consider the associated directed colored graph, for a given $h \geq 1$, and study the configuration model given by the colored degree sequence of this graph. The purpose is to relate the marked color blind versions of the directed colored graphs arising as realizations from this configuration model to the original marked graph we started with. The results we prove next are corollaries of Proposition 3.6.

Definition 3.5. *A marked or unmarked graph G is said to be h tree–like if, for all vertices v in G, the depth h local neighborhood of v in G, i.e. $(G, v)_h$, is a rooted tree. This condition is equivalent to requiring that there is no cycle of length $2h + 1$ or less in G.*

Corollary 3.2. *Let $n \in \mathbb{N}$. Recall that $\bar{\mathcal{G}}_n$ denotes the set of marked graphs on the vertex set $[n]$. For $h \geq 1$, assume that a marked h tree–like graph $G \in \bar{\mathcal{G}}_n$ is given. Let $\vec{D} = \vec{D}^{\mathsf{C}(G)}$ be the colored degree sequence associated to the directed colored version, $\mathsf{C}(G)$, of G. Let $\vec{\beta} := (\beta(v) : 1 \leq v \leq n)$ denote the vertex mark vector of G. Then, for any directed colored graph $H \in \mathcal{G}(\vec{D}, 2h + 1)$, we have $(\mathsf{MCB}_{\vec{\beta}}(H), v)_h \equiv (G, v)_h$ for all $v \in [n]$.*

Proof. By definition, since $H \in \mathcal{G}(\vec{D}, 2h+1) \subset \widehat{\mathcal{G}}(\vec{D})$, we have $D^H(v) = D(v) = D^{\mathsf{C}(G)}(v)$ for every vertex $v \in [n]$. Moreover, since $(G, v)_h$ is a rooted tree, using the rooted tree $(G, v)_h$ in Definition 3.4, we realize that the pair $(\beta(v), D^{\mathsf{C}(G)}(v))$ is graphical. On the other hand, since $H \in \mathcal{G}(\vec{D}, 2h+1)$, the colorblind graph $\mathsf{CB}(H)$ is h tree–like. Consequently, the set A_h in Proposition 3.6 coincides with $[n]$. Thus, the first part of Proposition 3.6 implies that for all $v \in [n]$, we have $(\mathsf{MCB}_{\vec{\beta}}(H), v)_h \equiv (G, v)_h$ which completes the proof. □

Corollary 3.3. *Let $n \in \mathbb{N}$ and $h \geq 1$. Let $G \in \bar{\mathcal{G}}_n$ be an h tree–like graph. Define*

$$N_h(G) := |\{G' \in \bar{\mathcal{G}}_n : U(G')_h = U(G)_h\}|. \tag{3.13}$$

Then, we have

$$N_h(G) = n(\vec{D}, \vec{\beta})|\mathcal{G}(\vec{D}, 2h + 1)|,$$

where $\vec{D} := \vec{D}^{\mathsf{C}(G)}$ and $\vec{\beta} = (\beta(i) : 1 \leq i \leq n)$ with $\beta(i) := \tau_G(i)$. Here $n(\vec{D}, \vec{\beta})$ denotes the number of distinct pairs $(\vec{D}^\pi, \vec{\beta}^\pi)$ where π ranges over the set of permutations $\pi : [n] \to [n]$ and where, for $1 \leq i \leq n$, $D^\pi(i) := D(\pi(i))$ and $\beta^\pi(i) := \beta(\pi(i))$.

Proof. For a permutation $\pi : [n] \to [n]$, define $G^\pi \in \bar{\mathcal{G}}_n$ to be the marked graph obtained from G by relabeling vertices using π. More precisely, for $v \in [n]$, we have $\tau_{G^\pi}(v) := \tau_G(\pi(v))$. Also, we place an edge between the vertices v and w in G^π if $\pi(v)$ and $\pi(w)$ are adjacent in G. In this case, we set $\xi_{G^\pi}(v, w) = \xi_G(\pi(v), \pi(w))$. With this, for any permutation $\pi : [n] \to [n]$ and $H \in \mathcal{G}(\vec{D}^\pi, 2h + 1)$, Corollary 3.2 implies that $U(\mathsf{MCB}_{\vec{\beta}^\pi}(H))_h = U(G^\pi)_h = U(G)_h$. On the other hand, if the permutations π and π' are such that $(\vec{D}^\pi, \vec{\beta}^\pi)$ and $(\vec{D}^{\pi'}, \vec{\beta}^{\pi'})$ are distinct, the sets $\{\mathsf{MCB}_{\vec{\beta}^\pi}(H) : H \in \mathcal{G}(\vec{D}^\pi, 2h+1)\}$ and $\{\mathsf{MCB}_{\vec{\beta}^{\pi'}}(H') : H' \in \mathcal{G}(\vec{D}^{\pi'}, 2h+1)\}$ are disjoint. Moreover, part 2 of Proposition 3.6 implies that for any permutation π and any $H \in \mathcal{G}(\vec{D}^\pi, 2h+1)$, $\mathsf{C}(\mathsf{MCB}_{\vec{\beta}^\pi}(H)) = H$. Thereby, distinct elements $H_1, H_2 \in \mathcal{G}(\vec{D}^\pi, 2h+1)$ yield distinct marked colorblind graphs $\mathsf{MCB}_{\vec{\beta}^\pi}(H_1)$ and $\mathsf{MCB}_{\vec{\beta}^\pi}(H_2)$. This establishes the inequality $N_h(G) \geq n(\vec{D}, \vec{\beta})|\mathcal{G}(\vec{D}, 2h+1)|$. The other direction can be seen by observing that if $G' \in \bar{\mathcal{G}}_n$ is such that $U(G')_h = U(G)_h$, then there exists a permutation $\pi : [n] \to [n]$ such that, for each vertex $v \in [n]$, we have $(G', v)_h \equiv (G, \pi(v))_h$. Consequently, for all vertices $v \in [n]$, we have $D^{\mathsf{C}(G')}(v) = D^{\mathsf{C}(G)}(\pi(v)) = D^\pi(v)$ and $\tau_{G'}(v) = \tau_G(\pi(v)) = \beta^\pi(v)$. Also, since G is h tree–like, so is G'. Hence, with $H := \mathsf{C}(G')$, we have $H \in \mathcal{G}(\vec{D}^\pi, 2h+1)$ and $G' = \mathsf{MCB}_{\vec{\beta}^\pi}(H)$. This shows that $N_h(G) \leq n(\vec{D}, \vec{\beta})|\mathcal{G}(\vec{D}, 2h+1)|$ and completes the proof. □

3.3.6 Realizing Admissible Probability Distributions with Finite Support

Next, using the tools developed above, we show that, for all $h \geq 1$ and any admissible probability distribution $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ having finite support, there exists a sequence of marked graphs which converges to P in the sense of local weak convergence. This result can be considered a generalization of Lemma 4.11 in [BC15].

Lemma 3.2. *Let $h \geq 1$ and $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$. Assume that P is admissible and has finite support. For $x, x' \in \Xi$, let $d_{x,x'} := \mathbb{E}_P\left[\deg_T^{x,x'}(o)\right]$ and $\vec{d} := (d_{x,x'} : x, x' \in \Xi)$. Moreover, for $\theta \in \Theta$, let q_θ be the probability of the mark at the root in P being θ and define $Q = (q_\theta : \theta \in \Theta)$. If $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are sequences of edge and vertex mark count vectors such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{d}, Q)$, then there exists a finite set $\Delta \subset \bar{\mathcal{T}}_*^h$ and a sequence of marked graphs $G_n \in \mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}$ such that the support of $U(G_n)_h$ is contained in Δ for each n, and $U(G_n)_h \Rightarrow P$.*

Proof. Let $S = \{r_1, \dots, r_k\} \subset \bar{\mathcal{T}}_*^h$ be the finite support of P. Since S is finite, we can construct, for each $n \in \mathbb{N}$, a sequence $(g^{(n)}(i) : 1 \leq i \leq n)$ where $g^{(n)}(i) \in S$, $1 \leq i \leq n$, and

$$\frac{1}{n} \sum_{i=1}^n \delta_{g^{(n)}(i)} \Rightarrow P. \tag{3.14}$$

Let δ be the maximum degree at the root over all the elements of S. Moreover, let $\mathcal{F} \subset \Xi \times \bar{\mathcal{T}}_*^{h-1}$ be the set comprised of $T[o, v]_{h-1}$ and $T[v, o]_{h-1}$ for each $[T, o] \in S$ and $v \sim_T o$. Since S is finite, $\mathcal{F}$ is finite, hence can be identified with $\{1, \dots, L\}$ with $L := |\mathcal{F}|$. With this, define the color set $\mathcal{C} := \mathcal{F} \times \mathcal{F}$.

For each $n \in \mathbb{N}$, define the sequences $\vec{\beta}^{(n)} = (\beta^{(n)}(i) : 1 \leq i \leq n)$ and $\vec{D}^{(n)} = (D^{(n)}(i) : 1 \leq i \leq n)$ as follows. For $1 \leq i \leq n$, let $\beta^{(n)}(i) \in \Theta$ be the mark at the root in $g^{(n)}(i)$. Further, let $D^{(n)}(i) \in \mathcal{M}_L^{(\delta)}$ be such that, for $c \in \mathcal{C}$, $D_c^{(n)}(i) = E_h(c)(g^{(n)}(i))$. Here, $E_h(c)(g^{(n)}(i)) = E_h(t, t')(g^{(n)}(i))$ with $c = (t, t')$, as was defined in (2.6).

Now, we try to construct directed colored graphs given $\vec{D}^{(n)}$. However, it might be the case that $\vec{D}^{(n)} \notin \mathcal{D}_n$. Therefore, we modify $\vec{D}^{(n)}$ slightly to get a sequence in $\mathcal{D}_n$. In order to do this, for $c \in \mathcal{C}$, let $S_c^{(n)} := \sum_{i=1}^n D_c^{(n)}(i)$. Moreover, for $c \in \mathcal{C}_=$, let $\widetilde{S}_c^{(n)} := 2\lfloor S_c^{(n)}/2 \rfloor$, and for $c \in \mathcal{C}_{\neq}$, let $\widetilde{S}_c^{(n)} := S_c^{(n)} \wedge S_{\bar{c}}^{(n)}$. Note that, because of (3.14), for all $c \in \mathcal{C}$, $S_c^{(n)}/n \to e_P(c)$ as $n \to \infty$. On the other hand, as P is admissible, we have $e_P(c) = e_P(\bar{c})$. Hence, $|\widetilde{S}_c^{(n)} - S_c^{(n)}| = o(n)$ for all $c \in \mathcal{C}$. Therefore, we can find a sequence $\vec{D}'^{(n)} = (D'^{(n)}(i) : 1 \leq i \leq n)$ such that for all $1 \leq i \leq n$ we have $D'^{(n)}(i) \in \mathcal{M}_L^{(\delta)}$, and for all $c \in \mathcal{C}$ we have $D_c'^{(n)}(i) \leq D_c^{(n)}(i)$, and we have $\sum_{i=1}^n D_c'^{(n)}(i) = \widetilde{S}_c^{(n)}$. Moreover, since $\sum_{c\in\mathcal{C}} |\widetilde{S}_c^{(n)} - S_c^{(n)}| = o(n)$, we may construct $\vec{D}'^{(n)}$ such that, for all but $o(n)$ vertices, we have $D'^{(n)}(i) = D^{(n)}(i)$. In particular, if $\widetilde{P} \in \mathcal{P}(\mathcal{M}_L^{(\delta)})$ is defined to be the law of $D = (D_c : c \in \mathcal{C})$, where $D_c = E_h(c)(r)$ with r having law P, we have

$$\frac{1}{n}\sum_{i=1}^n \delta_{D'^{(n)}(i)} \Rightarrow \widetilde{P}. \tag{3.15}$$

Indeed, due to (3.14), we have $(\sum_{i=1}^n \delta_{D^{(n)}(i)})/n \Rightarrow \tilde{P}$, which implies (3.15) since $D^{(n)}(i) = D'^{(n)}(i)$ for all but $o(n)$ many $1 \leq i \leq n$. Note that, by definition, $\widetilde{S}_c^{(n)}$ is even for $c \in \mathcal{C}_=$ and, for $c \in \mathcal{C}_{\neq}$, $S_c^{(n)} = S_{\bar{c}}^{(n)}$. Therefore, $\vec{D}'^{(n)} \in \mathcal{D}_n$.

Furthermore, since conditions (3.5a) and (3.5b) are both satisfied for $\vec{D}'^{(n)}$ and $\widetilde{P}$, Theorem 3.5 then implies that $\mathcal{G}(\vec{D}'^{(n)}, 2h+1)$ is non empty for n large enough. For such n, let $H^{(n)}$ be a member of $\mathcal{G}(\vec{D}'^{(n)}, 2h+1)$ and let $\widetilde{G}^{(n)} = \mathsf{MCB}_{\vec{\beta}^{(n)}}(H^{(n)})$. Since for each $1 \leq i \leq n$, $\beta^{(n)}(i)$ and $D^{(n)}(i)$ are defined based on $g^{(n)}(i) \in \bar{\mathcal{T}}_*^h$, they form a graphical pair in the sense of Definition 3.4. Also, $D'^{(n)}(i) = D^{(n)}(i)$ for all but $o(n)$ vertices. On the other hand, all the degrees in $\widetilde{G}^{(n)}$ are bounded by δ. Therefore, the number of vertices v in $\widetilde{G}^{(n)}$ such that $(\beta^{(n)}(w), D'^{(n)}(w))$ is graphical for all vertices w in the h–neighborhood of v is $n - o(n)$. Moreover, since $H^{(n)} \in \mathcal{G}(\vec{D}'^{(n)}, 2h+1)$, $\widetilde{G}^{(n)}$ has no cycle of length $2h+1$ or less, which means that $\widetilde{G}^{(n)}$ is h tree–like. Thereby, Proposition 3.6 implies that the number of vertices v in $\widetilde{G}^{(n)}$ such that $(\widetilde{G}^{(n)}, v)_h \equiv g^{(n)}(v)$ is $n - o(n)$. This means that $U(\widetilde{G}^{(n)})_h \Rightarrow P$.

Now, the only remaining step is to modify $\widetilde{G}^{(n)}$ to obtain a simple marked graph in $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}$. To do this, note that if $(\widetilde{m}^{(n)}(x,x') : x, x' \in \Xi)$ is the edge mark count vector of $\widetilde{G}^{(n)}$, we have

$$\widetilde{m}^{(n)}(x,x') = \begin{cases} \sum_{v=1}^n \vec{D}'^{(n)}_{x,x'}(v) & x \neq x', \\ \frac{1}{2}\sum_{v=1}^n \vec{D}'^{(n)}_{x,x}(v) & x = x', \end{cases}$$

where

$$D'^{(n)}_{x,x'}(v) := \sum_{\substack{t,t' \in \mathcal{F} \\ t[m]=x, t'[m]=x'}} D'^{(n)}_{t,t'}(v).$$

This together with condition (3.15), implies that for $x \neq x' \in \Xi$ we have $\widetilde{m}^{(n)}(x,x')/n \to d_{x,x'}$, and for $x \in \Xi$ we have $\widetilde{m}^{(n)}(x,x)/n \to d_{x,x}/2$. Consequently, $|\widetilde{m}^{(n)}(x,x') - m^{(n)}(x,x')| = o(n)$. On the other hand, if $(\widetilde{u}^{(n)}(\theta) : \theta \in \Theta)$ is the vertex mark count vector of $\widetilde{G}^{(n)}$, since

$\vec{\beta}^{(n)}$ is the vertex mark vector of $\widetilde{G}^{(n)}$, (3.14) implies that for $\theta \in \Theta$, $\widetilde{u}^{(n)}(\theta)/n \to q_\theta$ and hence $\sum_{\theta \in \Theta} |\widetilde{u}^{(n)}(\theta) - u^{(n)}(\theta)| = o(n)$. Now, we modify $\widetilde{G}^{(n)}$ to obtain $G^{(n)}$. In order to do this, for each $x \le x' \in \Xi$ such that $\widetilde{m}^{(n)}(x,x') < m^{(n)}(x,x')$, we add $m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')$ many edges with mark x, x'. We can do this for all such x, x' so that all vertices in the graph are connected to at most one of the newly added edges. This is possible for n large enough since Ξ is finite, $\sum_{x \le x' \in \Xi} |m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')| = o(n)$, and the total number of edges in $\widetilde{G}^{(n)}$ is $O(n)$. Next, for $x \le x' \in \Xi$ such that $\widetilde{m}^{(n)}(x,x') > m^{(n)}(x,x')$, we arbitrarily remove $\widetilde{m}^{(n)}(x,x') - m^{(n)}(x,x')$ many edges with mark x, x'. Moreover, since $\sum_{\theta \in \Theta} |u^{(n)}(\theta) - \widetilde{u}^{(n)}(\theta)| = o(n)$, we may change the vertex mark of all but $o(n)$ many vertices so that for all $\theta \in \Theta$, the number of vertices with mark θ becomes precisely equal to $u^{(n)}(\theta)$. Let $G^{(n)}$ be the resulting simple marked graph, which is indeed a member of $\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}$.

Note that, by construction, all the degrees in $G^{(n)}$ are bounded by $\delta + 1$. Hence, the support of $U(G^{(n)})_h$ is contained in the set Δ, defined as the set of $[T, o] \in \bar{\mathcal{T}}_*^h$ such that the degrees of all vertices in T are bounded by $\delta + 1$. Note that Δ is finite. Also, adding or removing each edge affects the h–neighborhood of at most $2(\delta+1)^{h+1}$ many vertices. Likewise, changing the mark of a vertex can affect the h–neighborhood of at most $(\delta+1)^{h+1}$ many vertices. Hence, $(G^{(n)}, v)_h = (\widetilde{G}^{(n)}, v)_h$ for all but $o(n)$ vertices $v \in [n]$. Consequently. $U(G^{(n)})_h \Rightarrow P$ and the proof is complete. □

3.3.7 Local Weak Convergence of a Sequence of Graphs obtained from a Colored Configuration Model

In Lemma 3.2 in the previous section, given $h \ge 1$ and an admissible $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ with finite support, we constructed a sequence of marked graphs $G^{(n)}$ such that $U(G^{(n)})_h \Rightarrow P$. In this section, we show how to use a colored configuration model based on this sequence to generate marked graphs which converge to $\mathsf{UGWT}_h(P)$ in the local weak sense. In the process of doing this, we also draw a connection between the marked unimodular Galton–Watson trees introduced in Section 2.7 and the colored unimodular Galton–Watson trees introduced in Section 3.3.3.

Fix $h \ge 1$. Let $\Delta \subset \bar{\mathcal{T}}_*^h$ be a fixed finite set. Let $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ be admissible with support contained in Δ. We write $\mathcal{F}$ for the set of $T[o,v]_{h-1}$ and $T[v,o]_{h-1}$ arising from $[T,o] \in \Delta$ and vertices $v \sim_T o$. Since Δ is finite, $\mathcal{F}$ is also finite. We use the notation $L := |\mathcal{F}|$. Define the color set $\mathcal{C} := \mathcal{F} \times \mathcal{F}$. Let δ be an upper bound for the degree of each vertex of each $[T,o] \in \Delta$. For $r \in \Delta$, define $D(r) \in \mathcal{M}_L^{(\delta)}$ to be the matrix such that, for $t, t' \in \mathcal{F}$, $D_{t,t'}(r) = E_h(t,t')(r)$. Furthermore, define $\theta(r) \in \Theta$ to be the mark at the root in r.

Proposition 3.7. *With the above setup, let $(\Gamma_n : n \in \mathbb{N})$ be a sequence of marked graphs, with Γ_n having the vertex set $[n]$ and the support of $U(\Gamma_n)_h$ contained in Δ for each n, and such that $U(\Gamma_n)_h \Rightarrow P$. Define $\vec{D}^{(n)} = (D^{(n)}(v) : v \in [n])$ where for $v \in [n]$, $D^{(n)}(v) \in \mathcal{M}_L^{(\delta)}$ is defined such that for $t, t' \in \mathcal{F}$, $D^{(n)}_{t,t'}(v) := E_h(t,t')(\Gamma_n, v)$. Moreover, define $\vec{\beta}^{(n)} = (\beta^{(n)}(v) : v \in [n])$ such that $\beta^{(n)}(v) := \tau_{\Gamma_n}(v)$ for $v \in [n]$. For $n \ge 1$, let H_n be a random directed* *colored graph uniformly distributed in* $\mathcal{G}(\vec{D}^{(n)}, 2h+1)$*, and assume that* $(H_n : n \in \mathbb{N})$ *are independent on a joint probability space. Let* $G_n := \mathsf{MCB}_{\vec{\beta}^{(n)}}(H_n)$*. Then, with probability one, we have* $U(G_n) \Rightarrow \mathsf{UGWT}_h(P)$*.*

We prove this proposition in two steps. First, in Lemma 3.3 below, we draw a connection between $\mathsf{UGWT}_h(P)$ and a colored unimodular Galton–Watson tree. Then, we use this to state Lemma 3.4, which will then complete the proof of the above statement. Before this, we need to set up some notation.

Let $\widetilde{P} \in \mathcal{P}(\mathcal{M}_L^{(\delta)})$ be the law of $D(r)$ where $r \sim P$. Since P is admissible, we have $\mathbb{E}_{\widetilde{P}}[D_c] = \mathbb{E}_{\widetilde{P}}[D_{\bar{c}}]$ for all $c \in \mathcal{C}$. Now, we generate a random rooted directed colored tree (F, o) using the procedure described in Section 3.3.3 by starting with $D^{(0)} = D(r^{(0)})$ with $r^{(0)} \sim P$ at the root, and then adding further layers as in the colored unimodular Galton–Watson tree. Let $Q \in \mathcal{P}(\Theta \times \mathcal{G}_*(\mathcal{C}))$ be the law of the pair $(\theta(r^{(0)}), [F, o])$. Furthermore, let Q_1 and Q_2 be the law of $\theta(r)$ and $[F, o]$, respectively. Note that $Q_2 = \mathsf{CUGWT}(\widetilde{P})$. For vertices v, w in F, let $c(v, w) \in \mathcal{F}$ be the first component of the color of the edge going from v towards w. For a vertex v in F other than the root, let $p(v)$ be the parent of v, and let $c(v)$ be the shorthand for $(c(v, p(v)), c(p(v), v))$. Moreover, for a vertex v, let $M(v) \in \mathcal{M}_L^{(\delta)}$ be such that for $c \in \mathcal{C}$,

$$M_c(v) := |\{w : p(w) = v, c(v, w) = c\}|.$$

In fact, $M(v)$ is the part of the colored degree matrix of v corresponding to its offspring, so that if $v \neq o$, $D^F(v) = M(v) + E^{c(v)}$ and $D^F(o) = M(o)$. Recall that $E^{c(v)} \in \mathcal{M}_L$ is the matrix with value 1 in entry $c(v)$ and zero elsewhere.

A matrix $D \in \mathcal{M}_L^{(\delta)}$ is said to be Δ–graphical if there exists $r \in \Delta$ such that $D = D(r)$. If $D \in \mathcal{M}_L^{(\delta)}$ is Δ–graphical and nonzero, define $\theta(D)$ to be the mark at the root for some $r \in \bar{\mathcal{T}}_*^h$ for which we have $D = D(r)$. To see why $\theta(D)$ is well-defined for $D \neq 0$, take $r, r' \in \Delta$ so that $D = D(r) = D(r')$. Since D is nonzero, there exist $t, t' \in \Xi \times \mathcal{F}$ such that $D_{t,t'} = D_{t,t'}(r) = D_{t,t'}(r') > 0$. Hence, the marks at the root in both r and r' are the same as the mark at the root in the subgraph part of t, i.e. $t[s]$. This shows that $\theta(D)$ is well defined. In fact this together with Lemma A.4 in Appendix A.2 implies that if $D \neq 0$ is Δ–graphical there is only one $r \in \Delta$ such that $D = D(r)$.

We say that a rooted directed colored graph $[F, o] \in \mathcal{G}_*(\mathcal{C})$ is Δ–graphical if for each vertex v in F, $D^F(v)$ is Δ–graphical. Let $\mathcal{H}$ be the subset of $\Theta \times \mathcal{G}_*(\mathcal{C})$ which consists of the pairs $(\theta, [F, o])$ such that $[F, o]$ is a Δ–graphical rooted directed colored graph, and if o is not isolated in F we have $\theta = \theta(D^F(o))$. For $(\theta, [F, o]) \in \mathcal{H}$, by an abuse of notation, we define $\mathsf{MCB}_\theta(F)$ to be the simple marked graph defined as follows. Let $\beta(o) := \theta$, and for $v \neq o$ in F, define $\beta(v) := \theta(D^F(v))$. Note that if v is a vertex other than the root, since F is connected by definition, v is not isolated and hence $D^F(v)$ is not the zero matrix. Thereby, $\theta(D^F(v))$ is well-defined. With this, let $\vec{\beta}$ be the vector consisting of $\beta(v)$ for vertices v in F, and define $\mathsf{MCB}_\theta(F) := \mathsf{MCB}_{\vec{\beta}}(F)$. Note that if $(\theta, [F, o]) \sim Q$ then, with probability one, we have $(\theta, [F, o]) \in \mathcal{H}$. The reason is that $D^F(o) = D(r^{(0)})$ and $\theta = \theta(r^{(0)})$, where $r^{(0)}$ is in the support of P and hence in Δ. Moreover, by the construction of $\mathsf{CUGWT}(\widetilde{P})$ and (3.6), with probability one, for all vertices $v \neq o$ in F, $D^F(v) = M(v) + E^{c(v)}$ is in the support of $\widetilde{P}$, and hence $D^F(v)$ is Δ–graphical.

Now we are ready to state two lemmas. Lemma 3.4 will prove Proposition 3.7, and itself depends on Lemma 3.3. The proposition will be proved assuming the truth of the lemmas, and then the lemmas will be proved.

Lemma 3.3. *If $(\theta, [F, o])$ has the law Q described above, then $[\mathsf{MCB}_\theta(F), o]$ has the law $\mathsf{UGWT}_h(P)$.*

Lemma 3.4. *With the above setup, assume that a sequence $\vec{D}^{(n)} \in \mathcal{D}_n$ together with a sequence $\vec{\beta}^{(n)} = (\beta^{(n)}(v) : v \in [n])$ are given such that $\beta^{(n)}(v) \in \Theta$ for all $v \in V$. Moreover, assume that for each $n \in \mathbb{N}$ and $v \in [n]$, we have $D^{(n)}(v) \in \mathcal{M}_L^{(\delta)}$ and $(\beta^{(n)}(v), D^{(n)}(v)) = (\theta(r), D(r))$ for some $r \in \Delta$. Also, with $\widetilde{Q}$ being the law of $(\theta(r), D(r))$ when $r \sim P$, assume that*

$$\frac{1}{n}\sum_{v=1}^{n} \delta_{(\beta^{(n)}(v), D^{(n)}(v))} \Rightarrow \widetilde{Q}. \tag{3.16}$$

With H_n uniformly distributed in $\mathcal{G}(\vec{D}^{(n)}, 2h+1)$ and independently for each n, define $G_n := \mathsf{MCB}_{\vec{\beta}^{(n)}}(H_n)$. Then, with probability one, we have $U(G_n) \Rightarrow \mathsf{UGWT}_h(P)$.

Proof of Proposition 3.7. Note that since $U(\Gamma_n)_h \Rightarrow P$, the sequences $\vec{\beta}^{(n)}$ and $\vec{D}^{(n)}$ obtained from Γ_n as in the statement of the proposition satisfy (3.16). Therefore, Lemma 3.4 completes the proof. □

Proof of Lemma 3.3. Note that, with probability one, $(\theta, [F, o]) \in \mathcal{H}$. Let $T := \mathsf{MCB}_\theta(F)$. Since $[F, o]$ is almost surely Δ–graphical and T is a simple marked tree, Proposition 3.6 implies that, for all vertices v in T, we have $[T, v]_h \in \Delta$ almost surely. Therefore, given $r \in \Delta$, using Proposition 3.6, we have $(T, o)_h \equiv r$ iff $M(o) = D(r)$ and θ is the mark at the root in r, i.e. $(\theta, M(o)) = (\theta(r), D(r))$. By the definition of Q, this has probability $P(r)$. To sum up, we have $\mathbb{P}\left((T, o)_h \equiv r\right) = P(r)$.

Now, assume that $v \sim_T o$ is an offspring of the root in T such that $T(o, v)_{h-1} \equiv t$ and $T(v, o)_{h-1} \equiv t'$. If $\tilde{t} \in \Xi \times \bar{\mathcal{T}}_*^h$ is such that $\tilde{t}_{h-1} = t$, Lemma A.2 in Appendix A.2 implies that $T(o, v)_h \equiv \tilde{t}$ iff $(T, v)_h \equiv \tilde{t} \oplus t'$. Since $[T, v]_h \in \Delta$ almost surely, $T(o, v)_h \equiv \tilde{t}$ has probability zero unless $\tilde{t} \oplus t' \in \Delta$. Assuming that $\tilde{t} \oplus t' \in \Delta$ is satisfied, by the construction of $\mathsf{MCB}_\theta(F)$ and Proposition 3.6, we know that $(T, v)_h \equiv \tilde{t} \oplus t'$ iff $D^F(v) = D(\tilde{t} \oplus t')$, or equivalently, $M(v) = D(\tilde{t} \oplus t') - E^{(t,t')}$. From (3.6), the probability of this is precisely

$$\widehat{\widetilde{P}}^{\overline{c(v)}}(M(v)) = \frac{(M_{c(v)}(v) + 1)\widetilde{P}(M(v) + E^{c(v)})}{e_P(\overline{c(v)})}.$$

Since $c(v) = (t, t')$, we have $M_{c(v)}(v) + 1 = D_{(t,t')}(\tilde{t} \oplus t') = E_h(t, t')(\tilde{t} \oplus t')$. On the other hand, $\widetilde{P}(M(v) + E^{c(v)}) = \widetilde{P}(D(\tilde{t} \oplus t')) = P(\tilde{t} \oplus t')$. Comparing this with (2.7), we realize that

$$\mathbb{P}\left(T(o,v)_h \equiv \tilde{t} | T(o,v)_{h-1} \equiv t, T(v,o)_{h-1} \equiv t'\right) = \mathbb{1}\left[\tilde{t} \oplus t' \in \Delta\right] \widehat{P}_{t,t'}(\tilde{t}) = \widehat{P}_{t,t'}(\tilde{t}),$$

where the last equality uses the fact that the support of P is contained in Δ. Comparing these with the definition of $\mathsf{UGWT}_h(P)$, the proof is complete by repeating this argument inductively for further depths in T and noting that the choice of $M(v)$ in F is done conditionally independently for vertices with the same parent. □

Proof of Lemma 3.4. From Theorem 3.6 we know that, with probability one, we have $U(H_n) \Rightarrow Q_2 = \mathsf{CUGWT}(\widetilde{P})$. Moreover, we claim that, with probability one,

$$\frac{1}{n}\sum_{v=1}^{n} \delta_{(\beta^{(n)}(v),[H_n,v])} \Rightarrow Q. \tag{3.17}$$

Recall that $[H_n, v] \in \mathcal{G}_*(\mathcal{C})$ is the isomorphism class of the connected component of v in H_n rooted at v. Since $H_n \in \mathcal{G}(\vec{D}^{(n)}, 2h+1)$, for each $v \in [n]$, it holds that $[H_n(v), v] \in \mathcal{G}_*(\mathcal{C})$ is a simple colored directed rooted graph. Here, to make sense of the weak convergence, we turn $\Theta \times \mathcal{G}_*(\mathcal{C})$ into a metric space with the metric

$$d((\theta, [H,o]), (\theta', [H', o'])) = d_\Theta(\theta, \theta') + d_{\mathcal{G}_*(\mathcal{C})}([H,o],[H',o']),$$

where d_Θ in the first term on the right hand side is an arbitrary metric on the finite set Θ, e.g. the discrete metric, and $d_{\mathcal{G}_*(\mathcal{C})}$ denotes the the local metric of $\mathcal{G}_*(\mathcal{C})$ from Section 3.3.3. To show (3.17), we take a bounded continuous function $f : \Theta \times \mathcal{G}_*(\mathcal{C}) \to \mathbb{R}$ and show that

$$\frac{1}{n}\sum_{v=1}^{n} f(\beta^{(n)}(v), [H_n, v]) \to \int f dQ \qquad \text{a.s..} \tag{3.18}$$

With such a function f, define $f_1 : \mathcal{G}_*(\mathcal{C}) \to \mathbb{R}$ as follows: for $[F,o] \in \mathcal{G}_*(\mathcal{C})$, if o is not isolated in F and $D^F(o)$ is Δ–graphical, define $f_1([F,o]) := f(\theta(D^F(o)), [F,o])$. Recall that, since $D^F(o)$ is nonzero and Δ–graphical, $\theta(D^F(o))$ is well-defined. Otherwise, if o is isolated in F or if $D^F(o)$ is not Δ–graphical, define $f_1([F,o])$ to be zero. Moreover, define $f_2 : \Theta \times \mathcal{M}_L \to \mathbb{R}$ as follows: for $\theta \in \Theta$ and $D \in \mathcal{M}_L$, if D is not the zero matrix, define $f_2(\theta, D)$ to be zero. Otherwise, define $f_2(\theta, D) := f(\theta, [F,o])$ where $[F,o] \in \mathcal{G}_*(\mathcal{C})$ is an isolated root. Now, take $(\theta, [F,o]) \in \Theta \times \mathcal{G}_*(\mathcal{C})$ such that $(\theta, D^F(o)) = (\theta(r), D(r))$ for some $r \in \Delta$. If o is isolated in F, $f_1([F,o]) = 0$ and $f_2(\theta, D^F(o)) = f(\theta, [F,o])$. Otherwise, $f_1([F,o]) = f(\theta, [F,o])$ and $f_2(\theta, D^F(o)) = 0$. In both cases, we have

$$f(\theta, [F,o]) = f_1([F,o]) + f_2(\theta, D^F(o)). \tag{3.19}$$

On the other hand, if $(\theta, [F,o]) \sim Q$, with probability one, we have $(\theta, D^F(o)) = (\theta(r), D(r))$ for some $r \in \Delta$. Thereby,

$$f(\theta, [F,o]) = f_1([F,o]) + f_2(\theta, D^F(o)) \qquad Q\text{–a.s..} \tag{3.20}$$

Note that, by assumption, for all $n \in \mathbb{N}$ and $v \in [n]$, we have $(\beta^{(n)}(v), D^{(n)}(v)) = (\theta(r), D(r))$ for some $r \in \Delta$. Also, we have $D^{H_n}(v) = D^{(n)}(v)$. Consequently, from (3.19), for all $n \in \mathbb{N}$ and $v \in [n]$, we have

$$f(\beta^{(n)}(v), [H_n, v]) = f_1([H_n, v]) + f_2(\beta^{(n)}(v), D^{(n)}(v)) \qquad \text{a.s..} \tag{3.21}$$

Moreover, if $(\theta, [F, o]) \sim Q$, $[F, o]$ is distributed according to Q_2 and $(\theta, D^F(o))$ is distributed according to $\widetilde{Q}$. Thereby, using (3.20), we have

$$\int f dQ = \int f_1 dQ_2 + \int f_2 d\widetilde{Q}. \tag{3.22}$$

It is easy to see that if f is continuous, both f_1 and f_2 are continuous. Therefore, using the fact that, with probability one, $U(H_n) \Rightarrow Q_2$, we realize that,

$$\frac{1}{n}\sum_{v=1}^{n} f_1([H_n, v]) = \int f_1 dU(H_n) \to \int f_1 dQ_2 \qquad \text{a.s..} \tag{3.23}$$

Also, due to (3.16), we have

$$\frac{1}{n}\sum_{v=1}^{n} f_2(\beta^{(n)}(v), D^{(n)}(v)) \to \int f_2 d\widetilde{Q}. \tag{3.24}$$

Substituting (3.23) and (3.24) into (3.21) and comparing with (3.22), we arrive at (3.18) which shows (3.17).

Now, define the function J that maps $(\theta, [F, o]) \in \mathcal{H}$ to $[\mathsf{MCB}_\theta(F), o] \in \bar{\mathcal{G}}_*$. It is easy to see that J is continuous. Moreover, Lemma 3.3 asserts that the pushforward of Q under the mapping J is precisely $\mathsf{UGWT}_h(P)$. On the other hand, since for all $n \in \mathbb{N}$ and $v \in [n]$ we have $(\beta^{(n)}(v), D^{(n)}(v)) = (\theta(r), D(r))$ for some $r \in \Delta$, we realize that, with probability one, $(\beta^{(n)}(v), [H_n, v]) \in \mathcal{H}$ and $J(\beta^{(n)}(v), [H_n, v]) = [\mathsf{MCB}_{\vec{\beta}^{(n)}}(H_n), v] = [G_n, v]$. Consequently, the pushforward of the left hand side in (3.17) under the map J is precisely $U(G_n)$, while the pushforward of its right hand side under the map J is $\mathsf{UGWT}(P)$. This means that, with probability one, $U(G_n) \Rightarrow \mathsf{UGWT}(P)$ and this completes the proof. □

3.4 Properties of the Entropy

In this section, we give the proof of steps taken in Section 3.2 in order to prove Theorems 3.1, 3.2 and 3.3. First, in Section 3.4.1, we prove Propositions 3.1 and 3.2, which specify conditions under which the entropy is $-\infty$. Afterwards, in Section 3.4.2, we prove the lower bound result of Proposition 3.3. In Section 3.4.3, we prove the upper bound result of Propositions 3.4. Finally, in Section 3.4.4, we prove the upper bound result of Proposition 3.5.

3.4.1 Conditions under which the entropy is $-\infty$

In this section, we prove Propositions 3.1 and 3.2. Before that, we state and prove the following useful lemma:

Lemma 3.5. *If, for integers n and $0 \le m \le \binom{n}{2}$, $\mathcal{G}_{n,m}$ denotes the set of simple unmarked graphs on the vertex set $[n]$ having exactly m edges, we have,*

$$\log |\mathcal{G}_{n,m}| = \log \left| \binom{\binom{n}{2}}{m} \right| \le m \log n + ns\left(\frac{2m}{n}\right),$$

where $s(x) := \frac{x}{2} - \frac{x}{2}\log x$ for $x > 0$ and $s(0) := 0$. Moreover, since $s(x) \le 1/2$ for all $x \ge 0$, we have in particular

$$\log |\mathcal{G}_{n,m}| \le m \log n + \frac{n}{2}.$$

Proof. Using the classical upper bound $\binom{r}{s} \le (re/s)^s$, we have

$$\log \left| \binom{\binom{n}{2}}{m} \right| \le m \log \frac{n^2 e}{2m} = m \log n + m \log \frac{ne}{2m} = m \log n + ns(2m/n),$$

which completes the first part. Also, it is easy to see that $s(x)$ is increasing for $x \le 1$, decreasing for $x > 1$ and attains its maximum value $1/2$ at $x = 1$. Therefore, $s(x) \le 1/2$. This completes the proof of the second statement. □

Proof of Proposition 3.1. Suppose $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} > -\infty$. Then, for all $\epsilon > 0$, $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)$ is non empty for infinitely many n. Therefore, there exists a sequence of integers n_i going to infinity together with simple marked graphs $G^{(n_i)} \in \mathcal{G}^{(n_i)}_{\vec{m}^{(n_i)},\vec{u}^{(n_i)}}$ such that $U(G^{(n_i)}) \Rightarrow \mu$. This already implies that if $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} > -\infty$, μ must be sofic and hence unimodular. In other words, if μ is not unimodular, $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$.

Consequently, it remains to show that if either $\vec{d} \neq \vec{\deg}(\mu)$ or $Q \neq \vec{\Pi}(\mu)$ we have $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$. Similar to the above, assume $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} > -\infty$ and take the above sequence of simple marked graphs $G^{(n_i)}$. First note that, for any $\alpha > 0$, and $x, x' \in \Xi$, the function $[G,o] \mapsto \deg_G^{x,x'}(o) \wedge \alpha$ is continuous and bounded on $\bar{\mathcal{G}}_*$. Thereby,

$$\begin{aligned}\int \deg_G^{x,x'}(o) dU(G^{(n_i)})([G,o]) &\ge \int (\deg_G^{x,x'}(o) \wedge \alpha) dU(G^{(n_i)})([G,o]) \\ &\to \int (\deg_G^{x,x'}(o) \wedge \alpha) d\mu([G,o]).\end{aligned}$$

Sending α to infinity on the right hand side and using the monotone convergence theorem, we realize that

$$\liminf_{i\to\infty} \int \deg_G^{x,x'}(o) dU(G^{(n_i)})([G,o]) \ge \deg_{x,x'}(\mu). \tag{3.25}$$

On the other hand, we have

$$\int \deg_G^{x,x'}(o) dU(G^{(n_i)})([G,o]) = \begin{cases} m^{(n_i)}(x,x')/n & x \neq x', \\ 2m^{(n_i)}(x,x')/n & x = x'. \end{cases}$$

We know that if $x \neq x'$ we have $m^{(n)}(x,x')/n \to d_{x,x'}$, and we also have $m^{(n)}(x,x)/n \to d_{x,x}/2$ for all x. Comparing this with (3.25), we realize that if $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} > -\infty$ then, for all $x, x' \in \Xi$, we have $d_{x,x'} \geq \deg_{x,x'}(\mu)$. Similarly, using the fact that for all $\theta \in \Theta$ the mapping $[G,o] \mapsto \mathbb{1}[\tau_G(o) = \theta]$ is continuous and bounded on $\bar{\mathcal{G}}_*$, we realize that, if $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} > -\infty$, with the sequence $G^{(n_i)}$ as above we have $u^{(n_i)}(\theta) \to \Pi_\theta(\mu)$. But $u^{(n)}(\theta)/n \to q_\theta$. This means that $Q = \vec{\Pi}(\mu)$. As a result, to complete the proof, we assume that for some $\tilde{x}, \tilde{x}' \in \Xi$, we have $d_{\tilde{x},\tilde{x}'} > \deg_{\tilde{x},\tilde{x}'}(\mu)$ and then we show that $\overline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$. In order to do this it suffices to prove that for any sequence $\epsilon_n \to 0$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \left(\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| - \|m^{(n)}\|_1 \log n \right) = -\infty. \tag{3.26}$$

For an integer $\Delta > 0$ define $A_\Delta := \{[G,o] \in \bar{\mathcal{G}}_* : \deg_G^{\tilde{x},\tilde{x}'}(o) > \Delta\}$. Recall that, by definition of the Lévy–Prokhorov distance, if $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$ then

$$U(G^{(n)})(A_\Delta) \leq \mu(A_\Delta^{\epsilon_n}) + \epsilon_n, \tag{3.27}$$

where $A_\Delta^{\epsilon_n}$ is the ϵ_n–extension of the set A_Δ. Note that if we have $d_*([G,o],[G',o']) < 1/2$ for $[G,o]$ and $[G',o']$ in $\bar{\mathcal{G}}_*$ then we have $[G,o]_1 \equiv [G',o']_1$ and hence $\deg_G^{\tilde{x},\tilde{x}'}(o) = \deg_{G'}^{\tilde{x},\tilde{x}'}(o')$. This implies that if $\epsilon_n < 1/2$, which indeed holds for n large enough, then $A_\Delta^{\epsilon_n} = A_\Delta$. Therefore, using (3.27), we realize that if n is large enough so that $\epsilon_n < 1/2$, for any $\Delta > 0$ we have

$$|\{v \in [n] : \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) > \Delta\}| \leq n(\mu(A_\Delta) + \epsilon_n). \tag{3.28}$$

A similar argument shows that, for n large enough such that $\epsilon_n < 1/2$, for any integer k and any $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$, we have

$$|\{v \in [n] : \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) = k\}| \leq n\left(\mu\left(\left\{[G,o] : \deg_G^{\tilde{x},\tilde{x}'}(o) = k\right\}\right) + \epsilon_n\right). \tag{3.29}$$

Now, fix a sequence of integers Δ_n such that as $n \to \infty$, $\Delta_n \to \infty$, but $\Delta_n^2 \epsilon_n \to 0$. For instance, one could make the choice $\Delta_n = \lceil \epsilon_n^{-1/3} \rceil$. Using (3.29) for $k = 0, \dots, \Delta_n$, we realize that, for n large enough and for any $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$, we have

$$\begin{aligned} \sum_{v \in [n] : \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) \leq \Delta_n} \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) &\leq \sum_{k=0}^{\Delta_n} kn \left(\mu\left(\left\{[G,o] : \deg_G^{\tilde{x},\tilde{x}'}(o) = k\right\}\right) + \epsilon_n\right) \\ &\leq n\left(\mathbb{E}_\mu\left[\deg_G^{\tilde{x},\tilde{x}'}(o) \mathbb{1}\left[\deg_G^{\tilde{x},\tilde{x}'}(o) \leq \Delta_n\right]\right] + \Delta_n^2 \epsilon_n\right) \\ &\leq n \deg_{\tilde{x},\tilde{x}'}(\mu) + n\Delta_n^2 \epsilon_n. \end{aligned} \tag{3.30}$$

On the other hand, for $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)$ we have

$$\sum_{v \in [n]} \deg_G^{\tilde{x},\tilde{x}'}(v) = \begin{cases} m^{(n)}(\tilde{x},\tilde{x}') & \tilde{x} \neq \tilde{x}', \\ 2m^{(n)}(\tilde{x},\tilde{x}') & \tilde{x} = \tilde{x}'. \end{cases}$$

Moreover, as $n \to \infty$, for $\tilde{x} \neq \tilde{x}'$ we have $m^{(n)}(\tilde{x},\tilde{x}')/n \to d_{\tilde{x},\tilde{x}'}$ and we also have $m^{(n)}(\tilde{x},\tilde{x})/n \to d_{\tilde{x},\tilde{x}}/2$ for all $\tilde{x}$. Consequently, we have

$$\sum_{v \in [n]} \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) = n(d_{\tilde{x},\tilde{x}'} + \alpha_n),$$

where α_n is a sequence such that $\alpha_n \to 0$ as $n \to \infty$. Comparing this with (3.30), we realize that for n large enough and $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)$ we have

$$\sum_{v \in [n] : \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) > \Delta_n} \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) \geq n(d_{\tilde{x},\tilde{x}'} - \deg_{\tilde{x},\tilde{x}'}(\mu) + \alpha_n - \Delta_n^2 \epsilon_n).$$

Recall that, by assumption, $d_{\tilde{x},\tilde{x}'} > \deg_{\tilde{x},\tilde{x}'}(\mu)$, $\alpha_n \to 0$ and $\Delta_n^2 \epsilon_n \to 0$. Hence, there exists $\delta > 0$ such that for n large enough and $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)$ we have

$$\sum_{v \in [n] : \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) > \Delta_n} \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) \geq n\delta. \tag{3.31}$$

Comparing this to (3.28), we realize that, for n large enough, $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)$ implies that for the subset $S_n \subset [n]$ defined as $S_n := \{v \in [n] : \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) > \Delta_n\}$ we have $\sum_{v \in S} \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) \geq n\delta$ and $|S_n| \leq n\beta_n$, where $\beta_n := \mu(A_{\Delta_n}) + \epsilon_n$. Note that, since $\Delta_n \to \infty$ and $\epsilon_n \to 0$, we have $\beta_n \to 0$. Observe that $\sum_{v \in S} \deg_{G^{(n)}}^{\tilde{x},\tilde{x}'}(v) \geq n\delta$ implies that there are at least $n\delta/2$ many edges in $G^{(n)}$ with mark $\tilde{x}, \tilde{x}'$ with at least one endpoint in the set S_n. Let $\mathcal{S}_n$ denote the family of subsets $S_n \subset [n]$ with $|S_n| \leq n\beta_n$. For $S_n \in \mathcal{S}_n$, let $B_n(S_n)$ denote the set of simple marked graphs $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}$ such that there are at least $n\delta/2$ many edges with mark $\tilde{x}, \tilde{x}'$ with at least one endpoint in S_n. The above discussion implies that, for n large enough, we have

$$\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n) \subset \bigcup_{S \in \mathcal{S}_n} B_n(S_n). \tag{3.32}$$

Now, in order to find an upper bound for the size of the set on the right hand side, note that for $S_n \in \mathcal{S}_n$ there are $\binom{|S_n|}{2} + |S_n|(n - |S_n|)$ many slots to choose for the edges with at least one endpoint in S_n and with marks $\tilde{x}, \tilde{x}'$. Since there are at least $n\delta/2$ many such edges, the number of ways to pick the $\tilde{x}, \tilde{x}'$ edges of a graph in $B_n(S_n)$ is at most

$$\binom{\binom{|S_n|}{2} + |S_n|(n - |S_n|)}{n\delta/2} \binom{\binom{n}{2}}{m^{(n)}(\tilde{x},\tilde{x}') - n\delta/2} 2^{m^{(n)}(\tilde{x},\tilde{x}')} =: C_n(S_n).$$

Here, the term $2^{m^{(n)}(\tilde{x},\tilde{x}')}$ is an upper bound for the number of ways we can apply the marks $\tilde{x}$ and $\tilde{x}'$ for the chosen edges (if $\tilde{x} = \tilde{x}'$, this number is in fact 1). Now, since $|S_n| \leq n\beta_n$ for $S_n \in \mathcal{S}_n$, using the standard bound $\binom{r}{s} \leq (re/s)^s$ and Lemma 3.5, if n is large enough so that $\beta_n \leq 1/2$, we get

$$\begin{aligned}\max_{S_n \in \mathcal{S}_n} \log C_n(S_n) &\leq \frac{n\delta}{2} \log\left(ne\frac{\frac{\beta_n^2}{2} + \beta_n(1-\beta_n)}{\delta/2} \right) + (m^{(n)}(\tilde{x},\tilde{x}') - n\delta/2)\log n \\ &\quad + \frac{n}{2} + m^{(n)}(\tilde{x},\tilde{x}')\log 2 \\ &= m^{(n)}(\tilde{x},\tilde{x}')\log n + n\Bigg(\frac{1}{2} + \frac{\delta}{2} - \frac{\delta}{2}\log\frac{\delta}{2} + \frac{m^{(n)}(\tilde{x},\tilde{x}')}{n}\log 2 \\ &\quad + \frac{\delta}{2}\log\left(\frac{\beta_n^2}{2} + \beta_n(1-\beta_n)\right)\Bigg).\end{aligned}$$

Note that $\delta > 0$ is fixed. On the other hand, as $n \to \infty$, $m^{(n)}(\tilde{x},\tilde{x}')/n$ either converges to $d_{\tilde{x},\tilde{x}'}$ or $d_{\tilde{x},\tilde{x}'}/2$, depending on whether $\tilde{x} \neq \tilde{x}'$ or $\tilde{x} = \tilde{x}'$ respectively. But, in any case, it remains bounded. However, $\beta_n \to 0$, hence $\delta\log(\beta_n^2/2 + \beta_n(1-\beta_n)) \to -\infty$. Consequently, we have

$$\lim_{n\to\infty} \frac{1}{n}\left(\max_{S \in \mathcal{S}_n} \log C_n(S_n) - m^{(n)}(\tilde{x},\tilde{x}')\log n \right) = -\infty. \tag{3.33}$$

Now, in order to find an upper bound for $|B_n(S_n)|$ given $S_n \in \mathcal{S}_n$, we multiply the term $C_n(S_n)$ defined above by the number of ways we can add vertex marks to the graph and also add edges with marks different from $\tilde{x}, \tilde{x}'$, to get

$$|B_n(S_n)| \leq C_n(S_n)|\Theta|^n \prod_{\substack{x \leq x' \in \Xi \\ (x,x') \neq (\tilde{x},\tilde{x}')}} \binom{\binom{n}{2}}{m^{(n)}(x,x')} 2^{m^{(n)}(x,x')}.$$

Using (3.33) and Lemma 3.5 for each term, we realize that

$$\lim_{n\to\infty} \frac{1}{n}\left(\max_{S \in \mathcal{S}_n} \log |B_n(S_n)| - \|\vec{m}^{(n)}\|_1 \log n \right) = -\infty. \tag{3.34}$$

Moreover, if n is large enough so that $\beta_n < 1/2$, we have

$$|\mathcal{S}_n| \leq \sum_{k=0}^{n\beta_n} \binom{n}{k} \leq (1 + n\beta_n)\binom{n}{n\beta_n}.$$

Observe that, since $\beta_n \to 0$, we have $\frac{1}{n}\log|\mathcal{S}_n| \to 0$ as $n \to \infty$. Putting this together with (3.34) and comparing with (3.32), we arrive at (3.26), which completes the proof. $\square$

Proof of Proposition 3.2. Let $\mathcal{G}_*$ denote the space of isomorphism classes of rooted simple unmarked connected graphs, which is defined in a similar way as $\bar{\mathcal{G}}_*$, with the difference that vertices and edges do not carry marks. We can equip $\mathcal{G}_*$ with a local metric similar to that of $\bar{\mathcal{G}}_*$. With this, let $F : \bar{\mathcal{G}}_* \to \mathcal{G}_*$ be such that $[G, o]$ is mapped to $[\widetilde{G}, o]$ under F, where $\widetilde{G}$ is the unmarked graph obtained from G by removing all vertex and edge marks. For $[G, o]$ and $[G', o']$ in $\bar{\mathcal{G}}_*$, let $\widetilde{G}$ and $\widetilde{G}'$ be obtained from G and G' by removing vertex and edge marks, respectively. Observe that if $[G, o]_h \equiv [G', o']_h$ for $h \geq 0$, then $[\widetilde{G}, o]_h \equiv [\widetilde{G}', o']_h$. This means that F is 1–Lipschitz, and in particular continuous.

Now, let $\vec{m}^{(n)}, \vec{u}^{(n)}$ be any sequences such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$ and define $m_n = \|\vec{m}^{(n)}\|_1$. Moreover, for integer n, let $\mathcal{G}_{n,m_n}$ be the set of simple unmarked graphs on the vertex set $[n]$ having m_n edges. Observe that if $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)$ for some $\epsilon > 0$ and $n \in \mathbb{N}$, and $\widetilde{G}^{(n)} \in \mathcal{G}_{n,m_n}$ is the unmarked graph obtained from $G^{(n)}$ by removing all vertex and edge marks, then $U(\widetilde{G}^{(n)})$ is the pushforward of $U(G^{(n)})$ under the mapping F. Let $\rho \in \mathcal{P}(\mathcal{G}_*)$ be the pushforward of μ under F. Since F is 1–Lipschitz, it is easy to see that for $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}$, we have $d_{\text{LP}}(U(\widetilde{G}^{(n)}), \rho) \leq d_{\text{LP}}(U(G^{(n)}), \mu) < \epsilon$. Therefore, if $\mathcal{G}_{n,m_n}(\rho, \epsilon)$ denotes the set of unmarked graphs $H \in \mathcal{G}_{n,m_n}$ such that $d_{\text{LP}}(U(H), \rho) < \epsilon$, the above discussion implies that for $G^{(n)} \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}$, we have $\widetilde{G}^{(n)} \in \mathcal{G}_{n,m_n}(\rho, \epsilon)$. Moreover, for a simple unmarked graph $H \in \mathcal{G}_{n,m_n}$, there are at most $(|\Xi|^2)^{m_n} |\Theta|^n$ many ways of adding marks to vertices and edges. Thereby,

$$|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)| \leq |\mathcal{G}_{n,m_n}(\rho, \epsilon)| (|\Xi|^2)^{m_n} |\Theta|^n.$$

Note that as $n \to \infty$, $m_n/n \to d/2$ where $d = \deg(\mu) = \deg(\rho)$. Consequently,

$$\limsup_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)| - \|\vec{m}^{(n)}\|_1 \log n}{n} \leq \limsup_{n\to\infty} \frac{\log |\mathcal{G}_{n,m_n}(\rho, \epsilon)| - m_n \log n}{n} + d \log |\Xi| + \log |\Theta|. \tag{3.35}$$

Now, the assumption $\mu(\bar{\mathcal{T}}_*) < 1$ implies that $\rho(\mathcal{T}_*) < 1$. Hence, Theorem 1.2 in [BC15] implies that the unmarked BC entropy of ρ is $-\infty$, i.e.

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{\log |\mathcal{G}_{n,m_n}(\rho, \epsilon)| - m_n \log n}{n} = -\infty.$$

Comparing this with (3.35), we realize that $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$ which completes the proof. □

3.4.2 Lower bound

In this section, we prove the lower bound result of Proposition 3.3.

Proof of Proposition 3.3. For $x, x' \in \Xi$, let $d_{x,x'} := \deg_{x,x'}(\mu)$ and $\vec{d} := (d_{x,x'} : x, x' \in \Xi)$. Furthermore, for $\theta \in \Theta$, let $q_\theta := \Pi_\theta(\mu)$ and $Q := (q_\theta : \theta \in \Theta)$. We prove the result in two steps: first we assume that P has a finite support, and then relax this assumption.

Case 1: P has a finite support: Using Lemma 3.2 from Section 3.3.6 we realize that there exists a finite set $\Delta \subset \bar{\mathcal{T}}_*^h$ containing the support of P, and a sequence of simple marked graphs $\Gamma_n \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}$ such that $U(\Gamma_n)_h \Rightarrow P$ and, for all n, the support of $U(\Gamma_n)_h$ is contained in Δ. To find a lower bound for $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)$, we may restrict ourselves to the graphs $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}$ such that $U(G)_h = U(\Gamma_n)_h$, since

$$|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)| \geq |\{G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}} : U(G)_h = U(\Gamma_n)_h, d_{\mathrm{LP}}(U(G), \mu) < \epsilon\}|. \tag{3.36}$$

In order to find a lower bound for the right hand side of (3.36), we employ the tools from Section 3.3.

More precisely, define $\mathcal{F} \subset \Xi \times \bar{\mathcal{T}}_*^{h-1}$ to be the set comprised of $T[o, v]_{h-1}$ and $T[v, o]_{h-1}$ for all $[T, o] \in \Delta$ and $v \sim_T o$. Since Δ is finite, $\mathcal{F}$ is also finite and hence can be identified with the set of integers $\{1, \dots, L\}$ where $L = |\mathcal{F}|$. Moreover, define the color set $\mathcal{C} := \mathcal{F} \times \mathcal{F}$. Also, let δ be the maximum degree at the root among the members of Δ. Since the support of $U(\Gamma_n)_h$ lies in Δ, the colored version of Γ_n, $\mathsf{C}(\Gamma_n)$, is a member of $\mathcal{G}(\mathcal{C})$. Let $\vec{D}^{(n)} := \vec{D}^{\mathsf{C}(\Gamma_n)}$ be the colored degree sequence of $\mathsf{C}(\Gamma_n)$. Recall that, for $t, t' \in \mathcal{F}$ and $v \in [n]$, we have $D^{(n)}_{t,t'}(v) = E_h(t, t')(\Gamma_n, v)$. Moreover, since the support of $U(\Gamma_n)_h$ lies in Δ, we have $\vec{D}^{(n)}(v) \in \mathcal{M}_L^{(\delta)}$ for all n and $v \in [n]$. Furthermore, define $\vec{\beta}^{(n)} = (\beta^{(n)}(v) : v \in [n])$ such that for $v \in [n]$, $\beta^{(n)}(v) := \tau_{\Gamma_n}(v)$.

From Corollary 3.3, we know that $N_h(\Gamma_n)$, which is the number of simple marked graphs G in $\bar{\mathcal{G}}_n$ such that $U(G)_h = U(\Gamma_n)_h$, is precisely $n(\vec{D}^{(n)}, \vec{\beta}^{(n)})|\mathcal{G}(\vec{D}^{(n)}, 2h+1)|$. Note that if $U(G)_h = U(\Gamma_n)_h$, then $\vec{m}_G = \vec{m}_{\Gamma_n} = \vec{m}^{(n)}$ and $\vec{u}_G = \vec{u}_{\Gamma_n} = \vec{u}^{(n)}$, thus $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}$. Moreover, from the proof of Corollary 3.3, we know that for two permutations π and π', if $((\vec{D}^{(n)})^\pi, (\vec{\beta}^{(n)})^\pi) \neq ((\vec{D}^{(n)})^{\pi'}, (\vec{\beta}^{(n)})^{\pi'})$, the sets $\{\mathsf{MCB}_{(\vec{\beta}^{(n)})^\pi}(H) : H \in \mathcal{G}((\vec{D}^{(n)})^\pi, 2h+1)\}$ and $\{\mathsf{MCB}_{(\vec{\beta}^{(n)})^{\pi'}}(H) : H \in \mathcal{G}((\vec{D}^{(n)})^{\pi'}, 2h+1)\}$ are disjoint. On the other hand, for $H_n \neq H'_n \in \mathcal{G}(\vec{D}^{(n)}, 2h+1)$, we have $\mathsf{MCB}_{\vec{\beta}^{(n)}}(H_n) \neq \mathsf{MCB}_{\vec{\beta}^{(n)}}(H'_n)$. These observations, together with (3.36), imply that with $\tilde{H}_n$ being uniformly distributed in $\mathcal{G}(\vec{D}^{(n)}, 2h+1)$ and $\tilde{G}_n := \mathsf{MCB}_{\vec{\beta}^{(n)}}(\tilde{H}_n)$, we have

$$\begin{aligned}
|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon)| &\geq n(\vec{D}^{(n)}, \vec{\beta}^{(n)})|\{H_n \in \mathcal{G}(\vec{D}^{(n)}, 2h+1) : d_{\mathrm{LP}}(U(\mathsf{MCB}_{\vec{\beta}^{(n)}}(H_n)), \mu) < \epsilon\}| \\
&= n(\vec{D}^{(n)}, \vec{\beta}^{(n)}) \left|\mathcal{G}^{(n)}(\vec{D}^{(n)}, 2h+1)\right| \mathbb{P}\left(d_{\mathrm{LP}}(U(\tilde{G}_n), \mu) < \epsilon\right) \\
&= N_h(\Gamma_n)\mathbb{P}\left(d_{\mathrm{LP}}(U(\tilde{G}_n), \mu) < \epsilon\right).
\end{aligned}$$

From Proposition 3.7, we know that, for any $\epsilon > 0$, $\mathbb{P}\left(d_{\mathrm{LP}}(U(\tilde{G}_n), \mu) < \epsilon\right) \to 1$ as $n \to \infty$.

Therefore, we have

$$\liminf_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)| - \|\vec{m}^{(n)}\|_1 \log n}{n} \geq \liminf_{n\to\infty} \frac{\log N_h(\Gamma_n) - \|\vec{m}^{(n)}\|_1 \log n}{n}.$$

Consequently, if we show that

$$\lim_{n\to\infty} \frac{1}{n} \left(\log N_h(\Gamma_n) - \|\vec{m}^{(n)}\|_1 \log n\right) = J_h(P), \tag{3.37}$$

then we can conclude that for $\epsilon > 0$, $\underline{\Sigma}_{\vec{d},Q}(\mu,\epsilon)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq J_h(P)$ and hence $\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \geq J_h(P)$, which completes the proof for this case. Thereby, it suffices to show (3.37).

In order to do this, first note that, as a result of Lemma A.4 in Appendix A.2, for $v, w \in [n]$ we have $(\Gamma_n, v)_h \equiv (\Gamma_n, w)_h$ iff $(\beta^{(n)}(v), D^{(n)}(v)) = (\beta^{(n)}(w), D^{(n)}(w))$. Thereby, since $U(\Gamma_n)_h \Rightarrow P$ and Δ is finite, we have

$$\lim_{n\to\infty} \frac{1}{n} \log n(\vec{D}^{(n)}, \vec{\beta}^{(n)}) = H(P). \tag{3.38}$$

Moreover, from Corollary 3.1 and Stirling's approximation, if, for $c \in \mathcal{C}$, $S^{(n)}_c$ denotes $\sum_{v=1}^n D^{(n)}_c(v)$, we have

$$\begin{aligned} \log |\mathcal{G}(\vec{D}^{(n)}, 2h+1)| &= \sum_{c\in\mathcal{C}_<} \left(S^{(n)}_c \log S^{(n)}_c - S^{(n)}_c\right) + \sum_{c\in\mathcal{C}_=} \left(\frac{S^{(n)}_c}{2} \log S^{(n)}_c - \frac{S^{(n)}_c}{2}\right) \\ &\quad - \sum_{c\in\mathcal{C}} \sum_{v=1}^n \log D^{(n)}_c(v)! + o(n) \\ &= \frac{1}{2} \sum_{c\in\mathcal{C}} \left(S^{(n)}_c \log S^{(n)}_c - S^{(n)}_c\right) - \sum_{c\in\mathcal{C}} \sum_{v=1}^n \log D^{(n)}_c(v)! + o(n). \end{aligned} \tag{3.39}$$

Here, we have used the following facts: (i) $\log k! = k \log k - k + o(k)$, (ii) $\log(k-1)!! = \frac{k}{2} \log k - \frac{k}{2} + o(k)$, (iii) for all $c \in \mathcal{C}$, $\limsup_{n\to\infty} S^{(n)}_c/n < \infty$ or equivalently $S^{(n)}_c = O(n)$, and (iv) for $c \in \mathcal{C}$, $S^{(n)}_c = S^{(n)}_{\bar{c}}$. Note that, since $U(\Gamma_n)_h \Rightarrow P$ and Δ is finite, for each $c = (t, t') \in \mathcal{C}$ we have

$$\frac{1}{n} S^{(n)}_c = \frac{1}{n} \sum_{v=1}^n D^{(n)}_c(v) \underset{n\to\infty}{\longrightarrow} \mathbb{E}_P \left[E_h(t,t')(T,o)\right] = e_P(t,t').$$

Likewise, for $c = (t, t') \in \mathcal{C}$, we have

$$\frac{1}{n} \sum_{v=1}^n \log D^{(n)}_c(v)! \underset{n\to\infty}{\longrightarrow} \mathbb{E}_P \left[\log E_h(t,t')(T,o)!\right].$$

Using these in (3.39) and simplifying, we get

$$\begin{aligned}
\log |\mathcal{G}(\vec{D}^{(n)}, 2h+1)| &= \frac{n}{2} \sum_{c \in \mathcal{C}} \left(\frac{S_c^{(n)}}{n} \log \frac{S_c^{(n)}}{n} + \frac{S_c^{(n)}}{n} \log n - \frac{S_c^{(n)}}{n} \right) \\
&\quad - n \sum_{c \in \mathcal{C}} \frac{1}{n} \sum_{v=1}^{n} \log D_c^{(n)}(v)! + o(n) \\
&= \|\vec{m}^{(n)}\|_1 \log n - \|\vec{m}^{(n)}\|_1 + \frac{n}{2} \sum_{t,t' \in \mathcal{F}} e_P(t,t') \log e_P(t,t') \\
&\quad - n \sum_{t,t' \in \mathcal{F}} \mathbb{E}_P \left[\log E_h(t,t')(T,o)! \right] + o(n),
\end{aligned} \tag{3.40}$$

where in the second line we have used $\sum_{c \in \mathcal{C}} S_c^{(n)} = 2\|\vec{m}^{(n)}\|_1$. Note that, since $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$, as $n \to \infty$ we have $\|\vec{m}^{(n)}\|_1/n \to \deg(\mu)/2$. From (3.38) and (3.40), with $d := \deg(\mu)$, we get

$$\begin{aligned}
\frac{\log N_h(\Gamma_n) - \|\vec{m}^{(n)}\|_1 \log n}{n} &= H(P) - \frac{d}{2} + \frac{1}{2} \sum_{t,t' \in \mathcal{F}} e_P(t,t') \log e_P(t,t') \\
&\quad - \sum_{t,t' \in \mathcal{F}} \mathbb{E}_P \left[\log E_h(t,t')(T,o)! \right] + o(1) \\
&= H(P) - \frac{d}{2} + \frac{d}{2} \sum_{t,t' \in \mathcal{F}} \frac{e_P(t,t')}{d} \left(\log d + \log \frac{e_P(t,t')}{d} \right) \\
&\quad - \sum_{t,t' \in \mathcal{F}} \mathbb{E}_P \left[\log E_h(t,t')(T,o)! \right] + o(1) \\
&\overset{(a)}{=} -s(d) + H(P) + \frac{d}{2} \sum_{t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}} \pi_P(t,t') \log \pi_P(t,t') \\
&\quad - \sum_{t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}} \mathbb{E}_P \left[\log E_h(t,t')(T,o)! \right] + o(1) \\
&= J_h(P) + o(1),
\end{aligned}$$

where in (a) we have used the facts that the support of P is contained in Δ and

$$\sum_{t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}} e_P(t,t') = d.$$

This shows (3.37) and thus completes the proof for the finite support case.

Case 2: For general P: We use a truncation procedure together with the proof in the above finite support case. More precisely, for an integer $k > 1$, we start from a random rooted marked tree (T, o) with law μ and, for all vertices v in T with degree more than k, we remove all the edges connected to v. Let $T^{(k)}$ denote the connected component of the root in the resulting forest. With this, define $\mu^{(k)}$ to be the law of $[T^{(k)}, o]$. It is easy to see that $\mu^{(k)}$ is unimodular. Furthermore, let $P_k := (\mu^{(k)})_h \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ be the law of the depth h neighborhood of the root in $\mu^{(k)}$. Since $\mu^{(k)}$ is unimodular, P_k is admissible. On the other hand, P_k has a finite support, and hence P_k is strongly admissible, i.e. $P_k \in \mathcal{P}_h$.

With the above construction, we have $\deg_{x,x'}(\mu^{(k)}) \le \deg_{x,x'}(\mu)$ for all $x, x' \in \Xi$ and $\vec{\Pi}(\mu^{(k)}) = \vec{\Pi}(\mu)$. We do not directly apply the result of the previous case to P_k, since the sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$ which might be different from $(\vec{\deg}(\mu^{(k)}), \vec{\Pi}(\mu^{(k)}))$. Instead, we modify $\mu^{(k)}$ to obtain a measure $\widetilde{\mu}^{(k)}$ such that $(\vec{\deg}(\widetilde{\mu}^{(k)}), \vec{\Pi}(\widetilde{\mu}^{(k)})) = (\vec{\deg}(\mu), \vec{\Pi}(\mu))$. In order to do this, for each pair of edge marks $x \le x' \in \Xi$, we choose an integer $\tilde{d}_{x,x'} > 2(|\Xi|^2 d_{x,x'} \vee 1)$. Moreover, define $\nu_{x,x'}$ to be the law of $[T, o] \in \bar{\mathcal{T}}_*$ where (T, o) is the random rooted marked $\tilde{d}_{x,x'}$–regular tree defined as follows. With probability 1/2, we have

$$\xi_T(v, w) = \begin{cases} x & \text{dist}_T(o, w) \text{ is even,} \\ x' & \text{dist}_T(o, w) \text{ is odd,} \end{cases} \qquad \forall v, w \in V(T),$$

and with probability 1/2, we have

$$\xi_T(v, w) = \begin{cases} x' & \text{dist}_T(o, w) \text{ is even,} \\ x & \text{dist}_T(o, w) \text{ is odd,} \end{cases} \qquad \forall v, w \in V(T).$$

Additionally, each vertex in T is independently given a mark with distribution $\vec{\Pi}(\mu)$. It is easy to check that $\nu_{x,x'}$ is unimodular, $\vec{\Pi}(\nu_{x,x'}) = \vec{\Pi}(\mu)$, and $\deg_{x,x'}(\nu_{x,x'}) = \deg_{x',x}(\nu_{x,x'}) = \tilde{d}_{x,x'}/2$. Let $U_{x,x'} := (\nu_{x,x'})_h \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ be the law of the depth h neighborhood of the root in $\nu_{x,x'}$. Due to the way we chose $\tilde{d}_{x,x'}$ for $x \le x' \in \Xi$, we can choose $p_k \in [0, 1]$ together with nonnegative numbers $(\alpha^k_{x,x'} : x \le x' \in \Xi)$ so that $p_k + \sum_{x \le x' \in \Xi} \alpha^k_{x,x'} = 1$ and such that with

$$\widetilde{P}_k := p_k P_k + \sum_{x \le x' \in \Xi} \alpha^k_{x,x'} U_{x,x'}, \tag{3.41}$$

we have $\mathbb{E}_{\widetilde{P}_k}\left[\deg_T^{x,x'}(o)\right] = d_{x,x'}$ for all $x, x' \in \Xi$. More precisely, with $d^k_{x,x'} := \deg_{x,x'}(\mu^{(k)})$, we may set

$$p_k := \frac{1 - \sum_{x \le x' \in \Xi} 2 d_{x,x'}/\tilde{d}_{x,x'}}{1 - \sum_{x \le x' \in \Xi} 2 d^k_{x,x'}/\tilde{d}_{x,x'}},$$

and, for $x \le x' \in \Xi$,

$$\alpha^k_{x,x'} := \frac{2(d_{x,x'} - p_k d^k_{x,x'})}{\tilde{d}_{x,x'}}.$$

Then, using $\tilde{d}_{x,x'} > 2(|\Xi|^2 d_{x,x'} \vee 1)$ and $d^k_{x,x'} < d_{x,x'}$, all the desired properties mentioned above would follow. On the other hand, since $\deg_{x,x'}(\mu^{(k)}) \uparrow \deg_{x,x'}(\mu)$ as $k \to \infty$, we have $p_k \to 1$

as $k \to \infty$. Furthermore, since P_k is admissible and $\nu_{x,x'}$ is unimodular, $\widetilde{P}_k$ is admissible, and in addition has a finite support. This implies that $\widetilde{P}_k$ is strongly admissible, i.e. $\widetilde{P}_k \in \mathcal{P}_h$. Thus, with $\widetilde{\mu}^{(k)} := \mathsf{UGWT}_h(\widetilde{P}_k)$, we have $\vec{\Pi}(\widetilde{\mu}^{(k)}) = \vec{\Pi}(\mu)$ and $\vec{\deg}(\widetilde{\mu}^{(k)}) = \vec{\deg}(\mu)$. Now, we claim that

$$\lim_{k\to\infty} e_{\widetilde{P}_k}(t,t') = e_P(t,t') \qquad \forall t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}. \tag{3.42}$$

In order to show this, note that from (3.41) we have

$$e_{\widetilde{P}_k}(t,t') = p_k e_{P_k}(t,t') + \sum_{x \le x' \in \Xi} \alpha^k_{x,x'} e_{U_{x,x'}}(t,t'), \qquad \forall t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}. \tag{3.43}$$

But $U_{x,x'}$ are fixed, Ξ is finite, and $\alpha^k_{x,x'} \to 0$. Hence, to show (3.42), it suffices to show that

$$\lim_{k\to\infty} e_{P_k}(t,t') = e_P(t,t') \qquad \forall t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}. \tag{3.44}$$

Observe that for $t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$ we have $e_{P_k}(t,t') = \mathbb{E}_\mu\left[E_h(t,t')([T^{(k)},o])\right]$. But, for $[T,o] \in \bar{\mathcal{T}}_*$, if k is large enough, $[T^{(k)},o]_h = [T,o]_h$. Thereby, $E_h(t,t')([T^{(k)},o]) \to E_h(t,t')([T,o])$ as $k \to \infty$. On the other hand, $E_h(t,t')([T^{(k)},o]) \le \deg_{T^{(k)}}(o) \le \deg_T(o)$. Hence,

$$\mathbb{E}_\mu\left[E_h(t,t')([T^{(k)},o])\right] \le \mathbb{E}_\mu\left[\deg_T(o)\right] < \infty.$$

This together with the dominated convergence theorem implies (3.44). Thus, we arrive at (3.42). On the other hand, we have $P_k \Rightarrow P$, and from (3.41) we have $\widetilde{P}_k \Rightarrow P$. Therefore Lemma 2.4 in Appendix A.4 implies that $\widetilde{\mu}^{(k)} \Rightarrow \mu$ as $k \to \infty$. Therefore, from Lemma 3.1 and the lower bound for the finite support case, we have

$$\underline{\Sigma}_{\vec{d},Q}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \ge \limsup_{k\to\infty} \underline{\Sigma}_{\vec{d},Q}(\mathsf{UGWT}_h(\widetilde{P}_k))|_{(\vec{m}^{(n)},\vec{u}^{(n)})} \ge \limsup_{k\to\infty} J_h(\widetilde{P}_k) \ge \liminf_{k\to\infty} J_h(\widetilde{P}_k). \tag{3.45}$$

Here, all the entropy terms are obtained via the same sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$. Therefore, it suffices to show that $\liminf_{k\to\infty} J_h(\widetilde{P}_k) \ge J_h(P)$. Note that, by definition, we have

$$J_h(\widetilde{P}_k) = -s(d) + H(\widetilde{P}_k) - \frac{d}{2}H(\pi_{\widetilde{P}_k}) - \sum_{t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}} \mathbb{E}_{\widetilde{P}_k}\left[\log E_h(t,t')!\right],$$

where $d = \deg(\mu)$. We claim that

$$\liminf_{k\to\infty} J_h(\widetilde{P}_k) \ge \liminf_{k\to\infty} J_h(P_k). \tag{3.46}$$

Note that P_k is admissible and is finitely supported, and hence $H(P_k) < \infty$. Furthermore, since $d > 0$, for k large enough P_k has positive expected degree at the root. Hence $J_h(P_k)$ is well defined for k large enough. In order to show (3.46), first note that if d_k is the average degree at the root in P_k then we have $d_k \to d$ as $k \to \infty$. Hence we have

$$\lim_{k\to\infty} s(d_k) = s(d). \tag{3.47}$$

On the other hand, using (3.41), we have

$$H(\widetilde{P}_k) = p_k \log \frac{1}{p_k} + \sum_{x \leq x' \in \Xi} \alpha^k_{x,x'} \log \frac{1}{\alpha^k_{x,x'}} + p_k H(P_k) + \sum_{x \leq x' \in \Xi} \alpha^k_{x,x'} H(U_{x,x'}).$$

Here, $U_{x,x'}$ are fixed distributions and have no dependence on k. Also, $p_k \to 1$ and $\alpha^k_{x,x'} \to 0$ for all $x \leq x' \in \Xi$. Hence, if we show that the sequence $H(P_k)$ is bounded, we can conclude that $\liminf_{k\to\infty} H(\widetilde{P}_k) \geq \liminf_{k\to\infty} H(P_k)$. In order to show that the sequence $H(P_k)$ is bounded, recall that P_k is the distribution of $[T^{(k)}, o]_h$. Observe that $[T^{(k)}, o]_h$ is a function of $[T, o]_{h+1}$. The reason is that, by definition, $T^{(k)}$ is obtained from T by removing all the edges connected to vertices with degree more than k, and the degree of a vertex with distance at most h from the root is completely determined by $[T, o]_{h+1}$. This means that $H(P_k) \leq H(R)$ where $R := \mu_{h+1} \in \mathcal{P}(\bar{\mathcal{T}}^{h+1}_*)$ is the law of the $h+1$ neighborhood of the root in μ. From Lemma 2.6, we have $R \in \mathcal{P}_{h+1}$ and hence $H(R) < \infty$. This shows that $H(P_k)$ is a bounded sequence and

$$\liminf_{k\to\infty} H(\widetilde{P}_k) \geq \liminf_{k\to\infty} H(P_k). \tag{3.48}$$

On the other hand, from (3.43), we have

$$\pi_{\widetilde{P}_k} = \frac{d_k}{d} p_k \pi_{P_k} + \sum_{x \leq x' \in \Xi} \frac{\tilde{d}_{x,x'}}{d} \alpha^k_{x,x'} \pi_{U_{x,x}}.$$

But, as $k \to \infty$, we have $d_k \to d$, $p_k \to 1$, and $\alpha^k_{x,x'} \to 0$ for all $x \leq x' \in \Xi$. Also, $U_{x,x'}$ for $x \leq x' \in \Xi$ are fixed and do not depend on k. Thereby, we conclude that

$$\limsup_{k\to\infty} H(\pi_{\widetilde{P}_k}) \leq \limsup_{k\to\infty} H(\pi_{P_k}). \tag{3.49}$$

Moreover, from (3.41), we have

$$\begin{aligned} \mathbb{E}_{\widetilde{P}_k}\left[\sum_{t,t' \in \Xi \times \bar{\mathcal{T}}^{h-1}_*} \log E_h(t,t')!\right] &= p_k \mathbb{E}_{P_k}\left[\sum_{t,t' \in \Xi \times \bar{\mathcal{T}}^{h-1}_*} \log E_h(t,t')!\right] \\ &\quad + \sum_{x \leq x' \in \Xi} \alpha^k_{x,x'} \mathbb{E}_{U_{x,x'}}\left[\sum_{t,t' \in \Xi \times \bar{\mathcal{T}}^{h-1}_*} \log E_h(t,t')!\right]. \end{aligned}$$

Again, as $k \to \infty$, we have $p_k \to 1$ and $\alpha^k_{x,x'} \to 0$ for all $x \leq x' \in \Xi$. Hence

$$\limsup_{k\to\infty} \mathbb{E}_{\widetilde{P}_k}\left[\sum_{t,t' \in \Xi \times \bar{\mathcal{T}}^{h-1}_*} \log E_h(t,t')!\right] \leq \limsup_{k\to\infty} \mathbb{E}_{P_k}\left[\sum_{t,t' \in \Xi \times \bar{\mathcal{T}}^{h-1}_*} \log E_h(t,t')!\right]. \tag{3.50}$$

Putting together (3.47), (3.48), (3.49) and (3.50), we arrive at (3.46). Comparing this with (3.45), in order to complete the proof, it suffices to show that

$$\liminf_{k\to\infty} J_h(P_k) \geq J_h(P). \tag{3.51}$$

Without loss of generality, for the rest of the proof, we may assume that $J_h(P) > -\infty$, otherwise nothing remains to be proved. In order to show (3.51), it suffices to show the following

$$\liminf_{k\to\infty} H(P_k) \geq H(P), \tag{3.52a}$$

$$\lim_{k\to\infty} \sum_{t,t'\in\Xi\times\bar{\mathcal{T}}_*^{h-1}} \mathbb{E}_{P_k}\left[\log E_h(t,t')!\right] = \sum_{t,t'\in\Xi\times\bar{\mathcal{T}}_*^{h-1}} \mathbb{E}_P\left[\log E_h(t,t')!\right], \tag{3.52b}$$

$$\limsup_{k\to\infty} H(\pi_{P_k}) \leq H(\pi_P). \tag{3.52c}$$

First, to show (3.52a), note that for all $[\tilde{T},\tilde{o}] \in \bar{\mathcal{T}}_*^h$, we have $P_k([\tilde{T},\tilde{o}]) = \int \mathbb{1}[[T^{(k)},o]_h = [\tilde{T},\tilde{o}]] d\mu([T,o])$. Therefore, the dominated convergence theorem implies that $P_k([\tilde{T},\tilde{o}]) \to P([\tilde{T},\tilde{o}])$ as $k \to \infty$. Hence, (3.52a) follows from this and lower semi–continuity of the Shannon entropy (see, for instance [HY10]).

Now we turn to showing (3.52b). Define $\mathcal{C} := (\Xi \times \bar{\mathcal{T}}_*^{h-1}) \times (\Xi \times \bar{\mathcal{T}}_*^{h-1})$. Moreover, for $r \in \bar{\mathcal{T}}_*$, let $F(r) := \sum_{c\in\mathcal{C}} \log E_h(c)(r)!$. With this, we have $\sum_{t,t'\in\Xi\times\bar{\mathcal{T}}_*^{h-1}} \mathbb{E}_{P_k}\left[\log E_h(t,t')!\right] = \mathbb{E}_\mu\left[F([T^{(k)},o])\right]$. Recall that $[T^{(k)},o] \in \bar{\mathcal{T}}_*$, as defined above, is the rooted tree obtained from $[T,o]$ by removing all edges connected to vertices with degree larger than k followed by taking the connected component of the root. Likewise, the right hand side of (3.52b) is precisely $\mathbb{E}_\mu\left[F([T,o])\right]$. Observe that for each $[T,o] \in \bar{\mathcal{T}}_*$, if k is large enough, $[T^{(k)},o]_h = [T,o]_h$. Thereby, $F([T^{(k)},o]) \to F([T,o])$ pointwise. Now, for $[T,o] \in \bar{\mathcal{T}}_*$, using the inequality $\log a! \leq a \log a$ that holds for any nonnegative integer a by interpreting $0 \log 0 = 0$, we get

$$\begin{aligned} F([T,o]) &\leq \sum_{c\in\mathcal{C}} E_h(c)(T,o) \log E_h(c)(T,o) \\ &\leq \sum_{c\in\mathcal{C}} E_h(c)(T,o) \log \deg_T(o) \\ &= \deg_T(o) \log \deg_T(o). \end{aligned}$$

Consequently,

$$\mathbb{E}_P\left[|F([T^{(k)},o])|\right] = \mathbb{E}_P\left[F([T^{(k)},o])\right] \leq \mathbb{E}_P\left[\deg_{T^{(k)}}(o) \log \deg_{T^{(k)}}(o)\right] \leq \mathbb{E}_P\left[\deg_T(o) \log \deg_T(o)\right].$$

The fact that $P \in \mathcal{P}_h$ implies that the right hand side is finite. Thereby, we arrive at (3.52b) using the dominated convergence theorem.

Next, we show (3.52c). Recall that, without loss of generality, we have assumed that $J_h(P) > -\infty$, which means $H(\pi_P) < \infty$. Consequently, we have

$$\begin{aligned}\sum_{c\in\mathcal{C}} |e_P(c)\log e_P(c)| &\le \sum_{c\in\mathcal{C}} |e_P(c)\log e_P(c) - e_P(c)\log d| + |e_P(c)\log d| \\ &= \sum_{c\in\mathcal{C}} e_P(c)\log\frac{d}{e_P(c)} + \sum_{c\in\mathcal{C}} e_P(c)\log d \\ &= dH(\pi_P) + d\log d < \infty,\end{aligned} \tag{3.53}$$

where, in the second line, we have used the fact that $e_P(c) \le d$ for all $c \in \mathcal{C}$. Therefore, the sequence $e_P(c)\log e_P(c)$ is absolutely summable. Hence, we may write

$$H(\pi_P) = \sum_{c\in\mathcal{C}} \frac{e_P(c)}{d}\log\frac{d}{e_P(c)} = \log d - \frac{1}{d}\sum_{c\in\mathcal{C}} e_P(c)\log e_P(c).$$

On the other hand, P_k has finite support. Hence, with $d_k = \sum_{c\in\mathcal{C}} e_{P_k}(c)$ being the expected degree at the root in P_k, we have

$$H(\pi_{P_k}) = \log d_k - \frac{1}{d_k}\sum_{c\in\mathcal{C}} e_{P_k}(c)\log e_{P_k}(c).$$

Therefore, as $d_k \uparrow d$, in order to show (3.52c), it suffices to show that

$$\liminf_{k\to\infty}\sum_{c\in\mathcal{C}} e_{P_k}(c)\log e_{P_k}(c) \ge \sum_{c\in\mathcal{C}} e_P(c)\log e_P(c). \tag{3.54}$$

Recall from (3.44) that for all $c \in \mathcal{C}$, we have $e_{P_k}(c) \to e_P(c)$ as $k \to \infty$. Now, for a nonnegative integer δ, define $\mathcal{C}^{(\delta)} \subset \mathcal{C}$ to be the set of $(t,t') \in \mathcal{C}$ such that all vertices in the subgraph components of t and t', i.e. $t[s]$ and $t'[s]$, have degrees bounded by δ. Therefore, due to (3.53) and the fact that $\mathcal{C} = \cup_{\delta=1}^{\infty}\mathcal{C}^{(\delta)}$, we have

$$\sum_{c\notin\mathcal{C}^{(\delta)}} |e_P(c)\log e_P(c)| < \epsilon_1(\delta), \tag{3.55}$$

where $\epsilon_1(\delta) \to 0$ as $\delta \to \infty$. Note that $\mathcal{C}^{(\delta)}$ is finite. This together with (3.44) and (3.55) implies that for $\delta > 0$ we have

$$\begin{aligned}\lim_{k\to\infty}\sum_{c\in\mathcal{C}^{(\delta)}} e_{P_k}(c)\log e_{P_k}(c) &= \sum_{c\in\mathcal{C}^{(\delta)}} e_P(c)\log e_P(c) \\ &\ge \sum_{c\in\mathcal{C}} e_P(c)\log e_P(c) - \epsilon_1(\delta).\end{aligned} \tag{3.56}$$

Hence, we may write

$$\liminf_{k\to\infty}\sum_{c\in\mathcal{C}} e_{P_k}(c)\log e_{P_k}(c) \ge \liminf_{k\to\infty}\sum_{c\in\mathcal{C}^{(\delta)}} e_{P_k}(c)\log e_{P_k}(c) + \liminf_{k\to\infty}\sum_{c\notin\mathcal{C}^{(\delta)}} e_{P_k}(c)\log e_{P_k}(c)$$

$$\geq \sum_{c\in\mathcal{C}} e_P(c)\log e_P(c) - \epsilon_1(\delta) + \liminf_{k\to\infty} \sum_{c\notin\mathcal{C}^{(\delta)}} e_{P_k}(c)\log e_{P_k}(c).$$

As this holds for all $\delta > 0$ and since $\epsilon_1(\delta) \to 0$ when $\delta \to 0$, in order to show (3.54) it suffices to prove that for all positive integers k and δ such that $k > \delta$, we have

$$\sum_{c\notin\mathcal{C}^{(\delta)}} e_{P_k}(c)\log e_{P_k}(c) \geq -\epsilon_2(\delta), \tag{3.57}$$

where $\epsilon_2(\delta) \to 0$ as $\delta \to \infty$. Now, we fix positive integers $k > \delta$ and show that (3.57) holds for an appropriate choice of $\epsilon_2(\delta)$. For an integer $r > 0$, let $\mathcal{A}_h^{(r)} \subset \bar{\mathcal{T}}_*^h$ be the set of marked rooted trees of depth at most h where all degrees are bounded by r. We define $\mathcal{A}_{h+1}^{(r)} \subset \bar{\mathcal{T}}_*^{h+1}$ similarly. Note that P_k has a finite support and so the left hand side of (3.57) is a finite sum. Indeed, P_k is supported on the finite set $\mathcal{A}_h^{(k)} \subset \bar{\mathcal{T}}_*^h$ and $e_{P_k}(c) = 0$ for $c \notin \mathcal{C}^{(k)}$. Consequently, we have

$$\begin{aligned}\sum_{c\notin\mathcal{C}^{(\delta)}} e_{P_k}(c)\log e_{P_k}(c) &= \sum_{c\in\mathcal{C}^{(k)}\setminus\mathcal{C}^{(\delta)}} \left(\sum_{s\in\mathcal{A}_h^{(k)}} P_k(s)E_h(c)(s)\right)\log\left(\sum_{s'\in\mathcal{A}_h^{(k)}} P_k(s')E_h(c)(s')\right)\\ &\geq \sum_{c\in\mathcal{C}^{(k)}\setminus\mathcal{C}^{(\delta)}}\sum_{s\in\mathcal{A}_h^{(k)}} P_k(s)E_h(c)(s)\log(P_k(s)E_h(c)(s)).\end{aligned} \tag{3.58}$$

Note that if $c \notin \mathcal{C}^{(\delta)}$ and $s \in \mathcal{A}_h^{(\delta)}$, we have $E_h(c)(s) = 0$. Thereby,

$$\begin{aligned}\sum_{c\notin\mathcal{C}^{(\delta)}} e_{P_k}(c)\log e_{P_k}(c) &\geq \sum_{s\in\mathcal{A}_h^{(k)}\setminus\mathcal{A}_h^{(\delta)}}\sum_{c\in\mathcal{C}^{(k)}\setminus\mathcal{C}^{(\delta)}} P_k(s)E_h(c)(s)\log(P_k(s)E_h(c)(s))\\ &\geq \sum_{s\in\mathcal{A}_h^{(k)}\setminus\mathcal{A}_h^{(\delta)}}\sum_{c\in\mathcal{C}^{(k)}\setminus\mathcal{C}^{(\delta)}} P_k(s)E_h(c)(s)\log P_k(s)\\ &\geq \sum_{[T,o]\in\mathcal{A}_h^{(k)}\setminus\mathcal{A}_h^{(\delta)}} \deg_T(o)P_k([T,o])\log P_k([T,o])\end{aligned} \tag{3.59}$$

where the last inequality uses the fact that for $[T,o] \in \mathcal{A}_h^{(k)} \setminus \mathcal{A}_h^{(\delta)}$, we have

$$\sum_{c\in\mathcal{C}^{(k)}\setminus\mathcal{C}^{(\delta)}} E_h(c)([T,o]) \leq \sum_{c\in\mathcal{C}} E_h(c)([T,o]) = \deg_T(o),$$

and $P_k([T,o])\log P_k([T,o]) \leq 0$. As we discussed above, for $[T,o] \in \bar{\mathcal{T}}_*$, $[T^{(k)},o]_h$ is determined by $[T,o]_{h+1}$, since the degree of all vertices up to depth h is determined by the structure of the tree up to depth $h+1$. Moreover, define $F_k : \bar{\mathcal{T}}_*^{h+1} \to \mathcal{A}_h^{(k)}$ such that for $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$, we have $F_k([T,o]) := [T^{(k)},o]_h$. With this, P_k is the pushforward of $R := \mu_{h+1}$

under the mapping F_k, i.e. for $[T,o] \in \mathcal{A}_h^{(k)}$, $P_k([T,o]) = R(F_k^{-1}([T,o]))$. On the other hand, for $[T,o] \in \mathcal{A}_h^{(k)}$, if $[T',o'] \in F_k^{-1}([T,o])$, then $R([T',o']) \leq R(F_k^{-1}([T,o])) = P_k([T,o])$. Using these in (3.59), we have

$$\begin{aligned}
\sum_{c \notin \mathcal{C}^{(\delta)}} e_{P_k}(c) \log e_{P_k}(c) &\geq \sum_{[T,o] \in \mathcal{A}_h^{(k)} \setminus \mathcal{A}_h^{(\delta)}} \sum_{[T',o'] \in F_k^{-1}([T,o])} \deg_T(o) R([T',o']) \log P_k([T,o]) \\
&\geq \sum_{[T,o] \in \mathcal{A}_h^{(k)} \setminus \mathcal{A}_h^{(\delta)}} \sum_{[T',o'] \in F_k^{-1}([T,o])} \deg_T(o) R([T',o']) \log R([T',o']) \\
&= \sum_{[T',o'] \in \bar{\mathcal{T}}_*^{h+1}} \mathbb{1}\left[F_k([T',o']) \notin \mathcal{A}_h^{(\delta)}\right] \deg_{F_k([T',o'])}(o') R([T',o']) \log R([T',o']).
\end{aligned}$$

Here, in the last equality, we were allowed to change the order of summations since all the terms are nonpositive. Also note that, by definition, for $[T',o'] \in \bar{\mathcal{T}}_*^{h+1}$ we have $F_k([T',o']) \in \mathcal{A}_h^{(k)}$. Since the mapping F_k decreases the degree of all the vertices, for all $[T',o'] \in \bar{\mathcal{T}}_*^{h+1}$ we have $\mathbb{1}\left[F_k([T',o']) \notin \mathcal{A}_h^{(\delta)}\right] \leq \mathbb{1}\left[[T',o'] \notin \mathcal{A}_{h+1}^{(\delta)}\right]$ and $\deg_{F_k([T',o'])}(o') \leq \deg_{[T',o']}(o')$. Using these observations in the above chain of inequalities, since all the terms in the summation are nonpositive, we get

$$\sum_{c \notin \mathcal{C}^{(\delta)}} e_{P_k}(c) \log e_{P_k}(c) \geq \sum_{[T',o'] \in \bar{\mathcal{T}}_*^{h+1} \setminus \mathcal{A}_{h+1}^{(\delta)}} \deg_{T'}(o') R([T',o']) \log R([T',o']). \tag{3.60}$$

Since P is strongly admissible, i.e. $P \in \mathcal{P}_h$, Lemma 2.6 implies that $R \in \mathcal{P}_{h+1}$, which means $H(R) < \infty$. Also, $\mathbb{E}_R\left[\deg_T(o) \log \deg_T(o)\right] = \mathbb{E}_P\left[\deg_T(o) \log \deg_T(o)\right] < \infty$. Therefore, from Lemma A.6, we have

$$\sum_{[T',o'] \in \bar{\mathcal{T}}_*^{h+1}} \deg_{T'}(o') R([T',o']) \log R([T',o']) > -\infty.$$

Since $\cup_\delta \mathcal{A}_{h+1}^{(\delta)} = \bar{\mathcal{T}}_*^{h+1}$, we have

$$\sum_{[T',o'] \in \bar{\mathcal{T}}_*^{h+1} \setminus \mathcal{A}_{h+1}^{(\delta)}} \deg_{T'}(o') R([T',o']) \log R([T',o']) \geq -\epsilon_2(\delta),$$

where $\epsilon_2(\delta) \to 0$ as $\delta \to \infty$. Putting this into (3.60), we arrive at (3.57), which completes the proof. □

3.4.3 Upper bound

In this section, we prove the upper bound result of Proposition 3.4.

Proof of Proposition 3.4. Let $P := \mu_h$. Since μ is unimodular and $d < \infty$, from Lemma 2.3 we see that P is admissible. Further, since $P \in \mathcal{P}(\bar{\mathcal{T}}_*)$ with $\mathbb{E}_P[\deg_T(o)] = d > 0$ and $H(P) < \infty$, we see that $J_h(P)$, as introduced in (2.9), is well-defined. From the local topology on $\bar{\mathcal{G}}_*$ one sees that, for all $h \geq 1$ and $\epsilon > 0$, there exists $\eta_1(\epsilon)$ such that, for all $\rho_1, \rho_2 \in \mathcal{P}(\bar{\mathcal{G}}_*)$, $d_{\mathrm{LP}}(\rho_1, \rho_2) < \epsilon$ implies $d_{\mathrm{TV}}((\rho_1)_h, (\rho_2)_h) < \eta_1(\epsilon)$. Here the function $\eta_1(.)$ depends only on h and has the property that $\eta_1(\epsilon) \to 0$ as $\epsilon \to 0$. Therefore, if for $\delta > 0$ we define

$$A^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P, \delta) := \{G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}} : d_{\mathrm{TV}}(U(G)_h, P) < \delta\},$$

we have

$$\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon) \subseteq A^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P, \eta_1(\epsilon)).$$

Hence, to show (3.3), it suffices to show that

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \left(\log |A^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P, \epsilon)| - \|\vec{m}^{(n)}\|_1 \log n \right) \leq J_h(P). \tag{3.61}$$

In order to do this, fix a finite subset $\Delta \subset \bar{\mathcal{T}}_*^h$ and define $\mathcal{F}(\Delta) \subset \Xi \times \bar{\mathcal{T}}_*^{h-1}$ to be the set of $T[o, v]_{h-1}$ and $T[v, o]_{h-1}$ for $[T, o] \in \Delta$ and $v \sim_T o$. Since Δ is finite, $\mathcal{F}(\Delta)$ is also finite and can be identified with the set of integers $\{1, \dots, L\}$ where $L = L(\Delta) := |\mathcal{F}(\Delta)|$. With this, define the color set $\mathcal{C}(\Delta) := \mathcal{F}(\Delta) \times \mathcal{F}(\Delta)$. Furthermore, let $\bar{\mathcal{F}}(\Delta) := \mathcal{F}(\Delta) \cup \{\star_x : x \in \Xi\}$, where $\star_x$ for $x \in \Xi$ are additional distinct elements not present in $\mathcal{F}(\Delta)$. Note that $\bar{\mathcal{F}}(\Delta)$ is finite, thus can be identified with the set of integers $\{1, \dots, \bar{L}\}$ where $\bar{L} = \bar{L}(\Delta) = L(\Delta) + |\Xi|$, where the first L elements represent $\mathcal{F}(\Delta)$. Finally, extend the color set $\mathcal{C}(\Delta)$ to $\bar{\mathcal{C}}(\Delta) := \bar{\mathcal{F}}(\Delta) \times \bar{\mathcal{F}}(\Delta)$.

Now, for a fixed Δ as above, given a simple marked graph G on the vertex set $[n]$, we construct a simple directed colored graph $\widetilde{G} \in \mathcal{G}(\bar{\mathcal{C}}(\Delta))$ on the same vertex set $[n]$, with color set $\bar{\mathcal{C}}(\Delta)$, as follows. For each edge between vertices u and v in G, if $\varphi^h_G(u, v) \in \mathcal{C}(\Delta)$, we place an edge in $\widetilde{G}$ directed from u towards v with color $\varphi^h_G(u, v)$, and another edge directed from v towards u with color $\varphi^h_G(v, u)$. Otherwise, if $\varphi^h_G(u, v) \notin \mathcal{C}(\Delta)$, we place an edge in $\widetilde{G}$ directed from u towards v with color $(\star_x, \star_{x'})$ and an edge directed from v towards u with color $(\star_{x'}, \star_x)$, where $x = \xi_G(v, u)$ and $x' = \xi_G(u, v)$. Let $\vec{D}^{\widetilde{G}}$ be the colored degree sequence of $\widetilde{G}$. More precisely, for a vertex v, $D^{\widetilde{G}}(v) \in \mathcal{M}_{\bar{L}(\Delta)}$ has the following form. For $c \in \mathcal{C}(\Delta)$, $D^{\widetilde{G}}_c(v)$ is the number of vertices $w \sim_G v$ such that $\varphi^h_G(v, w) = c$. Moreover, for $x, x' \in \Xi$, $D^{\widetilde{G}}_{(\star_x, \star_{x'})}(v)$ is the number of vertices $w \sim_G v$ such that $\varphi^h_G(v, w) \notin \mathcal{C}(\Delta)$, $\xi_G(w, v) = x$ and $\xi_G(v, w) = x'$. Note that, for $x \in \Xi$ and $t \in \mathcal{F}(\Delta)$, we have $D^{\widetilde{G}}_{(\star_x, t)}(v) = D^{\widetilde{G}}_{(t, \star_x)}(v) = 0$.

With an abuse of notation, for a marked rooted graph (G, o) on a finite or countable vertex set, and $c \in \bar{\mathcal{C}}(\Delta)$ of the form $c = (\star_x, \star_{x'})$, $x, x' \in \Xi$, we define

$$E_h(\star_x, \star_{x'})(G, o) = |\{v \sim_G o : \varphi^h_G(o, v) \notin \mathcal{C}(\Delta), \xi_G(v, o) = x, \xi_G(o, v) = x'\}|, \tag{3.62}$$

and, for $x \in \Xi$, we define

$$E_h(\star_x, t)(G, o) = E_h(t, \star_x)(G, o) = 0, \qquad \forall t \in \mathcal{F}(\Delta). \tag{3.63}$$

With this convention, define the map $F_\Delta : \bar{\mathcal{G}}_*^h \to \Theta \times \mathcal{M}_{\bar{L}(\Delta)}$, such that for $[G,o] \in \bar{\mathcal{G}}_*^h$, $F([G,o]) := (\theta, D)$ where $\theta = \tau_G(o)$ is the mark at the root in G, and for $c \in \bar{\mathcal{C}}(\Delta)$, $D_c = E_h(c)(G,o)$. Moreover, let $\bar{P}^\Delta \in \mathcal{P}(\Theta \times \mathcal{M}_{\bar{L}(\Delta)})$ be the law of $F_\Delta([T,o])$ when $[T,o] \sim P$. Note that if G is a marked graph on the vertex set $[n]$, with the directed colored graph $\widetilde{G}$ defined above then for all vertices v in G, we have $(\tau_G(v), D^{\widetilde{G}}(v)) = F_\Delta([G,v]_h)$. Therefore, if $G \in A^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\epsilon)$, since $d_{\text{TV}}(U(G)_h, P) < \epsilon$, we have

$$d_{\text{TV}}\left(\frac{1}{n}\sum_{i=1}^n \delta_{(\tau_G(v), D^{\widetilde{G}}(v))}, \bar{P}^\Delta\right) < \epsilon. \tag{3.64}$$

Let $B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$ be the set of pairs of sequences $(\vec{\beta}, \vec{D})$, $\vec{\beta} = (\beta(i) : 1 \le i \le n)$ and $\vec{D} = (D(i) : 1 \le i \le n) \in \mathcal{D}_n$ where, for $1 \le i \le n$, $\beta(i) \in \Theta$, $D(i) \in \mathcal{M}_{\bar{L}(\Delta)}$ are such that with

$$R(\vec{\beta}, \vec{D}) := \frac{1}{n}\sum_{i=1}^n \delta_{(\beta(i), D(i))}, \tag{3.65}$$

we have

$$d_{\text{TV}}(R(\vec{\beta}, \vec{D}), \bar{P}^\Delta) < \epsilon, \tag{3.66a}$$

$$\sum_{i=1}^n \sum_{c \in \bar{\mathcal{C}}(\Delta)} D_c(i) = 2\|\vec{m}^{(n)}\|_1, \tag{3.66b}$$

$$\sum_{i=1}^n \mathbb{1}\left[\beta(i) = \theta\right] = u^{(n)}(\theta), \quad \forall \theta \in \Theta, \tag{3.66c}$$

$$D_{(\star_x, t)}(v) = D_{(t, \star_x)}(v) = 0, \quad \forall x \in \Xi, t \in \mathcal{F}(\Delta), i \in [n]. \tag{3.66d}$$

Let $\vec{\tau}_G$ denote $(\tau_G(v) : v \in V(G))$. Note that, from (3.64), for $G \in A^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\epsilon)$ we have $(\vec{\tau}_G, \vec{D}^{\widetilde{G}}) \in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$. Now, we claim that for $(\vec{\beta}, \vec{D}) \in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$ and a colored directed graph $H \in \mathcal{G}(\vec{D}, 2)$, there is at most one marked graph $G \in A^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\epsilon)$ such that $\vec{\tau}_G = \vec{\beta}$ and $\widetilde{G} = H$. The reason is that, to start with, the condition $\vec{\tau}_G = \vec{\beta}$ uniquely determines vertex marks for G. Moreover, since $\widetilde{G} = H$, vertices v and w are adjacent in G iff they are adjacent in H. Let v and w be adjacent vertices in H with $c \in \bar{\mathcal{C}}(\Delta)$ being the color of the edge directed from v towards w. Note that, by the definition of the set $B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$, c is either of the form (t,t') where $t,t' \in \mathcal{F}(\Delta)$, or $c = (\star_{x_1}, \star_{x_2})$ for some $x_1, x_2 \in \Xi$. Since $\widetilde{G} = H$, in the former case, we have $\xi_G(w,v) = t[m]$ and $\xi_G(v,w) = t'[m]$, while, in the latter case, we have $\xi_G(w,v) = x_1$ and $\xi_G(v,w) = x_2$. This implies that there is at most one marked graph satisfying $\vec{\tau}_G = \vec{\beta}$ and $\widetilde{G} = H$. Consequently, we have

$$|A^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\epsilon)| \le |B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)| \max_{(\vec{\beta},\vec{D}) \in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} |\mathcal{G}(\vec{D}, 2)|. \tag{3.67}$$

Now, we find bounds for the two terms on the right hand side of (3.67).

Bounding $|B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)|$: we claim that

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log|B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)| \le H(P). \tag{3.68}$$

Note that P and $\bar{P}^\Delta$ are not necessarily finitely supported, so this is not a direct consequence of (3.66a) and requires some work. In order establish the claim, fix a finite subset $X \subset \Theta \times \mathcal{M}_{\bar{L}(\Delta)}$ of the form $X = \{(\beta^1, D^1), \dots, (\beta^{|X|}, D^{|X|})\}$. With this, for $1 \le j \le |X|$, let $\bar{p}_j := \bar{P}^\Delta(\beta^j, D^j)$. Furthermore, define $\bar{p}_0 := 1 - \sum_{j=1}^{|X|}\bar{p}_j = 1 - \bar{P}^\Delta(X)$. Now, fix $(\vec{\beta}, \vec{D}) \in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$ and let $I_j := \{i \in [n] : (\beta(i), D(i)) = (\beta^j, D^j)\}$ for $1 \le j \le |X|$. Additionally, let $I_0 := \{i \in [n] : (\beta(i), D(i)) \notin X\}$. Moreover, define $a_j := |I_j|/n$ for $0 \le j \le |X|$. Then, because of (3.66a), we have $|a_j - \bar{p}_j| < \epsilon$ for $0 \le j \le |X|$. Also, due to (3.66b), we have

$$\begin{aligned}
\sum_{i\in I_0}\sum_{c\in\bar{\mathcal{C}}(\Delta)} D_c(i) &= 2\|\vec{m}^{(n)}\|_1 - \sum_{j=1}^{|X|}\sum_{i\in I_j}\sum_{c\in\bar{\mathcal{C}}(\Delta)} D_c(i)\\
&= 2\|\vec{m}^{(n)}\|_1 - \sum_{j=1}^{|X|} na_j \sum_{c\in\bar{\mathcal{C}}(\Delta)} D^j_c\\
&\le 2\|\vec{m}^{(n)}\|_1 - n\sum_{j=1}^{|X|}\bar{p}_j\sum_{c\in\bar{\mathcal{C}}(\Delta)} D^j_c + n\epsilon\sum_{j=1}^{|X|}\sum_{c\in\bar{\mathcal{C}}(\Delta)} D^j_c\\
&= 2\|\vec{m}^{(n)}\|_1 - n\bar{d}(X) + n\epsilon\alpha(X),
\end{aligned} \tag{3.69}$$

where

$$\bar{d}(X) := \sum_{j=1}^{|X|}\bar{p}_j\sum_{c\in\bar{\mathcal{C}}(\Delta)} D^j_c = \mathbb{E}_{\bar{P}^\Delta}\left[\mathbb{1}\left[(\beta, D)\in X\right]\sum_{c\in\bar{\mathcal{C}}(\Delta)} D_c\right],$$

and

$$\alpha(X) := \sum_{j=1}^{|X|}\sum_{c\in\bar{\mathcal{C}}(\Delta)} D^j_c.$$

Motivated by this, in order to find an upper bound for $B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$, we may first count the number of choices for $I_0, \dots, I_{|X|}$, and then the number of pairs $(\vec{\beta}, \vec{D})$ consistent with each of them. For this, let $Y^{(n)}_\epsilon$ be the set of $(a_0, \dots, a_{|X|})$ such that $\sum_{j=0}^n a_j = 1$ and such that for $0 \le j \le |X|$ we have $na_j \in \mathbb{Z}_+$ and $|a_j - \bar{p}_j| < \epsilon$. We can see that $|Y^{(n)}_\epsilon| \le (2n\epsilon)^{1+|X|}$. Moreover, given $(a_0, \dots, a_{|X|})$, there are $\binom{n}{na_0 \dots na_{|X|}}$ many ways to chose a partition $I_0, \dots, I_{|X|}$ of $[n]$ such that $|I_j| = na_j$, $0 \le j \le |X|$. Fixing such a partition, for $i \in I_j$, $j \neq 0$, we must have $(\beta(i), D(i)) = (\beta^j, D^j)$. Hence, we only need to count the number of choices for $\{(\beta(i), D(i)) : i \in I_0\}$. Note that, there are at most $|\Theta|^{|I_0|} \leq |\Theta|^{n(\bar{p}_0+\epsilon)}$ many ways to choose $\beta(i)$ for $i \in I_0$. On the other hand, for $(D_c(i) : i \in I_0, c \in \bar{\mathcal{C}}(\Delta))$ there are $|I_0|\bar{\mathcal{C}}(\Delta) = na_0|\bar{\mathcal{C}}(\Delta)|$ many nonnegative integers satisfying (3.69). Hence, there are at most

$$\binom{2\|\vec{m}^{(n)}\|_1 - n\bar{d}(X) + n\epsilon\alpha(X) + na_0|\bar{\mathcal{C}}(\Delta)|}{na_0|\bar{\mathcal{C}}(\Delta)|}$$

many ways to choose $D(i)$ for $i \in I_0$. Putting all these together, we have

$$\begin{aligned}\log |B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P, \Delta, \epsilon)| \leq (1 + |X|) \log(2n\epsilon) + \max_{(a_0,\dots,a_{|X|}) \in Y^{(n)}_\epsilon} \log \binom{n}{na_0 \dots na_{|X|}} + n(\bar{p}_0 + \epsilon) \log |\Theta| \\ + \max_{(a_0,\dots,a_{|X|}) \in Y^{(n)}_\epsilon} \log \binom{2\|\vec{m}^{(n)}\|_1 - n\bar{d}(X) + n\epsilon\alpha(X) + na_0|\bar{\mathcal{C}}(\Delta)|}{na_0|\bar{\mathcal{C}}(\Delta)|}.\end{aligned} \tag{3.70}$$

Furthermore, using Stirling's approximation, one can show that for $(a_0, \dots, a_{|X|}) \in Y^{(n)}_\epsilon$, we have

$$\begin{aligned}\log \binom{n}{na_0 \dots na_{|X|}} &\leq 1 + \frac{1}{2} \log n - n \sum_{j=0}^{|X|} a_j \log a_j \\ &\leq 1 + \frac{1}{2} \log n - n \sum_{j=0}^{|X|} \bar{p}_j \log \bar{p}_j + n\eta_2(\epsilon),\end{aligned} \tag{3.71}$$

where, in the second inequality, $\eta_2(\epsilon) \to 0$ as $\epsilon \to 0$, and we have used the fact that $x \mapsto x \log x$ is uniformly continuous on $(0, 1]$ and also the assumption that $|a_j - \bar{p}_j| < \epsilon$ for $0 \leq j \leq |X|$. Note that

$$\begin{aligned}-\sum_{j=0}^{|X|} \bar{p}_j \log \bar{p}_j &= -\left(\sum_{(\beta,D) \in X} \bar{P}^\Delta(\beta, D) \log \bar{P}^\Delta(\beta, D)\right) - (1 - \bar{P}^\Delta(X)) \log(1 - \bar{P}^\Delta(X)) \\ &\leq H(\bar{P}^\Delta) \leq H(P),\end{aligned} \tag{3.72}$$

where the last inequality follows from the fact that $\bar{P}^\Delta$ is the pushforward of P under F_Δ. Putting (3.72) back in (3.71), we get

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \max_{(a_0,\dots,a_{|X|}) \in Y^{(n)}_\epsilon} \log \binom{n}{na_0 \dots na_{|X|}} \leq H(P). \tag{3.73}$$

On the other hand, using the general inequality $\log \binom{r}{s} \leq \log 2^r \leq r$, which holds for integers $r \geq s \geq 0$, together with $a_0 \leq \bar{p}_0 + \epsilon$, we realize that for all $(a_0, \dots, a_{|X|}) \in Y^{(n)}_\epsilon$, we have

$$\frac{1}{n} \log \binom{2\|\vec{m}^{(n)}\|_1 - n\bar{d}(X) + n\epsilon\alpha(X) + na_0|\bar{\mathcal{C}}(\Delta)|}{na_0|\bar{\mathcal{C}}(\Delta)|} \leq 2\frac{\|\vec{m}^{(n)}\|_1}{n} - \bar{d}(X) + \epsilon\alpha(X) + (\bar{p}_0 + \epsilon)|\bar{\mathcal{C}}(\Delta)|.$$

Since the sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are such that $(\vec{m}^{(n)}, \vec{u}^{(n)})$ is adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$, as $n \to \infty$ we have $\|\vec{m}^{(n)}\|_1/n \to d/2$, where $d = \deg(\mu)$ is the average degree at the root in μ. Therefore,

$$\begin{aligned}\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\max_{(a_0,\dots,a_{|X|})\in Y_\epsilon^{(n)}}\log\binom{2\|\vec{m}^{(n)}\|_1 - n\bar{d}(X) + n\epsilon\alpha(X) + na_0|\bar{\mathcal{C}}(\Delta)|}{na_0|\bar{\mathcal{C}}(\Delta)|}\\ \le d - \bar{d}(X) + (1-\bar{P}^\Delta(X))|\bar{\mathcal{C}}(\Delta)|. \qquad (3.74)\end{aligned}$$

Using (3.73) and (3.74) in (3.70), we get

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log|B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)| \le H(P) + d - \bar{d}(X) + (1-\bar{P}^\Delta(X))(\log|\Theta| + |\bar{\mathcal{C}}(\Delta)|). \qquad (3.75)$$

Since this holds for any finite $X \subset \Theta \times \mathcal{M}_{\bar{L}(\Delta)}$, we may take a nested sequence X_k converging to $\Theta \times \mathcal{M}_{\bar{L}(\Delta)+1}$ so that $\bar{P}^\Delta(X_k) \to 1$ and

$$\bar{d}(X_k) = \mathbb{E}_{\bar{P}^\Delta}\left[\mathbb{1}\left[(\beta, D) \in X_k\right]\sum_{c\in\bar{\mathcal{C}}(\Delta)} D_c\right] \to \mathbb{E}_{\bar{P}^\Delta}\left[\sum_{c\in\bar{\mathcal{C}}(\Delta)} D_c\right] = \mathbb{E}_P\left[\deg_T(o)\right] = \deg(\mu) = d.$$

Using this in (3.75), we arrive at (3.68).

Bounding $|\mathcal{G}(\vec{D}, 2)|$: Now, we find an upper bound for the second term in the right hand side of (3.67). We claim that for $(\vec{\beta}, \vec{D}) \in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P, \Delta, \epsilon)$, we have

$$|\mathcal{G}(\vec{D}, 2)| \le \frac{\prod_{c\in\bar{\mathcal{C}}(\Delta)_<} S_c^{(n)}(\vec{D})! \prod_{c\in\bar{\mathcal{C}}(\Delta)_=}(S_c^{(n)}(\vec{D}) - 1)!!}{\prod_{c\in\bar{\mathcal{C}}(\Delta)}\prod_{v=1}^n D_c(v)!}, \qquad (3.76)$$

where $S_c^{(n)}(\vec{D}) = \sum_{v=1}^n D_c(v)$. In order to show this, we take a simple directed colored graph $H \in \mathcal{G}(\vec{D}, 2)$ and construct $N := \prod_{c\in\bar{\mathcal{C}}(\Delta)}\prod_{v=1}^n D_c(v)!$ many configurations in the space Σ. Recall from Section 3.3.2 that Σ is the set of all possible matchings of half edges. Note that by the definition of the set $B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P, \Delta, \epsilon)$, we have $\vec{D} \in \mathcal{D}_n$ and Σ is well-defined. For $v \in [n]$ and $c \in \mathcal{C}(\Delta)$, we consider all the possible numberings of $D_c(v)$ many edges going out of the vertex v, which is $D_c(v)!$. It is easy to see that since H does not have loops and there is at most one directed edge between each two pair of vertices, the N many objects constructed in this way are all distinct members of Σ. Also, for distinct simple directed colored graphs $H \neq H' \in \mathcal{G}(\vec{D}, 2)$, the N many objects corresponding to H are indeed distinct from the N many objects corresponding to H'. Hence, $|\mathcal{G}(\vec{D}, 2)|N \le |\Sigma|$. But $|\Sigma|$ is precisely $\prod_{c\in\bar{\mathcal{C}}(\Delta)_<} S_c^{(n)}(\vec{D})! \prod_{c\in\bar{\mathcal{C}}(\Delta)_=}(S_c^{(n)}(\vec{D}) - 1)!!$. This establishes (3.76). Applying

Stirling's approximation to (3.76), in a manner similar to what we did in (3.39), we get

$$\begin{aligned}\log|\mathcal{G}(\vec{D},2)| &\leq \frac{1}{2}\sum_{c\in\bar{\mathcal{C}}(\Delta)}\left(S_c^{(n)}(\vec{D})\log S_c^{(n)}(\vec{D}) - S_c^{(n)}(\vec{D})\right) - \sum_{v=1}^{n}\sum_{c\in\bar{\mathcal{C}}(\Delta)}\log D_c(v)! + o(n) \\ &= \frac{n}{2}\sum_{c\in\bar{\mathcal{C}}(\Delta)}\left(\frac{S_c^{(n)}(\vec{D})}{n}\log\frac{S_c^{(n)}(\vec{D})}{n} - \frac{S_c^{(n)}(\vec{D})}{n}\right) - \sum_{v=1}^{n}\sum_{c\in\bar{\mathcal{C}}(\Delta)}\log D_c(v)! \\ &\quad + \frac{1}{2}\sum_{c\in\bar{\mathcal{C}}(\Delta)} S_c^{(n)}(\vec{D})\log n + o(n) \\ &= \frac{n}{2}\sum_{c\in\bar{\mathcal{C}}(\Delta)}\left(\frac{S_c^{(n)}(\vec{D})}{n}\log\frac{S_c^{(n)}(\vec{D})}{n} - \frac{S_c^{(n)}(\vec{D})}{n}\right) - \sum_{v=1}^{n}\sum_{c\in\bar{\mathcal{C}}(\Delta)}\log D_c(v)! \\ &\quad + \|\vec{m}^{(n)}\|_1\log n + o(n),\end{aligned} \tag{3.77}$$

where the $o(n)$ term does not depend on $\vec{D}$. Now, we claim that, for $c \in \bar{\mathcal{C}}(\Delta)$,

$$\limsup_{\epsilon\to 0}\limsup_{n\to\infty}\max_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} |S_c^{(n)}(\vec{D})/n - \mathbb{E}_{\bar{P}^\Delta}[D_c]| = 0. \tag{3.78}$$

Note that for $c \in \bar{\mathcal{C}}(\Delta)$, we have $\mathbb{E}_{\bar{P}^\Delta}[D_c] \leq d = \deg(\mu)$, which is finite. On the other hand, we have $S_c^{(n)}(\vec{D})/n = \mathbb{E}_{R(\vec{\beta},\vec{D})}[D_c]$. Therefore, condition (3.66a) implies that for any integer $k > 0$ and any $(\vec{\beta},\vec{D}) \in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$, we have

$$S_c^{(n)}(\vec{D})/n \geq \mathbb{E}_{R(\vec{\beta},\vec{D})}[D_c \wedge k] \geq \mathbb{E}_{\bar{P}^\Delta}[D_c \wedge k] - 2k\epsilon.$$

Taking the lim inf as $n \to \infty$ and then sending ϵ to zero, we get

$$\liminf_{\epsilon\to 0}\liminf_{n\to\infty}\min_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} \frac{S_c^{(n)}(\vec{D})}{n} \geq \mathbb{E}_{\bar{P}^\Delta}[D_c \wedge k].$$

Furthermore, sending $k \to \infty$, we get

$$\liminf_{\epsilon\to 0}\liminf_{n\to\infty}\min_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} \frac{S_c^{(n)}(\vec{D})}{n} \geq \mathbb{E}_{\bar{P}^\Delta}[D_c]. \tag{3.79}$$

Now, we show that a matching upper bound exists. To do this, note that due to (3.66b), for $c \in \bar{\mathcal{C}}(\Delta)$, we have $S_c^{(n)}(\vec{D}) = 2\|\vec{m}^{(n)}\|_1 - \sum_{\substack{c'\in\bar{\mathcal{C}}(\Delta)\\ c'\neq c}} S_{c'}^{(n)}(\vec{D})$. Using $2\|\vec{m}^{(n)}\|_1/n \to d =$

$\deg(\mu) \in (0,\infty)$ and (3.79), since $\bar{\mathcal{C}}(\Delta)$ is finite, we get

$$
\begin{aligned}
\limsup_{\epsilon\to 0}\limsup_{n\to\infty} \max_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} \frac{S^{(n)}_c(\vec{D})}{n} &\le d - \sum_{\substack{c'\in\bar{\mathcal{C}}(\Delta)\\ c'\neq c}} \liminf_{\epsilon\to 0}\liminf_{n\to\infty} \min_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} \frac{S^{(n)}_{c'}(\vec{D})}{n}\\
&\le d - \sum_{\substack{c'\in\bar{\mathcal{C}}(\Delta)\\ c'\neq c}} \mathbb{E}_{\bar{P}^\Delta}\left[D_{c'}\right]\\
&= \sum_{c''\in\bar{\mathcal{C}}(\Delta)} \mathbb{E}_{\bar{P}^\Delta}\left[D_{c''}\right] - \sum_{\substack{c'\in\bar{\mathcal{C}}(\Delta)\\ c'\neq c}} \mathbb{E}_{\bar{P}^\Delta}\left[D_{c'}\right]\\
&= \mathbb{E}_{\bar{P}^\Delta}\left[D_c\right].
\end{aligned}
$$

This together with (3.79) completes the proof of (3.78).

On the other hand, observe that for $(\vec{\beta},\vec{D}) \in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)$ and $c \in \bar{\mathcal{C}}(\Delta)$, $\frac{1}{n}\sum_{v=1}^n \log D_c(v)! = \mathbb{E}_{R(\vec{\beta},\vec{D})}\left[\log D_c!\right]$. Therefore, a similar truncation argument as in (3.79) implies that

$$
\liminf_{\epsilon\to 0}\liminf_{n\to\infty} \min_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} \frac{1}{n}\sum_{v=1}^n \log D_c(v)! \ge \mathbb{E}_{\bar{P}^\Delta}\left[\log D_c!\right]. \tag{3.80}
$$

Note that $\log D_c! \ge 0$, hence $\mathbb{E}_{\bar{P}^\Delta}\left[\log D_c!\right]$ is well-defined, although it can be ∞. Also, $\bar{\mathcal{C}}(\Delta)$ is finite. Therefore, using (3.80) together with (3.78) in (3.77) and simplifying, we get

$$
\begin{aligned}
&\limsup_{\epsilon\to 0}\limsup_{n\to\infty} \max_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} \frac{1}{n}\left(\log|\mathcal{G}(\vec{D},2) - \|\vec{m}^{(n)}\|_1 \log n\right)\\
&\le \frac{1}{2}\sum_{c\in\bar{\mathcal{C}}(\Delta)} \left(\mathbb{E}_{\bar{P}^\Delta}\left[D_c\right]\log \mathbb{E}_{\bar{P}^\Delta}\left[D_c\right] - \mathbb{E}_{\bar{P}^\Delta}\left[D_c\right]\right) - \sum_{c\in\bar{\mathcal{C}}(\Delta)} \mathbb{E}_{\bar{P}^\Delta}\left[\log D_c!\right]\\
&= -s(d) + \frac{d}{2}\sum_{c\in\bar{\mathcal{C}}(\Delta)} \frac{\mathbb{E}_{\bar{P}^\Delta}\left[D_c\right]}{d}\log\frac{\mathbb{E}_{\bar{P}^\Delta}\left[D_c\right]}{d} - \sum_{c\in\bar{\mathcal{C}}(\Delta)} \mathbb{E}_{\bar{P}^\Delta}\left[\log D_c!\right].
\end{aligned}
$$

Note that, for each $c \in \bar{\mathcal{C}}(\Delta)$, $0 \le \mathbb{E}_{\bar{P}^\Delta}\left[D_c\right] \le d < \infty$, hence each term in the first summation is nonpositive and finite. Also, $\mathbb{E}_{\bar{P}^\Delta}\left[\log D_c!\right] \ge 0$ for $c \in \bar{\mathcal{C}}(\Delta)$. As a result, the bound on the right hand side is well-defined, although it can be $-\infty$. Also, since each term in the first summation is nonpositive while each term in the second summation is nonnegative, we may restrict both the summations to $\mathcal{C}(\Delta) \subset \bar{\mathcal{C}}(\Delta)$ to find an upper bound for the right hand side. But for $c \in \mathcal{C}(\Delta)$, we have $\mathbb{E}_{\bar{P}^\Delta}\left[D_c\right] = e_P(c)$ and $\mathbb{E}_{\bar{P}^\Delta}\left[\log D_c!\right] = \mathbb{E}_P\left[\log E_h(c)!\right]$, which yields

$$
\begin{aligned}
&\limsup_{\epsilon\to 0}\limsup_{n\to\infty} \max_{(\vec{\beta},\vec{D})\in B^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(P,\Delta,\epsilon)} \frac{1}{n}\left(\log|\mathcal{G}(\vec{D},2)| - \|\vec{m}^{(n)}\|_1 \log n\right)\\
&\qquad\le -s(d) + \frac{d}{2}\sum_{c\in\mathcal{C}(\Delta)} \pi_P(c)\log \pi_P(c) - \sum_{c\in\mathcal{C}(\Delta)} \mathbb{E}_P\left[\log E_h(c)!\right].
\end{aligned} \tag{3.81}
$$

Again, note that the terms in the first summation are finite and nonpositive, while the terms in the second summation are nonnegative, but possibly $+\infty$. Thereby, the above bound is well-defined, although it can be $-\infty$.

By assumption, we have $H(P) < \infty$. Therefore, we can put the bounds in (3.81) and (3.68) back in (3.67) to get

$$\begin{aligned} &\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \left(\log |A^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(P, \epsilon)| - \|\vec{m}^{(n)}\|_1 \log n \right) \\ &\quad \le -s(d) + H(P) - \frac{d}{2} \sum_{c \in \mathcal{C}(\Delta)} \pi_P(c) \log \frac{1}{\pi_P(c)} - \sum_{c \in \mathcal{C}(\Delta)} \mathbb{E}_P \left[\log E_h(c)! \right]. \end{aligned} \tag{3.82}$$

Note that this holds for any finite $\Delta \subset \bar{\mathcal{T}}_*^h$, and that $\pi_P(c) \log \frac{1}{\pi_P(c)} \ge 0$ and $\mathbb{E}_P \left[\log E_h(c)! \right] \ge 0$ for all $c \in (\Xi \times \bar{\mathcal{T}}_*^{h-1}) \times (\Xi \times \bar{\mathcal{T}}_*^{h-1})$. Interpreting the summations on the right hand side of (3.82) as integrals, restricted to $\mathcal{C}(\Delta)$, with respect to the uniform measure on $(\Xi \times \bar{\mathcal{T}}_*^{h-1}) \times (\Xi \times \bar{\mathcal{T}}_*^{h-1})$, by sending Δ to $\bar{\mathcal{T}}_*^h$ and using the monotone convergence theorem, we arrive at (3.61) which completes the proof. □

3.4.4 Proof of Proposition 3.5

In this section, we prove the upper bound result of Proposition 3.5.

Proof of Proposition 3.5. Let $P := \mu_1 \in \mathcal{P}(\bar{\mathcal{T}}_*^1)$ be the distribution of the depth–1 neighborhood of the root in μ. Borrowing the idea in the proof of Corollary 2.1, note that each rooted tree equivalence class $[T, o] \in \bar{\mathcal{T}}_*^1$ is uniquely determined by knowing the integers

$$N_{x,x'}^{\theta,\theta'}(T, o) := |\{ v \sim_T o : \xi_T(v, o) = x, \tau_T(o) = \theta, \xi_T(o, v) = x', \tau_T(v) = \theta' \}|, \tag{3.83}$$

for each $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$. Now, for $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$, we have $\mathbb{E}_P \left[N_{x,x'}^{\theta,\theta'}(T, o) \right] \le \mathbb{E}_P \left[\deg_T(o) \right] < \infty$. Consequently, when $[T, o] \sim P$, the entropy of the random variable $N_{x,x'}^{\theta,\theta'}(T, o)$ is finite for all $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$. Therefore, since Ξ and Θ are finite sets, we conclude that $H(P) < \infty$. Hence, using Proposition 3.4 for $h = 1$, we have

$$\overline{\Sigma}_{\vec{\deg}(\mu), \vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)}, \vec{u}^{(n)})} \le -s(d) + H(P) - \frac{d}{2} H(\pi_P) - \sum_{t, t' \in \Xi \times \bar{\mathcal{T}}_*^0} \mathbb{E}_P \left[\log E_1(t, t')! \right]. \tag{3.84}$$

Now we show that there exist t and t' in $\Xi \times \bar{\mathcal{T}}_*^0$ such that $\mathbb{E}_P \left[\log E_1(t, t')! \right] = \infty$. Since every element of $\bar{\mathcal{T}}_*^0$ is a marked isolated vertex, $\bar{\mathcal{T}}_*^0$ can be identified with Θ. With an abuse of notation, we may therefore write $t, t' \in \Xi \times \bar{\mathcal{T}}_*^0$ as $t = (x, \theta)$ and $t' = (x', \theta')$ respectively, where $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$. With this, for $[T, o] \in \bar{\mathcal{T}}_*$, we have $E_h(t, t')(T, o) = N_{x,x'}^{\theta,\theta'}(T, o)$. Therefore, from (3.84), it suffices to prove that $\mathbb{E}_P \left[\log N_{x,x'}^{\theta,\theta'}(T, o)! \right] = \infty$ for some $x, x' \in \Xi$, $\theta, \theta' \in \Theta$.

We prove this by contradiction. Assume that $\mathbb{E}_P\left[\log N_{x,x'}^{\theta,\theta'}(T,o)!\right] < \infty$ for all $x, x' \in \Xi$, $\theta, \theta' \in \Theta$. Using Stirling's approximation, for $k \geq 0$, we have $\log k! \geq k \log k - k$, where $0 \log 0$ is interpreted as 0. Therefore, for $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$, we have

$$\infty > \mathbb{E}_P\left[\log N_{x,x'}^{\theta,\theta'}(T,o)!\right] \geq \mathbb{E}_P\left[N_{x,x'}^{\theta,\theta'}(T,o) \log N_{x,x'}^{\theta,\theta'}(T,o)\right] - \mathbb{E}_P\left[N_{x,x'}^{\theta,\theta'}(T,o)\right]. \quad (3.85)$$

On the other hand, $\deg_T(o) = \sum_{\substack{x,x' \in \Xi \\ \theta,\theta' \in \Theta}} N_{x,x'}^{\theta,\theta'}(T,o)$ for all $[T,o] \in \bar{\mathcal{T}}_*$. Also, we have $\mathbb{E}_P[\deg_T(o)] = \deg(\mu) < \infty$. Hence $\mathbb{E}_P\left[N_{x,x'}^{\theta,\theta'}(T,o)\right] < \infty$ for all $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$. Using this in (3.85), we realize that

$$\mathbb{E}_P\left[N_{x,x'}^{\theta,\theta'}(T,o) \log N_{x,x'}^{\theta,\theta'}(T,o)\right] < \infty \qquad \forall x, x' \in \Xi \text{ and } \theta, \theta' \in \Theta \quad (3.86)$$

Moreover, for $[T,o] \in \bar{\mathcal{T}}_*$, using $\deg_T(o) = \sum_{\substack{x,x' \in \Xi \\ \theta,\theta' \in \Theta}} N_{x,x'}^{\theta,\theta'}(T,o)$ and the convexity of $x \mapsto x \log x$, we have

$$\frac{\deg_T(o)}{|\Xi|^2|\Theta|^2} \log \frac{\deg_T(o)}{|\Xi|^2|\Theta|^2} \leq \frac{1}{|\Xi|^2|\Theta|^2} \sum_{\substack{x,x' \in \Xi \\ \theta,\theta' \in \Theta}} N_{x,x'}^{\theta,\theta'}(T,o) \log N_{x,x'}^{\theta,\theta'}(T,o),$$

where as usual, we interpret $0 \log 0$ as 0. Taking the expectation with respect to P on both sides we realize that $\mathbb{E}_P[\deg_T(o) \log \deg_T(o)] < \infty$, which is a contradiction. Hence, there must exist $x, x' \in \Xi$ and $\theta, \theta' \in \Theta$ such that $\mathbb{E}_P\left[\log N_{x,x'}^{\theta,\theta'}!\right] = \infty$. Finally, using this in (3.84) implies $\overline{\Sigma}_{\vec{\deg}(\mu),\vec{\Pi}(\mu)}(\mu)|_{(\vec{m}^{(n)},\vec{u}^{(n)})} = -\infty$ and completes the proof. □

3.5 Conclusion

In this chapter, we introduced the marked BC entropy, which is a notion of entropy for probability distributions on the space of marked rooted graphs, and discussed its properties. This is a generalization of the notion of entropy defined in [BC15] for the marked regime. We saw that the marked BC entropy is only interesting for probability distributions which are unimodular and are supported on marked rooted trees. As we will see in Chapters 4 and 5, the marked BC entropy can be considered as a natural counterpart of the Shannon entropy rate in the context of local weak convergence.

Part II

Compression of Graphical Data

Chapter 4

Universal Lossless Compression of Graphical Data

The problem of graph compression has drawn a lot of attention in different fields. There have been two main type of work in the literature under the subject of graph compression. Some authors assume that the graphical data is generated from a certain statistical model and the encoder aims to achieve the entropy of the input distribution. For instance, Choi and Szpankowski studied the structural entropy of the Erdős–Rényi model, i.e. the entropy associated to the isomorphism class of such graphs [CS12]. Moreover, they proposed a compression scheme which asymptotically achieves the structural entropy up to the first two terms. Aldous and Ross studied the asymptotics of the entropy of several models of random graphs, including the sparse Erdős–Rényi ensemble [AR14]. Abbe studied the asymptotic behavior of the entropy of stochastic block models, and discussed the optimal compression rate for such models up to the first order term [Abb16]. They also considered the case where vertices in a stochastic block model can carry data which is conditionally independent given their community membership. Łuczak et al. studied the asymptotics of the entropy associated to the preferential attachment model, both for labeled and unlabeled regimes, and built upon their analysis to design optimal compression schemes [ŁMS19]. Turowski et al. studied the information content of the duplication model [TMS20]. They analyzed the asymptotic behavior of the entropy of such models for both the labeled and unlabeled regimes, and designed compression algorithms to achieve their entropy bounds.

A second line of research offers an alternative approach to compress specific types of graphical data, such as web graphs [BBH+98, SMHM99, BKM+00, BV04], social networks [CKL+09, MP10, BRSV11, Mas12], or biological networks [DWvW12, ADK12, KK14, SSA+16, HPP16]. These works usually take advantage of some properties specific to a specific data source, where such properties are often inferred through observing real-world data samples. For instance, the web graph framework of [BV04] employ the *locality* and *similarity* properties existing in web graphs to design efficient compression algorithm tailored for such data. The *locality* property refers to the fact that a web page usually refers to other web pages whose URLs have a long prefix in common, while the *similarity* property refers to the fact that if we sort web pages based on the lexicographic order of their URLs, web pages that are close to each other tend to have many successors in common. Therefore, techniques such as *reference encoding* and *gap encoding* are useful in compressing web graphs. Due to the nature of this approach, results in this category of work usually do not come with information theoretic guarantees of optimality. The reader is referred to [BH18] for a survey on graph compression methods.

The key property distinguishing our approach from the existing ones is *universality*. More precisely, in this chapter, we introduce a compression scheme which roughly speaking has the property that if a sequence of marked graphs is given to the encoder which is converging in the local weak sense of Chapter 2, the normalized codeword length associated to this sequence does not asymptotically exceed the marked BC entropy (as was defined in Chapter 3) associated to the limit. This scheme is universal in the sense that the above condition holds without a priori knowledge of the local weak limit of the sequence. In addition to this, recalling from Chapter 3, the notion of marked BC entropy captures the per vertex growth rate of the size of the typical graphs after carefully separating out the leading term. This is while the existing literature usually consider the random graph ensemble entropy up to only the leading term. Finally, since we allow for marked graphs instead of simple unmarked graphs, our framework is capable of modeling information associated to vertices and edges in the graph on top of the connectivity structure. This in particular is highly advantageous from a practical point of view.

This chapter is organized as follows. Then in Section 4.1, we precisely formulate the problem and state the main results. In Section 4.2, we introduce our universal compression scheme and provide the proof of its optimality. Finally, we conclude the chapter in Section 4.3.

To close this discussion we introduce some notation that will be needed when we develop our compression algorithm for graphical data. For a locally finite graph G and integer Δ, let G^Δ be the graph with the same vertex set that includes only those edges of G such that the degrees of both their endpoints are at most Δ (without reference to their marks). Another way to put this is that to arrive at G^Δ from G we remove all the edges in G that are connected to vertices with degree strictly bigger than Δ. This construction is used as a technical device in the proof of the main result, the main point being that the maximum degree in G^Δ is at most Δ.

4.1 Main Results

Recall that $\bar{\mathcal{G}}_n$ is the set of marked graphs on the vertex set $\{1, \dots, n\}$, with edge marks from Ξ and vertex marks from Θ. Our goal is to design a compression scheme, comprised of compression and decompression functions f_n and g_n for each n, such that f_n maps $\bar{\mathcal{G}}_n$ to $\{0,1\}^* - \emptyset$ and g_n maps $\{0,1\}^* - \emptyset$ to $\bar{\mathcal{G}}_n$, with the condition that $g_n \circ f_n(G) = G$ for all $G \in \bar{\mathcal{G}}_n$. Motivated by the notion of entropy introduced in Chapter 3, we want our compression scheme to be universally optimal in the following sense: if $\mu \in \mathcal{P}(\bar{\mathcal{T}}_*)$ is unimodular and $G^{(n)}$ is a sequence of marked graphs with local weak limit μ, then, with $\vec{m}^{(n)} := \vec{m}_{G^{(n)}}$, we have

$$\limsup_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \le \Sigma(\mu). \tag{4.1}$$

In Section 4.2, we design such a universally optimal compression scheme and prove its optimality. This is stated formally in the next theorem.

Theorem 4.1. *There is a compression scheme that is optimal in the above sense for all* $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ *such that* $\deg(\mu) \in (0,\infty)$.

We also prove the following converse theorem, which justifies the claim of optimality for compression schemes that satisfy the condition in (4.1).

Theorem 4.2. *Assume that a compression scheme* $\{f_n, g_n\}_{n=1}^\infty$ *is given. Fix some unimodular* $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ *such that* $\deg(\mu) \in (0,\infty)$ *and* $\Sigma(\mu) > -\infty$. *Moreover, fix a sequence* $\vec{m}^{(n)}$ *and* $\vec{u}^{(n)}$ *of edge mark count vectors and vertex mark count vectors respectively, such that* $(\vec{m}^{(n)}, \vec{u}^{(n)})$ *is adapted to* $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. *Then, there exists a sequence of positive real numbers* ϵ_n *going to zero, together with a sequence of independent graph–valued random variables* $\{G^{(n)}\}_{n=1}^\infty$ *defined on a joint probability space, with* $G^{(n)}$ *being uniform in* $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$, *such that with probability one*

$$\liminf_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \ge \Sigma(\mu).$$

Proof. First note that any marked graph on n vertices can be represented with $O(n^2)$ bits. Hence, without loss of generality, we may assume that, for some finite positive constant c, we have $\mathsf{nats}(f_n(G^{(n)})) \le cn^2$ for all $G^{(n)}$ on n vertices. Consequently, by adding a header of size $O(\log n^2) = O(\log n)$ to the beginning of each codeword in f_n, in order to describe its length, we can make f_n prefix–free. Thus, without loss of generality, we may assume that f_n is prefix–free for all n.

From the definition of $\underline{\Sigma}(\mu)$, one can find a sequence of positive numbers ϵ_n going to zero, such that

$$\underline{\Sigma}(\mu) = \liminf_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| - \|\vec{m}^{(n)}\|_1 \log n}{n}.$$

From Theorem 3.2, we have $\underline{\Sigma}(\mu) = \Sigma(\mu)$, and since $\Sigma(\mu) > -\infty$ by assumption, $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$ is nonempty once n is large enough. Using Kraft's inequality, we have

$$\sum_{G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)} e^{-\mathsf{nats}(f_n(G))} \le 1.$$

With $G^{(n)}$ being uniform in $\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$, the Markov inequality then implies that

$$\mathbb{P}\left(\mathsf{nats}(f_n(G^{(n)})) < \log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| - 2\log n\right) \le \frac{1}{n^2}.$$

From this, using the Borel-Cantelli lemma, we have $\mathsf{nats}(f_n(G^{(n)})) \geq \log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)| - 2\log n$ eventually. Therefore, with probability 1, we have

$$\liminf_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \geq \liminf_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n)| - \|\vec{m}^{(n)}\|_1 \log n}{n} = \underline{\Sigma}(\mu) = \Sigma(\mu),$$

which completes the proof. □

Remark 4.1. *Note that the existence of a sequence of graph–valued random variables $\{G^{(n)}\}_{n=1}^{\infty}$ for which with probability one the normalized codeword length is asymptotically no less than the BC entropy $\Sigma(\mu)$, as is implied by Theorem 4.2 above, in particular implies the existence of a sequence of deterministic graphs for which the normalized codeword length is asymptotically no less than the BC entropy. This draws a connection between our converse setup of Theorem 4.2 and the achievability result of Theorem 4.1 in which a sequence of deterministic graphs is considered which is convergent in the local weak limit sense.*

4.2 The Universal Compression Scheme

In this section, we propose our compression scheme. First, in Section 4.2.1, we introduce our compression scheme under certain assumptions. Then, in Section 4.2.2, we relax these assumptions.

4.2.1 A First–step Scheme

We first give an outline of the compression scheme, then illustrate it via an example, and finally formally describe it and prove its optimality. Fix two sequences of integers k_n and Δ_n as design parameters, which will be specified in Section 4.2.2. Given a marked graph $G^{(n)}$ on n vertices, with maximum degree no more than Δ_n, we first encode its depth–k_n empirical distribution, i.e. $U(G^{(n)})_{k_n}$ (defined in (2.2)). We do this by counting the number of times each marked rooted graph with depth at most k_n and maximum degree at most Δ_n appears in the graph $G^{(n)}$. Then, in the set of all graphs which result in these counts, we specify the target graph $G^{(n)}$. Figure 4.1 illustrates an example of this procedure. In this example, the marked graph on $n = 4$ vertices in Figure 4.1a is given and the design parameters $k_n = 1$ and $\Delta_n = 2$ are chosen. We then list all the rooted marked graphs with depth at most $k_n = 1$ and maximum degree at most $\Delta_n = 2$, and count the number of times each of these patterns appears in the graph, as depicted in Figure 4.1b. Finally, we consider all the graphs that would result in the same counts if we run this procedure on them (shown in Figure 4.1c for this example), and specify the input graph within this collection of graphs. In principle, this scheme is similar to the conventional universal coding for sequential data in which we

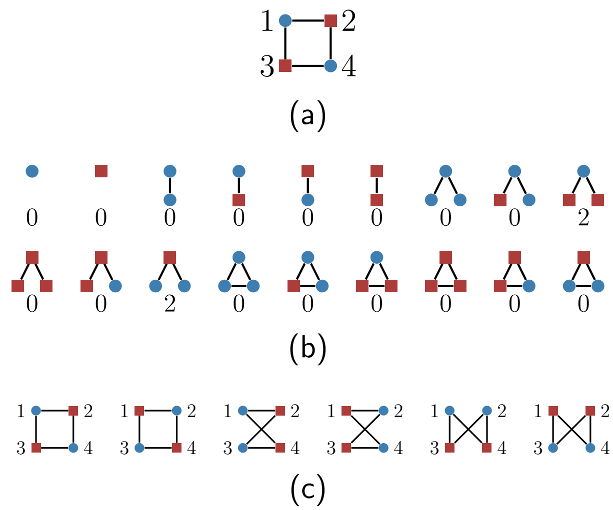

Figure 4.1: An example of encoding via the compression function associated to our compression scheme with the parameter $k = 1$ and the graph $G^{(4)}$ on $n = 4$ vertices, with vertex mark set $\Theta = \{\bullet, \blacksquare\}$ and edge mark set Ξ with cardinality 1, shown for (a) acting as the input. (b) depicts all members in the set $\mathcal{A}_{1,2}$ with the corresponding number of times each of them appears in the graph, i.e. the vector $(|\psi^{(4)}_{G^{(4)}}([G,o])|, [G,o] \in \mathcal{A}_{1,2})$ and (c) illustrates all the graphs with the same count vector, i.e. W_4.

first specify the type of a given sequence and then specify the sequence itself among all the sequences that have this type.

Before formally explaining the compression scheme, we need some definitions. For integers k and Δ, let $\mathcal{A}_{k,\Delta}$ be the set of equivalence classes of rooted marked graphs $[G,o] \in \bar{\mathcal{G}}_*$ with depth at most k and maximum degree at most Δ. Note that since k and Δ are finite and the mark sets are also finite sets, $\mathcal{A}_{k,\Delta}$ is a finite set.

For a marked graph $G^{(n)}$ on the vertex set $\{1, \dots, n\}$, with maximum degree at most Δ_n, and for $[G,o] \in \mathcal{A}_{k_n,\Delta_n}$, we denote the set $\{1 \le i \le n : [G^{(n)}, i]_{k_n} = [G,o]\}$ by $\psi^{(n)}_{G^{(n)}}([G,o])$. This is the set of vertices in $G^{(n)}$ with local structure $[G,o]$ up to depth k_n. Recall that $[G^{(n)}, i]_{k_n} = [G,o]$ means that $G^{(n)}$ rooted at i is isomorphic to (G,o) up to depth k_n. Note that when the maximum degree in $G^{(n)}$ is no more than Δ_n, $[G^{(n)}, i]_{k_n}$ is a member of $\mathcal{A}_{k_n,\Delta_n}$, for all $1 \le i \le n$. Therefore, the subsets $\psi^{(n)}_{G^{(n)}}([G,o])$, as $[G,o]$ ranges over $\mathcal{A}_{k_n,\Delta_n}$, form a partition of $\{1, \dots, n\}$.

We encode a marked graph $G^{(n)}$ with vertex set $\{1, \dots, n\}$ and maximum degree no more than Δ_n as follows:

1. Encode the vector $(|\psi^{(n)}_{G^{(n)}}([G,o])|, [G,o] \in \mathcal{A}_{k_n,\Delta_n})$. Since we have $|\psi^{(n)}_{G^{(n)}}([G,o])| \le n$ for all $[G,o] \in \mathcal{A}_{k_n,\Delta_n}$, this can be done with at most $|\mathcal{A}_{k_n,\Delta_n}|(1 + \lfloor \log_2 n \rfloor) \log 2$ nats.

2. Let W_n be the set of marked graphs G on the vertex set $\{1, \dots, n\}$ with degrees bounded

by Δ_n such that

$$|\psi_G^{(n)}([G', o'])| = |\psi_{G^{(n)}}^{(n)}([G', o'])|, \quad \forall\ [G', o'] \in \mathcal{A}_{k_n, \Delta_n}. \tag{4.2}$$

Specify $G^{(n)}$ among the elements of W_n by sending $(1 + \lfloor \log_2 |W_n| \rfloor) \log 2$ nats to the decoder.

See Figure 4.1 for an example of the running of this procedure.

Remark 4.2. *The vector* $(\psi_{G^{(n)}}^{(n)}([G, o]) : [G, o] \in \mathcal{A}_{k_n, \Delta_n})$ *is directly compressed in the above scheme; therefore, we are capable of making local queries in the compressed form without going through the decompression process. An example of such a query is "how many triangles exist in the graph?"*

Now we show the optimality of this compression scheme under an assumption on k_n and Δ_n that allows us to bound the size of the set $\mathcal{A}_{k_n, \Delta_n}$. Lemma 4.4, which is proved in Appendix C.1, shows that the assumptions made in Proposition 4.1 below are not vacuous. To avoid confusion, we use $\tilde{f}_n$ for the compression function in this section and f_n for that of Section 4.2.2 (which does not require any a priori assumed bound on the maximum degree of the graph on n vertices, $G^{(n)}$).

Proposition 4.1. *Fix the parameters* k_n *and* Δ_n *such that* $|\mathcal{A}_{k_n, \Delta_n}| = o(\frac{n}{\log n})$*, and* $k_n \to \infty$ *as* $n \to \infty$*. Assume that a sequence of marked graphs* $\{G^{(n)}\}_{n=1}^{\infty}$ *is given such that* $G^{(n)}$ *is on the vertex set* $\{1, \dots, n\}$*, the maximum degree in* $G^{(n)}$ *is no more than* Δ_n*, and* $\{G^{(n)}\}_{n=1}^{\infty}$ *converges in the local weak sense to a unimodular* $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ *as* $n \to \infty$*. Furthermore, assume that* $\deg(\mu) \in (0, \infty)$*. Then we have*

$$\limsup_{n \to \infty} \frac{\mathsf{nats}(\tilde{f}_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \leq \Sigma(\mu), \tag{4.3}$$

where $\vec{m}^{(n)} := \vec{m}_{G^{(n)}}$*.*

Before proving this proposition, we need the following tools. Lemma 4.1 is stated in a way which is stronger than what we need here, but this stronger form will prove useful later on. The proofs of Lemmas 4.1 and 4.2 are given in Appendix C.1.

Lemma 4.1. *Let* G *and* G' *be marked graphs on the vertex set* $\{1, \dots, n\}$*. For a permutation* $\pi \in \mathcal{S}_n$ *and an integer* $h \geq 0$*, let* L *be the number of vertices* $1 \leq i \leq n$ *such that* $(G, i)_h \equiv (G', \pi(i))_h$*. Then, we have*

$$d_{LP}(U(G), U(G')) \leq \max\left\{\frac{1}{1+h}, 1 - \frac{L}{n}\right\}.$$

Lemma 4.2. *Assume that a unimodular $\mu \in \mathcal{P}_u(\bar{\mathcal{T}}_*)$ is given such that $\deg(\mu) \in (0, \infty)$. Moreover, assume that sequences of edge and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ respectively are given such that*

$$\liminf_{n\to\infty} \frac{m^{(n)}(x,x')}{n} \geq \deg_{x,x'}(\mu), \qquad \forall x \neq x' \in \Xi; \tag{4.4a}$$

$$\liminf_{n\to\infty} \frac{m^{(n)}(x,x)}{n} \geq \frac{\deg_{x,x}(\mu)}{2}, \qquad \forall x \in \Xi; \tag{4.4b}$$

$$\lim_{n\to\infty} \frac{u^{(n)}(\theta)}{n} = \Pi_\theta(\mu), \qquad \forall \theta \in \Theta. \tag{4.4c}$$

Then, for any sequence ϵ_n of positive reals converging to zero, we have

$$\limsup_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| - \left\|\vec{m}^{(n)}\right\|_1 \log n}{n} \leq \Sigma(\mu). \tag{4.5}$$

Proof of Proposition 4.1. For our compression scheme, we have

$$\begin{aligned}\limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} &\leq \limsup_{n\to\infty} \frac{|\mathcal{A}_{k_n,\Delta_n}|(\log 2 + \log n)}{n} \\ &\quad + \frac{\log 2 + \log |W_n| - \|\vec{m}^{(n)}\|_1 \log n}{n} \\ &= \limsup_{n\to\infty} \frac{\log |W_n| - \|\vec{m}^{(n)}\|_1 \log n}{n},\end{aligned} \tag{4.6}$$

where the last equality employs the assumption $|\mathcal{A}_{k_n,\Delta_n}| = o(n/\log n)$. We now show that

$$W_n \subseteq \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}\left(\mu, \epsilon_n + \frac{1}{1+k_n}\right), \tag{4.7}$$

where $\epsilon_n := d_{\mathrm{LP}}(U(G^{(n)}),\mu)$ and $\vec{u}^{(n)} := \vec{u}_{G^{(n)}}$. For this, let $G \in W_n$. By definition, for all $[G',o'] \in \mathcal{A}_{k_n,\Delta_n}$, we have $|\psi^{(n)}_{G^{(n)}}([G',o'])| = |\psi^{(n)}_G([G',o'])|$. Hence there exists a permutation π on the set of vertices $\{1,\dots,n\}$ such that $(G^{(n)},i)_{k_n} \equiv (G,\pi(i))_{k_n}$ for all $1 \leq i \leq n$. Using Lemma 4.1 above with $h = k_n$ and $L = n$, we have

$$d_{\mathrm{LP}}(U(G^{(n)}), U(G)) \leq \frac{1}{1+k_n}.$$

Consequently,

$$d_{\mathrm{LP}}(U(G),\mu) < \epsilon_n + 1/(1+k_n). \tag{4.8}$$

We claim that for $G \in W_n$ we have $\vec{m}_G = \vec{m}^{(n)}$ and $\vec{u}_G = \vec{u}^{(n)}$. To see this, note that for $\theta \in \Theta$ we have

$$u_G(\theta) = \sum_{i=1}^n \mathbb{1}\left[\tau_G(i) = \theta\right]$$

$$= \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n}} \sum_{i \in \psi_G^{(n)}([G',o'])} \mathbb{1}\left[\tau_G(i) = \theta\right].$$

Note that for $i \in \psi_G^{(n)}([G', o'])$ we have $\tau_G(i) = \tau_{G'}(o')$. Therefore,

$$u_G(\theta) = \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n} : \tau_{G'}(o') = \theta} |\psi_G^{(n)}([G',o'])|.$$

A similar argument implies

$$u^{(n)}(\theta) = \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n} : \tau_{G'}(o') = \theta} |\psi_{G^{(n)}}^{(n)}([G',o'])|.$$

Hence $u^{(n)}(\theta) = u_G(\theta)$. Likewise, for $G \in W_n$ and $x \neq x' \in \Xi$, we can write, for n large enough,

$$\begin{aligned}
m_G(x,x') &= \sum_{i=1}^n \deg_G^{x,x'}(i) \\
&= \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n}} \deg_{G'}^{x,x'}(o') |\psi_G^{(n)}([G',o'])| \\
&= \sum_{[G',o'] \in \mathcal{A}_{k_n,\Delta_n}} \deg_{G'}^{x,x'}(o') |\psi_{G^{(n)}}^{(n)}([G',o'])| \\
&= m^{(n)}(x,x').
\end{aligned}$$

The proof of $m_G(x,x) = m^{(n)}(x,x)$ for $x \in \Xi$ is similar. This, together with (4.8), implies that $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu, \epsilon_n + 1/(1+k_n))$ which completes the proof of (4.7).

Note that, for fixed $t > 0$ and $x, x' \in \Xi$, the mapping $[G,o] \mapsto \deg_G^{x,x'}(o) \wedge t$ is bounded and continuous. Therefore, for $x \neq x' \in \Xi$, we have

$$\begin{aligned}
\frac{m^{(n)}(x,x')}{n} &= \int \deg_G^{x,x'}(o) dU(G^{(n)})([G,o]) \\
&\geq \int (\deg_G^{x,x'}(o) \wedge t) dU(G^{(n)})([G,o]) \\
&\xrightarrow{n \to \infty} \int (\deg_G^{x,x'}(o) \wedge t) d\mu.
\end{aligned}$$

Sending t to infinity, we get

$$\liminf_{n \to \infty} \frac{m^{(n)}(x,x')}{n} \geq \deg_{x,x'}(\mu).$$

Similarly, for $x \in \Xi$, we have $\liminf_{n\to\infty} m_n(x,x)/n \geq \deg_{x,x}(\mu)/2$. On the other hand, for $\theta \in \Theta$, the mapping $[G,o] \mapsto \mathbb{1}\left[\tau_G(o) = \theta\right]$ is bounded and continuous. This implies that

$$\lim_{n \to \infty} u^{(n)}(\theta) = \Pi_\theta(\mu).$$

Thus, substituting (4.7) into (4.6), using the fact that $\epsilon_n + 1/(1+k_n) \to 0$, and using Lemma 4.2 above, we get

$$\limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} \le \limsup_{n\to\infty} \frac{\log \left| \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}} \left(\mu, \epsilon_n + \frac{1}{1+k_n}\right) \right| - \|\vec{m}^{(n)}\|_1 \log n}{n}$$
$$\le \Sigma(\mu),$$

which completes the proof. □

4.2.2 Step 2: The General Compression Scheme

In Section 4.2.1 we introduced a compression scheme which achieves the BC entropy of μ by focusing on the depth k_n empirical distribution of the graph $G^{(n)}$ in the sequence of graphs $\{G^{(n)}\}_{n=1}^\infty$, under the assumption that the maximum degree of $G^{(n)}$ is bounded above by Δ_n which does not grow too fast, in the sense that $|\mathcal{A}_{k_n,\Delta_n}| = o(n/\log n)$. In principle, we can choose the design parameter k_n, but we have no control over the maximum degree Δ_n. In order to overcome this and drop the restriction on the compression scheme in Section 4.2.1, we first choose k_n and Δ_n and then trim the input graph by removing some edges to make its maximum degree no more than Δ_n. Then, we encode the resulting trimmed graph by the compression function in Section 4.2.1. Finally, we encode the removed edges separately. More precisely, we encode a graph $G^{(n)} \in \bar{\mathcal{G}}_n$ as follows:

1. Define $\Delta_n := \log\log n$.
2. Let $\widetilde{G}^{(n)} := (G^{(n)})^{\Delta_n}$ be the trimmed graph obtained by removing each edge connected to any vertex with degree more than Δ_n. Moreover, define
$$R_n := \{1 \le i \le n : \deg_{G^{(n)}}(i) > \Delta_n \text{ or } \deg_{G^{(n)}}(j) > \Delta_n \text{ for some } j \sim_{G^{(n)}} i\},$$
which consists of the endpoints of the removed edges.
3. Encode the graph $\widetilde{G}^{(n)}$ by the compression function introduced in Section 4.2.1, with $k_n = \sqrt{\log\log n}$.
4. Encode $|R_n|$ using at most $(1 + \lfloor \log_2 n \rfloor)\log 2$ nats.
5. Encode the set R_n using at most $(1 + \lfloor \log_2 \binom{n}{|R_n|} \rfloor)\log 2$ nats.
6. Let $\vec{m}^{(n)} = \vec{m}_{G^{(n)}}$ and $\vec{\widetilde{m}}^{(n)} = \vec{m}_{\widetilde{G}^{(n)}}$. Note that the edges present in $G^{(n)}$ but not in $\widetilde{G}^{(n)}$ have both endpoints in the set R_n. So we can first encode $m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')$ for all $x, x' \in \Xi$ by $|\Xi|^2(1 + \lfloor \log_2 n^2 \rfloor)\log 2 \le 2|\Xi|^2(1 + \lfloor \log_2 n \rfloor)\log 2$ nats and then encode these removed edges using
$$\sum_{x \le x' \in \Xi} \left(1 + \left\lfloor \log_2 \binom{\binom{|R_n|}{2}}{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')} \right\rfloor \right) \log 2$$

nats by specifying the removed edges of each pair of marks separately.

Now we show that this general compression scheme asymptotically achieves the upper BC entropy rate, as was stated in Theorem 4.1. Before this, we need the results of the following lemmas. We postpone the proofs of these lemmas to Appendix C.1.

Lemma 4.3. *Assume that $\{G^{(n)}\}_{n=1}^\infty$ is a sequence of marked graphs with local weak limit $\mu \in \mathcal{P}(\bar{\mathcal{T}}_*)$, where $G^{(n)}$ is on the vertex set $\{1, \dots, n\}$. If Δ_n is a sequence of integers going to infinity as $n \to \infty$, μ is also the local weak limit of the trimmed sequence $\{(G^{(n)})^{\Delta_n}\}_{n=1}^\infty$.*

Lemma 4.4. *If $\Delta_n \leq \log \log n$ and $k_n \leq \sqrt{\log \log n}$, then $|\mathcal{A}_{k_n, \Delta_n}| = o(n / \log n)$.*

Lemma 4.5. *Assume that $\{G^{(n)}\}_{n=1}^\infty$ is a sequence of marked graphs with local weak limit $\mu \in \mathcal{P}(\bar{\mathcal{T}}_*)$, where $G^{(n)}$ is on the vertex set $\{1, \dots, n\}$. Let $\{\Delta_n\}_{n=1}^\infty$ be a sequence of integers such that $\Delta_n \to \infty$ and define*

$$R_n := \{1 \leq i \leq n : \deg_{G^{(n)}}(i) > \Delta_n \text{ or } \deg_{G^{(n)}}(j) > \Delta_n \text{ for some } j \sim_{G^{(n)}} i\}.$$

Then $|R_n|/n \to 0$ as n goes to infinity.

Proof of Theorem 4.1. Let $\tilde{f}_n$ be the compression function of the scheme in Section 4.2.1. We have

$$\begin{aligned} \mathsf{nats}(f_n(G^{(n)})) \leq \mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) &+ \log n + \log \binom{n}{|R_n|} \\ &+ 2|\Xi|^2 \log n \\ &+ \sum_{x \leq x' \in \Xi} \log \binom{\binom{|R_n|}{2}}{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')} \\ &+ C \log 2, \end{aligned}$$

where $C = 2 + 3|\Xi|^2$. Using the inequality $\binom{r}{s} \leq (re/s)^s$ and Lemma 3.5 above, we have

$$\begin{aligned} \mathsf{nats}(f_n(G^{(n)})) \leq \mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) &+ (1 + 2|\Xi|^2) \log n \\ &+ |R_n| \log \frac{ne}{|R_n|} \\ &+ (\|\vec{m}^{(n)}\|_1 - \|\vec{\widetilde{m}}^{(n)}\|_1) \log |R_n| \\ &+ \frac{|R_n||\Xi|^2}{2} + C \log 2. \end{aligned}$$

Using the fact that $|R_n| \leq n$, this gives

$$\mathsf{nats}(f_n(G^{(n)})) \leq \mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) + (1 + 2|\Xi|^2) \log n$$

$$+ |R_n| \log \frac{ne}{|R_n|}$$
$$+ (\|\vec{m}^{(n)}\|_1 - \|\vec{\tilde{m}}^{(n)}\|_1) \log n$$
$$+ \frac{|R_n||\Xi|^2}{2} + C \log 2.$$

Hence,

$$\begin{aligned} \limsup_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)}) - \|\vec{m}^{(n)}\|_1 \log n}{n} &\leq \limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)}) - \|\vec{\tilde{m}}^{(n)}\|_1 \log n}{n} \\ &\quad + \limsup_{n\to\infty} \frac{|R_n||\Xi|^2}{2n} + \limsup_{n\to\infty} \frac{|R_n|}{n} \log \frac{ne}{|R_n|}. \end{aligned} \tag{4.9}$$

Now, we claim that the conditions of Proposition 4.1 hold for the sequence $\widetilde{G}^{(n)}$ and the parameters k_n and Δ_n defined above. To show this, note that both k_n and Δ_n go to infinity by definition. Lemma 4.3 then implies that μ is also the local weak limit of the sequence $\widetilde{G}^{(n)}$. Moreover, by Lemma 4.4, $|\mathcal{A}_{k_n,\Delta_n}| = o(n/\log n)$. On the other hand, the maximum degree in $\widetilde{G}^{(n)}$ is at most Δ_n. Therefore, all the conditions of Proposition 4.1 are satisfied and

$$\limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)})) - \|\vec{\tilde{m}}^{(n)}\|_1 \log n}{n} \leq \Sigma(\mu).$$

Furthermore, all the other terms in (4.9) go to zero, since, by Lemma 4.5, $|R_n|/n \to 0$, and the function $\delta \mapsto \delta \log \delta$ goes to zero as $\delta \to 0$. Therefore,

$$\begin{aligned} \limsup_{n\to\infty} \frac{\mathsf{nats}(f_n(G^{(n)})) - \|\vec{m}^{(n)}\|_1 \log n}{n} &\leq \limsup_{n\to\infty} \frac{\mathsf{nats}(\tilde{f}_n(\widetilde{G}^{(n)}) - \|\vec{\tilde{m}}^{(n)}\|_1 \log n}{n} \\ &\leq \Sigma(\mu), \end{aligned}$$

which completes the proof. □

Remark 4.3. *From Lemma 4.5 above, for typical graphs, $|R_n| = o(n)$. Hence, similar to our discussion in Remark 4.2, we are capable of answering local queries with an error of $o(n)$ without needing to go through the decompression process.*

4.3 Conclusion

In this chapter, employing the local weak convergence framework from Chapter 2 and the marked BC entropy from Chapter 3, we formalized the problem of compressing graphical data without assuming prior knowledge of its stochastic properties. More precisely, we proposed a universal compression scheme which is asymptotically optimal in the size of the underlying graph, where optimality is characterized using the marked BC entropy. Moreover, this compression scheme is capable of performing local data queries in the compressed form, with an error negligible compared to the number of vertices.

Chapter 5

Distributed Compression of Graphical Data

In Chapter 4, we introduced a universal compression scheme for a single source of graphical data. As the data is not always available in one location, it is also important to consider distributed compression of graphical data. This latter question is the focus of this chapter. Traditionally, when dealing with time series, distributed lossless compression is modeled using two (or more) possibly dependent jointly stationary and ergodic processes representing the components of the data at the individual locations. In this case, the rate region, which characterizes how efficiently the data can be compressed, is given by the Slepian–Wolf Theorem [CT12]. We adopt an analogous framework, namely that two jointly defined marked random graphs on the same vertex set are presented to two encoders, one to each encoder. Each encoder is then required to individually compress its data such that a third party, having access to the two compressed representations, can recover both marked graph realizations with a vanishing probability of error in the asymptotic limit of the size of the data.

We characterize the compression rate region for two scenarios, namely, a sequence of marked sparse Erdős–Rényi ensembles and a sequence of marked configuration model ensembles. We employ the framework of local weak convergence of Chapter 2 as a counterpart of stochastic processes for sparse marked graphs. Our characterization of the rate region is best understood in terms of the marked BC entropy which was introduced in Chapter 3.

This chapter is organized as follows. In Section 5.1 we introduce the notation and formally state the problem. In Section 5.2, we study the local weak limits of the marked sparse Erdős–Rényi and the marked configuration model ensembles, we analyze the asymptotic behavior of their entropy, and connect this asymptotic behavior to the marked BC entropy of their respective local weak limits. Then, in Section 5.3, we characterize the rate region for distributed lossless compression in the scenarios of Section 5.1. Also, in Section 5.3.5, we generalize this result to the case where there are more than two graphical sources. Finally, we conclude the chapter in Section 5.4.

5.1 Problem Statement

In this chapter, for the sake of simplicity, we assume that in all marked graphs, and in all edges in such a marked graph, the two marks towards the endpoints of that edge are the same. In other words, each edge effectively carries one mark. Therefore, with an abuse of notation, an edge mark count vector in this setting is a vector of nonnegative integers $\vec{m} = (m(x) : x \in \Xi)$, where $m(x)$ denotes the number of edges with mark x.

Let G be a marked graph on a finite vertex set with edges and vertices carrying marks in the sets Ξ and Θ, respectively. With an abuse of notation, we denote the edge mark count vector of G by $\vec{m}_G = \{m_G(x)\}_{x\in\Xi}$, where $m_G(x)$ is the number of edges in G carrying mark x. In fact, comparing to the notation in Section 2.3, we have $m_G(x) = m_G(x, x)$, and $m_G(x, x') = 0$ when $x \neq x'$. Also, recall from Section 2.3 that we denote the vertex mark count vector of G by $\vec{u}_G = \{u_G(\theta)\}_{\theta\in\Theta}$, where $u_G(\theta)$ denotes the number of vertices in G carrying mark θ. Additionally, for a graph G on the vertex set $[n]$, we denote the degree sequence of G by $\overrightarrow{\mathrm{dg}}_G = (\deg_G(1), \dots, \deg_G(n))$, where $\deg_G(i)$ denotes the degree of vertex i. For a degree sequence $\vec{d} = (d(1), \dots, d(n))$ and a nonnegative integer k, we define

$$c_k(\vec{d}) := |\{1 \le i \le n : d(i) = k\}|. \tag{5.1}$$

Also, for two degree sequences $\vec{d} = (d(1), \dots, d(n))$ and $\vec{d'} = (d'(1), \dots, d'(n))$, and two nonnegative integers k and l, we define

$$c_{k,l}(\vec{d}, \vec{d'}) := |\{1 \le i \le n : d(i) = k, d'(i) = l\}|. \tag{5.2}$$

Given a degree sequence $\vec{d} = (d(1), \dots, d(n))$, we let $\mathcal{G}^{(n)}_{\vec{d}}$ denote the set of simple unmarked graphs G on the vertex set $[n]$ such that $\deg_G(i) = d(i)$ for $1 \le i \le n$.

When discussing distributed compression of graphical data with two sources, we assume that Ξ_1 and Ξ_2 are two fixed and finite sets of edge marks and Θ_1 and Θ_2 are two fixed and finite sets of vertex marks. For $i \in \{1, 2\}$ and $n \in \mathbb{N}$, let $\mathcal{G}^{(n)}_i$ denote the set of marked graphs on the vertex set $[n]$ with edge and vertex mark sets Ξ_i and Θ_i respectively. For two graphs $G_1 \in \mathcal{G}^{(n)}_1$ and $G_2 \in \mathcal{G}^{(n)}_2$, $G_1 \oplus G_2$ denotes the superposition of G_1 and G_2 which is a marked graph defined as follows: a vertex $1 \le v \le n$ in $G_1 \oplus G_2$ carries the mark (θ_1, θ_2) where θ_i is the mark of v in G_i. Furthermore, we place an edge in $G_1 \oplus G_2$ between vertices v and w if there is an edge between them in at least one of G_1 of G_2, and mark this edge (x_1, x_2), where, for $1 \le i \le 2$, x_i is the mark of the edge (v, w) in G_i if it exists and $\circ_i$ otherwise. Here $\circ_1$ and $\circ_2$ are auxiliary marks not present in $\Xi_1 \cup \Xi_2$. Note that $G_1 \oplus G_2$ is a marked graph with edge and vertex mark sets $\Xi_{1,2} := (\Xi_1 \cup \{\circ_1\}) \times (\Xi_2 \cup \{\circ_2\}) \setminus \{(\circ_1, \circ_2)\}$ and $\Theta_{1,2} := \Theta_1 \times \Theta_2$, respectively. We use the terminology *jointly marked graph* to refer to a marked graph with edge and vertex mark sets $\Xi_{1,2}$ and $\Theta_{1,2}$ respectively. With this, let $\mathcal{G}^{(n)}_{1,2}$ denote the set of jointly marked graphs on the vertex set $[n]$. Moreover, for $i \in \{1, 2\}$, we say that a graph is in the i–th domain if its edge and vertex marks come from Ξ_i and Θ_i respectively. For a jointly marked graph $G_{1,2}$ and $1 \le i \le 2$, the i–th marginal of $G_{1,2}$, denoted by G_i, is the marked graph in the i–th domain obtained by projecting all vertex and edge marks onto Ξ_i and Θ_i, respectively, followed by removing edges with mark $\circ_i$. Note that any jointly marked graph $G_{1,2}$ is uniquely determined by its marginals G_1 and G_2, because $G_{1,2} = G_1 \oplus G_2$. Given an edge mark count vector $\vec{m} = \{m(x)\}_{x \in \Xi_{1,2}}$, for $x_1 \in \Xi_1 \cup \{\circ_1\}$ and $x_2 \in \Xi_2 \cup \{\circ_2\}$, with an abuse of notation we define

$$m(x_1) := \sum_{(x_1',x_2') \in \Xi_{1,2} : x_1' = x_1} m((x_1', x_2')), \qquad m(x_2) := \sum_{(x_1',x_2') \in \Xi_{1,2} : x_2' = x_2} m((x_1', x_2')). \tag{5.3}$$

Likewise, given a vertex mark count vector $\vec{u} = \{u(\theta)\}_{\theta \in \Theta_{1,2}}$, we define, for $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$,

$$u(\theta_1) := \sum_{\theta_2' \in \Theta_2} u((\theta_1, \theta_2')), \qquad u(\theta_2) := \sum_{\theta_1' \in \Theta_1} u((\theta_1', \theta_2)). \tag{5.4}$$

Assume that we have a sequence of random marked graphs $G_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$, defined for all n sufficiently large, drawn for each n according to some ensemble distribution on $\mathcal{G}_{1,2}^{(n)}$. Additionally, assume that there are two encoders who want to compress realizations of such jointly marked graphs in a distributed fashion. Namely, the i–th encoder, $1 \le i \le 2$, has only access to the i–th marginal $G_i^{(n)}$. We assume that the distribution of $G_{1,2}^{(n)}$ is known.

Definition 5.1. *A sequence of $\langle n, L_1^{(n)}, L_2^{(n)} \rangle$ codes is a sequence of triples $(f_1^{(n)}, f_2^{(n)}, g^{(n)})$, defined for all sufficiently large n, such that*

$$f_i^{(n)} : \mathcal{G}_i^{(n)} \to [L_i^{(n)}], \qquad i \in \{1, 2\},$$

and

$$g^{(n)} : [L_1^{(n)}] \times [L_2^{(n)}] \to \mathcal{G}_{1,2}^{(n)}.$$

The probability of error for this code corresponding to the ensemble of $G_{1,2}^{(n)}$, which is denoted by $P_e^{(n)}$, is defined as

$$P_e^{(n)} := \mathbb{P}\left(g^{(n)}(f_1^{(n)}(G_1^{(n)}), f_2^{(n)}(G_2^{(n)})) \neq G_{1,2}^{(n)} \right).$$

Now we define our achievability criterion.

Definition 5.2. *A rate tuple $(\alpha_1, R_1, \alpha_2, R_2) \in \mathbb{R}^4$ is said to be achievable for distributed compression of the sequence of random graphs $G_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$ if there is a sequence of $\langle n, L_1^{(n)}, L_2^{(n)} \rangle$ codes such that*

$$\limsup_{n \to \infty} \frac{\log L_i^{(n)} - (\alpha_i n \log n + R_i n)}{n} \le 0, \qquad i \in \{1, 2\}, \tag{5.5}$$

and also $P_e^{(n)} \to 0$. The rate region $\mathcal{R} \in \mathbb{R}^4$ is defined as follows: for fixed α_1 and α_2, if there are sequences $R_1^{(m)}$ and $R_2^{(m)}$ with limit points R_1 and R_2 in $\mathbb{R}$, respectively, such that for each m the rate tuple $(\alpha_1, R_1^{(m)}, \alpha_2, R_2^{(m)})$ is achievable, then we include $(\alpha_1, R_1, \alpha_2, R_2)$ in the set $\mathcal{R}$.

In this chapter, we characterize the above rate region for the following two sequences of ensembles:

A sequence of Erdős–Rényi ensembles: Assume that nonnegative real numbers $\vec{p} = \{p_x\}_{x \in \Xi_{1,2}}$ together with a probability distribution $\vec{q} = \{q_\theta\}_{\theta \in \Theta_{1,2}}$ are given such that, for all $x_1 \in \Xi_1$ and $x_2 \in \Xi_2$, we have

$$\sum_{\substack{(x'_1,x'_2) \in \Xi_{1,2} \\ x'_1 = x_1}} p_{(x'_1,x'_2)} > 0 \qquad \text{and} \qquad \sum_{\substack{(x'_1,x'_2) \in \Xi_{1,2} \\ x'_2 = x_2}} p_{(x'_1,x'_2)} > 0, \tag{5.6}$$

and, for all $(\theta_1, \theta_2) \in \Theta_{1,2}$, we have

$$\sum_{\theta'_2 \in \Theta_2} q_{(\theta_1,\theta'_2)} > 0 \qquad \text{and} \qquad \sum_{\theta'_1 \in \Theta_1} q_{(\theta'_1,\theta_2)} > 0. \tag{5.7}$$

For $n \in \mathbb{N}$ large enough, we define the probability distribution $\mathcal{G}(n; \vec{p}, \vec{q})$ on $\mathcal{G}^{(n)}_{1,2}$ as follows: for each pair of vertices $1 \leq i < j \leq n$, the edge (i,j) is present in the graph and has mark $x \in \Xi_{1,2}$ with probability p_x/n, and is not present with probability $1 - \sum_{x \in \Xi_{1,2}} p_x/n$. Furthermore, each vertex in the graph is given a mark $\theta \in \Theta_{1,2}$ with probability q_θ. The choice of edge and vertex marks is done independently.

The conditions in (5.7) and the conditions for $x_i \in \Xi_i$, $i = 1, 2$, in (5.6) are required only to ensure that the sets of vertex marks and edge marks are chosen to be as small as possible, and these conditions could be relaxed if desired.

A sequence of configuration model ensembles: Fix $\Delta \in \mathbb{N}$. Suppose that a probability distribution $\vec{r} = \{r_k\}_{k=0}^{\Delta}$ supported on the set $\{0, \dots, \Delta\}$ is given, such that $r_0 < 1$. Moreover, assume that probability distributions $\vec{\gamma} = \{\gamma_x\}_{x \in \Xi_{1,2}}$ and $\vec{q} = \{q_\theta\}_{\theta \in \Theta_{1,2}}$ on the sets $\Xi_{1,2}$ and $\Theta_{1,2}$, respectively, are given. We assume that, for all $x_1 \in \Xi_1 \cup \{\circ_1\}$ and $x_2 \in \Xi_2 \cup \{\circ_2\}$, we have

$$\sum_{\substack{(x'_1,x'_2) \in \Xi_{1,2} \\ x'_1 = x_1}} \gamma_{(x'_1,x'_2)} > 0 \qquad \text{and} \qquad \sum_{\substack{(x'_1,x'_2) \in \Xi_{1,2} \\ x'_2 = x_2}} \gamma_{(x'_1,x'_2)} > 0, \tag{5.8}$$

and, for all $(\theta_1, \theta_2) \in \Theta_{1,2}$, we have

$$\sum_{\theta'_2 \in \Theta_2} q_{(\theta_1,\theta'_2)} > 0 \qquad \text{and} \qquad \sum_{\theta'_1 \in \Theta_1} q_{(\theta'_1,\theta_2)} > 0. \tag{5.9}$$

Furthermore, for each n, the degree sequence $\vec{d}^{(n)} = \{d^{(n)}(1), \dots, d^{(n)}(n)\}$ is given such that, for all $1 \leq i \leq n$, we have $d^{(n)}(i) \leq \Delta$ and also $\sum_{i=1}^n d^{(n)}(i)$ is even. Let $m_n := (\sum_{i=1}^n d^{(n)}(i))/2$. Additionally, if, for $0 \leq k \leq \Delta$, $c_k(\vec{d}^{(n)})$ denotes the number of $1 \leq i \leq n$ such that $d^{(n)}(i) = k$, we assume that, for some constant $K > 0$, we have

$$\sum_{k=0}^{\Delta} |c_k(\vec{d}^{(n)}) - n r_k| \leq K n^{1/2}. \tag{5.10}$$

Now, for fixed $\vec{r}$, $\vec{\gamma}$ and $\vec{q}$ as above, and a sequence $\vec{d}^{(n)}$ satisfying (5.10), we define the law $\mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$ on $\mathcal{G}_{1,2}^{(n)}$, for $n \in \mathbb{N}$ large enough, as follows. First, we pick an unmarked graph on the vertex set $[n]$ uniformly at random among the set of graphs G with maximum degree Δ such that for each $0 \leq k \leq \Delta$, $c_k(\overrightarrow{\text{dg}}_G) = c_k(\vec{d}^{(n)})$.[1] Then, we assign i.i.d. marks with law $\vec{\gamma}$ on the edges and i.i.d. marks with law $\vec{q}$ on the vertices.

The conditions in (5.9) and the conditions for $x_i \in \Xi_i$, $i = 1, 2$, in (5.8) are required only to ensure that the sets of vertex marks and edge marks are chosen to be as small as possible, and these conditions could be relaxed if desired. However, the conditions in (5.8) for $x_i = \circ_i$, $i = 1, 2$, are essential, as will be pointed out at the appropriate point in the proofs, since they ensure that neither of the two underlying unmarked graphs is a subgraph of the other.

As we will discuss in Section 5.2 below, the sequence of Erdős–Rényi ensembles defined above converges in the local weak sense to a marked Poisson Galton Watson tree. Moreover, the sequence of configuration model ensembles converges in the same sense to a marked Galton Watson process with degree distribution $\vec{r}$. In Section 5.3, we will characterize the achievable rate regions for lossless distributed compression of graphical data modeled as coming from one of the two sequences of ensembles above in terms of these limiting objects for the above two sequences of ensembles respectively. The formulation of this result will be in terms of the marked BC entropy which we discussed in Chapter 3.

Remark 5.1. *It should be pointed out that a rate region in the sense of Definition 5.2 need not be a closed set, in contrast to what one is used to in the discussion of the Slepian-Wolf region in the traditional case. Further, while α_1 and α_2 can be restricted to being nonnegative, R_1 and R_2 should be thought of as real numbers. Indeed, the rate regions for the two sequences of ensembles considered in this chapter, which are characterized in Theorem 5.1, are not closed sets. The correct way to think of such a rate region is in terms of the subsets of $(R_1, R_2) \in \mathbb{R}^2$, parametrized by $(\alpha_1, \alpha_2) \in \mathbb{R}^2$, for which $(\alpha_1, R_1, \alpha_2, R_2)$ lies in the rate region, and each such subset is closed as a subset of $\mathbb{R}^2$. Further, for any $(\alpha_1, R_1, \alpha_2, R_2)$ in the rate region, if $\alpha_1' > \alpha_1$ then $(\alpha_1', R_1', \alpha_2, R_2)$ lies in the rate region for all $R_1' \in \mathbb{R}$, and a similar statement holds if one replaces the index 1 by the index 2.*

5.2 Asymptotic of the Erdős–Rényi and the configuration model ensembles

In this section, we first study the local weak limits of the marked sparse Erdős–Rényi and the marked configuration model ensembles introduced in Section 5.1. Then, we analyze the asymptotic behavior of the entropy of each of the two models, and connect it to the marked BC entropy of the limit associated to that model.

[1] The fact that each degree is bounded by Δ, $r_0 < 1$ and the sum of degrees is even implies that $\vec{d}^{(n)}$ is a graphic sequence for $n \in \mathbb{N}$ large enough. This is, for instance, a consequence of Theorem 4.5 in [BC15].

5.2.1 The local weak limit of the Erdős–Rényi and the configuration model ensembles

Let $G_{1,2}^{(n)}$ be a random jointly marked graph with law $\mathcal{G}(n; \vec{p}, \vec{q})$ and let v_n be a vertex chosen uniformly at random in the set $[n]$. A simple Poisson approximation implies that $D_x(v_n)$, the number of edges adjacent to v_n with mark $x \in \Xi_{1,2}$, converges in distribution to a Poisson random variable with mean p_x, as n goes to infinity. Moreover, $\{D_x(v_n)\}_{x \in \Xi_{1,2}}$ are asymptotically mutually independent. A similar argument can be repeated for any other vertex in the neighborhood of v_n. Also, it can be shown that the probability of having cycles of any fixed length converges to zero. In fact, the isomorphism class of $(G_{1,2}^{(n)}, v_n)_h$ converges in distribution to that of a rooted marked Poisson Galton Watson tree with depth h.

More precisely, let $(T_{1,2}^{\mathrm{ER}}, o)$ be a rooted jointly marked tree defined as follows. First, the mark of the root is chosen with distribution $\vec{q}$. Then, for $x \in \Xi_{1,2}$, we independently generate D_x with law Poisson(p_x). We then add D_x many edges with mark x to the root o. For each offspring, i.e. vertex at the other end of an edge connected to the root, we repeat the same procedure independently, i.e. choose its vertex mark according to the distribution $\vec{q}$ and then attach additional edges with each edge mark from the corresponding Poisson distribution with mean p_x, independently for each edge mark in $\Xi_{1,2}$. Recursively repeating this, we get a connected jointly marked tree $T_{1,2}^{\mathrm{ER}}$ rooted at o, which has possibly countably infinitely many vertices. Let $\mu_{1,2}^{\mathrm{ER}}$ denote the law of the isomorphism class $[T_{1,2}^{\mathrm{ER}}, o]$. Note that $\mu_{1,2}^{\mathrm{ER}}$ is a probability distribution on $\mathcal{G}_*(\Xi_{1,2}, \Theta_{1,2})$. $\mu_{1,2}^{\mathrm{ER}}$ depends on the underlying choice of the parameters $(\vec{p}, \vec{q})$, but we suppress this from the notation, for readability. The above discussion implies that, for all $h \geq 0$, $[G_{1,2}^{(n)}, v_n]_h$ converges in distribution to $[T_{1,2}^{\mathrm{ER}}, o]_h$. In fact, even a stronger statement can be proved, which is the following: If we consider the sequence of random graphs $G_{1,2}^{(n)}$ independently on a joint probability space, $U(G_{1,2}^{(n)})$ converges weakly to $\mu_{1,2}^{\mathrm{ER}}$ with probability one. With this, we say that, almost surely, $\mu_{1,2}^{\mathrm{ER}}$ is the *local weak limit* of the sequence $G_{1,2}^{(n)}$, where the term "local" is meant to indicate that we require the convergence in distribution of the isomorphism class of each fixed depth neighborhood of a typical vertex (i.e. a vertex chosen uniformly at random).

With the construction above, let T_i^{ER} be the i–th marginal of $T_{1,2}^{\mathrm{ER}}$, for $1 \leq i \leq 2$. Moreover, let μ_i^{ER} be the law of $[T_i^{\mathrm{ER}}(o), o]$. Therefore, μ_i^{ER} is a probability distribution on $\mathcal{G}_*(\Xi_i, \Theta_i)$. Similarly to the argument above, one can see that, almost surely, μ_i^{ER} is the local weak limit of the sequence $G_i^{(n)}$.

A similar picture also holds for the configuration model. Let $(T_{1,2}^{\mathrm{CM}}, o)$ be a rooted jointly marked random tree constructed as follows. First, we generate the degree of the root o with law $\vec{r}$. Then, for each offspring w of o, we independently generate the offspring count of w with law $\vec{r'} = \{r'_k\}_{k=0}^{\Delta-1}$ defined as

$$r'_k = \frac{(k+1)r_{k+1}}{\mathbb{E}[X]}, \qquad 0 \leq k \leq \Delta - 1,$$

where X has law $\vec{r}$. We continue this process recursively, i.e. for each vertex other than the root, we independently generate its offspring count with law $\vec{r'}$. The distribution $\vec{r'}$ is called the *size-biased* distribution, and takes into account the fact that each vertex other than the root has an extra edge by virtue of its being defined via an edge to an earlier defined vertex, and hence its degree should be biased in order to get the correct degree distribution $\vec{r}$. Then, for each vertex and edge existing in the graph $T_{1,2}^{\mathrm{CM}}$, we generate marks independently with laws $\vec{q}$ and $\vec{\gamma}$, respectively. Let $\mu_{1,2}^{\mathrm{CM}}$ be the law of $[T_{1,2}^{\mathrm{CM}}, o]$. Moreover, for $1 \leq i \leq 2$, let μ_i^{CM} be the law of $[T_i^{\mathrm{CM}}(o), o]$. It can be shown that if $G_{1,2}^{(n)}$ has law $\mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$, with these random graphs being constructed independently on a joint probability space, then, almost surely, $\mu_{1,2}^{\mathrm{CM}}$ is the local weak limit of $G_{1,2}^{(n)}$, and μ_i^{CM} is the local weak limit of $G_i^{(n)}$, for $1 \leq i \leq 2$. $\mu_{1,2}^{\mathrm{CM}}$ depends on the choice of the underlying parameters $(\vec{\gamma}, \vec{q}, \vec{r})$, but we suppress this from the notation, for readability.

5.2.2 Asymmptotic behavior of the entropy for the Erdős–Rényi and the configuration model ensembles

The following general lemma, whose proof is straightforward using Stirling's approximation, is often used in this chapter. See Appendix D.1 for a proof.

Lemma 5.1. *Let $k \in \mathbb{N}$. Let a_n and $b_1^n, \dots, b_k^n$ be sequences of integers, defined for all sufficiently large n.*

1. *Assume that $a_n = \sum_{i=1}^k b_k^n$ for all n. If $a_n/n \to a > 0$ and, for each $1 \leq i \leq k$, $b_i^n/n \to b_i \geq 0$ where $a = \sum_{i=1}^k b_i$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \binom{a_n}{\{b_i^n\}_{1\leq i\leq k}} = aH\left(\left\{\frac{b_i}{a}\right\}_{1\leq i\leq k}\right).$$

2. *Assume that $a_n \geq \sum_{i=1}^k b_k^n$ for all n. If $a_n/\binom{n}{2} \to 1$ and $b_i^n/n \to b_i \geq 0$, we have*

$$\lim_{n\to\infty} \frac{\log \binom{a_n}{\{b_i^n\}_{1\leq i\leq k}} - \left(\sum_{i=1}^k b_i^n\right)\log n}{n} = \sum_{i=1}^k s(2b_i),$$

where $s(x)$ is defined to be $\frac{x}{2} - \frac{x}{2}\log x$ for $x > 0$ and 0 if $x = 0$.

We next connect the asymptotic behavior of the entropy of the ensembles defined in Section 5.1 to the marked BC entropy of their local weak limits. We first consider a sequence of Erdős–Rényi ensembles. Let $n \in \mathbb{N}$ be large enough, and assume that $G_{1,2}^{(n)}$ has law $\mathcal{G}(n; \vec{p}, \vec{q})$. Let $d_{1,2}^{\mathrm{ER}} := \deg(\mu_{1,2}^{\mathrm{ER}}) = \sum_{x\in\Xi_{1,2}} p_x$. For $x_i \in \Xi_i$ and $\theta_i \in \Theta_i$, $1 \leq i \leq 2$, let

$$\begin{aligned} p_{x_1} &:= \sum_{x_2' \in \Xi_2 \cup \{\circ_2\}} p_{(x_1, x_2')}, \quad p_{x_2} := \sum_{x_1' \in \Xi_1 \cup \{\circ_1\}} p_{(x_1', x_2)}, \\ q_{\theta_1} &:= \sum_{\theta_2' \in \Theta_2} q_{(\theta_1, \theta_2')}, \quad q_{\theta_2} := \sum_{\theta_1' \in \Theta_1} q_{(\theta_1', \theta_2)}. \end{aligned} \tag{5.11}$$

For $1 \leq i \leq 2$, let $d_i^{\mathrm{ER}} := \deg(\mu_i^{\mathrm{ER}}) = \sum_{x_i \in \Xi_i} p_{x_i}$. If $Q = (Q_1, Q_2)$ has law $\vec{q}$, it can be verified that we have

$$H(G_{1,2}^{(n)}) = \frac{d_{1,2}^{\mathrm{ER}}}{2} n \log n + n \left(H(Q) + \sum_{x \in \Xi_{1,2}} s(p_x) \right) + o(n), \tag{5.12a}$$

$$H(G_1^{(n)}) = \frac{d_1^{\mathrm{ER}}}{2} n \log n + n \left(H(Q_1) + \sum_{x_1 \in \Xi_1} s(p_{x_1}) \right) + o(n), \tag{5.12b}$$

$$H(G_2^{(n)}) = \frac{d_2^{\mathrm{ER}}}{2} n \log n + n \left(H(Q_2) + \sum_{x_2 \in \Xi_2} s(p_{x_2}) \right) + o(n). \tag{5.12c}$$

Using Theorem 3.3 from Chapter 3, it can be seen that the coefficients of n in equations (5.12a)–(5.12c) are $\Sigma(\mu_{1,2}^{\mathrm{ER}})$, $\Sigma(\mu_1^{\mathrm{ER}})$ and $\Sigma(\mu_2^{\mathrm{ER}})$, respectively.

Before discussing configuration model ensembles, we state two lemmas, which are used at several points. The proof of the following Lemma 5.2 is straightforward, and is therefore omitted.

Lemma 5.2. *Let $\Delta \in \mathbb{N}$. Let Y be a random variable taking values in $\{0, 1, \ldots, \Delta\}$, and let $0 \leq \epsilon \leq 1$. Let $\{V_i\}_{i \geq 1}$ be a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(V_i = 1) = \epsilon$, and let $Y_1 := \sum_{i=1}^{Y} V_i$, where $Y_1 = 0$ when $Y = 0$. Then, we have*

$$H(Y_1, Y - Y_1) = H(Y_1, Y) = H(Y) + \mathbb{E}[Y] H(V_1) - \mathbb{E}\left[\log \binom{Y}{Y_1} \right].$$

□

The proof of the following Lemma 5.3 is given in Appendix D.2.

Lemma 5.3. *Let $\Delta \in \mathbb{N}$. Let Y be a random variable taking values in $\{0, 1, \ldots \Delta\}$, such that $d := \mathbb{E}[Y] > 0$. For all $n \in \mathbb{N}$ large enough, let $\vec{a}^{(n)} = (a^{(n)}(1), \ldots, a^{(n)}(n))$ be a degree sequence of length n with entries bounded by Δ such that $b_n := \sum_{i=1}^{n} a^{(n)}(i)$ is even and, for $0 \leq k \leq \Delta$, we have $c_k(\vec{a}^{(n)})/n \to \mathbb{P}(Y = k)$. Then, we have*

$$\lim_{n \to \infty} \frac{\log |\mathcal{G}_{\vec{a}^{(n)}}^{(n)}| - \frac{b_n}{2} \log n}{n} = -s(d) - \mathbb{E}[\log Y!],$$

where we recall that $\mathcal{G}_{\vec{a}^{(n)}}^{(n)}$ denotes the set of simple unmarked graphs G on the vertex set $[n]$ such that $\deg_G(i) = a^{(n)}(i)$ for $1 \leq i \leq n$.

Remark 5.2. *The assumption $\mathbb{E}[Y] > 0$ in the above lemma is crucial and cannot be relaxed. To see this, consider the following example: let $Y = 0$ with probability one, and let $\vec{a}^{(n)}$ be such that $a^{(n)}(1) = a^{(n)}(2) = 3$ and $a^{(n)}(i) = 0$ for $i > 2$. Then, although b_n is even, $\vec{a}^{(n)}$ is not graphic and $\mathcal{G}_{\vec{a}^{(n)}}^{(n)}$ is empty. Therefore, the above limit of interest is $-\infty$ and the equality does not hold.*

Consider now a sequence of configuration model ensembles. Namely, for all $n \in \mathbb{N}$ large enough, let $G_{1,2}^{(n)}$ be distributed according to $\mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$. Let X be a random variable with law $\vec{r}$ and $\Gamma^k = (\Gamma_1^k, \Gamma_2^k)$, $1 \le k \le \Delta$, an i.i.d. sequence distributed according to $\vec{\gamma}$. With this, let

$$X_1 := \sum_{k=1}^{X} \mathbb{1}\left[\Gamma_1^k \neq \circ_1\right], \qquad X_2 := \sum_{k=1}^{X} \mathbb{1}\left[\Gamma_2^k \neq \circ_2\right], \tag{5.13}$$

where $X_1 = X_2 = 0$ if $X = 0$. Then, if $d_{1,2}^{\mathrm{CM}} := \deg(\mu_{1,2}^{\mathrm{CM}})$ and, for $1 \le i \le 2$, $d_i^{\mathrm{CM}} := \deg(\mu_i^{\mathrm{CM}})$, it can be seen that

$$\begin{aligned} H(G_{1,2}^{(n)}) = \frac{d_{1,2}^{\mathrm{CM}}}{2} n \log n + n\Big(- s(d_{1,2}^{\mathrm{CM}}) + H(X) - \mathbb{E}\left[\log X!\right] \\ + H(Q) + \frac{d_{1,2}^{\mathrm{CM}}}{2} H(\Gamma)\Big) + o(n), \end{aligned} \tag{5.14a}$$

$$\begin{aligned} H(G_1^{(n)}) = \frac{d_1^{\mathrm{CM}}}{2} n \log n + n\Big(- s(d_1^{\mathrm{CM}}) + H(X_1) - \mathbb{E}\left[\log X_1!\right] \\ + H(Q_1) + \frac{d_1^{\mathrm{CM}}}{2} H(\Gamma_1 | \Gamma_1 \neq \circ_1)\Big) + o(n), \end{aligned} \tag{5.14b}$$

$$\begin{aligned} H(G_2^{(n)}) = \frac{d_2^{\mathrm{CM}}}{2} n \log n + n\Big(- s(d_2^{\mathrm{CM}}) + H(X_2) - \mathbb{E}\left[\log X_2!\right] \\ + H(Q_2) + \frac{d_2^{\mathrm{CM}}}{2} H(\Gamma_2 | \Gamma_2 \neq \circ_2)\Big) + o(n), \end{aligned} \tag{5.14c}$$

where Γ is distributed according to $\vec{\gamma}$. Also, using Theorem 3.3 from Chapter 3, it can be seen that the coefficients of n in equations (5.14a)–(5.14c) are $\Sigma(\mu_{1,2}^{\mathrm{CM}})$, $\Sigma(\mu_1^{\mathrm{CM}})$ and $\Sigma(\mu_2^{\mathrm{CM}})$, respectively. The proof of equations (5.14a)–(5.14c), which is given in Appendix D.3, and depends on both Lemma 5.2 and Lemma 5.3.

If $\mu_{1,2}$ is any one of the two distributions $\mu_{1,2}^{\mathrm{ER}}$ or $\mu_{1,2}^{\mathrm{CM}}$, and μ_1 and μ_2 are its marginals, we define the *conditional marked BC entropies* as $\Sigma(\mu_2|\mu_1) := \Sigma(\mu_{1,2}) - \Sigma(\mu_1)$ and $\Sigma(\mu_1|\mu_2) := \Sigma(\mu_{1,2}) - \Sigma(\mu_2)$.

5.3 Main Results

Now, we are ready to state our main result, which is to characterize the rate region in Definition 5.2 for a sequence of Erdős–Rényi ensembles and a sequence of configuration model ensembles. In the following, for pairs of reals (α, R) and (α', R'), we write $(\alpha, R) \succ (\alpha', R')$ if either $\alpha > \alpha'$, or $\alpha = \alpha'$ and $R > R'$. We also write $(\alpha, R) \succeq (\alpha', R')$ if either $(\alpha, R) \succ (\alpha', R')$ or $(\alpha, R) = (\alpha', R')$.

Theorem 5.1. *Assume $\mu_{1,2}$ is a member of either of the two families of distributions $\mu_{1,2}^{ER}$ (parametrized by $(\vec{p}, \vec{q})$) or $\mu_{1,2}^{CM}$ (parametrized by $(\vec{\gamma}, \vec{q}, \vec{r})$) defined in Section 5.2. Then,* *if $\mathcal{R}$ is the rate region for the sequence of ensembles corresponding to $\mu_{1,2}$, as defined in Section 5.1, a rate tuple $(\alpha_1, R_1, \alpha_2, R_2) \in \mathcal{R}$ if and only if*

$$(\alpha_1, R_1) \succeq ((d_{1,2} - d_2)/2, \Sigma(\mu_1|\mu_2)), \tag{5.15a}$$

$$(\alpha_2, R_2) \succeq ((d_{1,2} - d_1)/2, \Sigma(\mu_2|\mu_1)), \tag{5.15b}$$

$$(\alpha_1 + \alpha_2, R_1 + R_2) \succeq (d_{1,2}/2, \Sigma(\mu_{1,2})), \tag{5.15c}$$

where $d_{1,2} := \deg(\mu_{1,2})$, $d_1 := \deg(\mu_1)$ and $d_2 := \deg(\mu_2)$.

We prove the achievability for the Erdős–Rényi case and the configuration model case in Sections 5.3.1 and 5.3.2, respectively. Subsequently, we prove the converses for the two cases in Sections 5.3.3 and 5.3.4, respectively.

As is the case for the classical Slepian–Wolf theorem, one can generalize the above result to more than two sources. The definition of the rate region as well as its characterization can be naturally extended to this case. In Section 5.3.5 below, we generalize the Erdős–Rényi and configuration model ensembles to more than two sources, define the corresponding Slepian-Wolf rate region, and characterize the rate region for each of these cases in Theorem 5.2. The proof structure is similar to that for the scenario with two sources, and is highlighted in Appendix D.6.

5.3.1 Proof of Achievability for the Erdős–Rényi case

Here we show that a rate tuple $(\alpha_1, R_1, \alpha_2, R_2)$ is achievable for the Erdős–Rényi ensemble if it satisfies the following

$$(\alpha_1, R_1) \succ ((d_{1,2}^{\mathrm{ER}} - d_2^{\mathrm{ER}})/2, \Sigma(\mu_1^{\mathrm{ER}}|\mu_2^{\mathrm{ER}})), \tag{5.16a}$$

$$(\alpha_2, R_2) \succ ((d_{1,2}^{\mathrm{ER}} - d_1^{\mathrm{ER}})/2, \Sigma(\mu_2^{\mathrm{ER}}|\mu_1^{\mathrm{ER}})), \tag{5.16b}$$

$$(\alpha_1 + \alpha_2, R_1 + R_2) \succ (d_{1,2}^{\mathrm{ER}}/2, \Sigma(\mu_{1,2}^{\mathrm{ER}})). \tag{5.16c}$$

Note that if a rate tuple $(\alpha_1', R_1', \alpha_2', R_2')$ satisfies the weak inequalities (5.15a)–(5.15c) then, for any $\epsilon > 0$, $(\alpha_1', R_1' + \epsilon, \alpha_2', R_2' + \epsilon)$ satisfies the strict inequalities (5.16a)–(5.16c). As we show below, this implies that $(\alpha_1', R_1' + \epsilon, \alpha_2', R_2' + \epsilon)$ is achievable. Hence, after sending $\epsilon \to 0$, we get $(\alpha_1', R_1', \alpha_2', R_2') \in \mathcal{R}$.

We show that any $(\alpha_1, R_1, \alpha_2, R_2)$ satisfying (5.16a)–(5.16c) is achievable by employing a random binning method. More precisely, for $i \in \{1, 2\}$, we set $L_i^{(n)} = \lfloor \exp(\alpha_i n \log n + R_i n) \rfloor$ and for each $G_i \in \mathcal{G}_i^{(n)}$, we assign $f_i^{(n)}(G_i)$ uniformly at random in the set $[L_i^{(n)}]$ and independent of everything else.

To describe our decoding scheme, we first need to set up some notation. Let $\mathcal{M}^{(n)}$ denote the set of edge count vectors $\vec{m} = \{m(x)\}_{x \in \Xi_{1,2}}$ such that

$$\sum_{x \in \Xi_{1,2}} |m(x) - np_x/2| \leq n^{2/3}.$$

Moreover, let $\mathcal{U}^{(n)}$ denote the set of vertex mark count vectors $\vec{u} = \{u(\theta)\}_{\theta \in \Theta_{1,2}}$ such that

$$\sum_{\theta \in \Theta_{1,2}} |u(\theta) - nq_\theta| \leq n^{2/3}.$$

Furthermore, we define $\mathcal{G}^{(n)}_{\vec{p},\vec{q}}$ to be the set of graphs $H^{(n)}_{1,2} \in \mathcal{G}^{(n)}_{1,2}$ such that $\vec{m}_{H^{(n)}_{1,2}} \in \mathcal{M}^{(n)}$ and $\vec{u}_{H^{(n)}_{1,2}} \in \mathcal{U}^{(n)}$. Upon receiving $(i,j) \in [L^{(n)}_1] \times [L^{(n)}_2]$, we form the set of graphs $H^{(n)}_{1,2} \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}$ such that $f^{(n)}_1(H^{(n)}_1) = i$ and $f^{(n)}_2(H^{(n)}_2) = j$, where $H^{(n)}_1$ and $H^{(n)}_2$ are the marginals of $H^{(n)}_{1,2}$. If this set has only one element, we output this element as the decoded graph; otherwise, we report an error.

In what follows, assume that $G^{(n)}_{1,2}$ is a random graph with law $\mathcal{G}(n;\vec{p},\vec{q})$. We consider the following four error events corresponding to the above scheme:

$$\begin{aligned}
&\mathcal{E}^{(n)}_1 := \{G^{(n)}_{1,2} \notin \mathcal{G}^{(n)}_{\vec{p},\vec{q}}\},\\
&\mathcal{E}^{(n)}_2 := \{\exists H^{(n)}_{1,2} \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}} : H^{(n)}_1 \neq G^{(n)}_1, H^{(n)}_2 \neq G^{(n)}_2, f^{(n)}_i(H^{(n)}_i) = f^{(n)}_i(G^{(n)}_i), i \in \{1,2\}\},\\
&\mathcal{E}^{(n)}_3 := \{\exists H^{(n)}_2 \neq G^{(n)}_2 : G^{(n)}_1 \oplus H^{(n)}_2 \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}, f^{(n)}_2(H^{(n)}_2) = f^{(n)}_2(G^{(n)}_2)\},\\
&\mathcal{E}^{(n)}_4 := \{\exists H^{(n)}_1 \neq G^{(n)}_1 : H^{(n)}_1 \oplus G^{(n)}_2 \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}, f^{(n)}_1(H^{(n)}_1) = f^{(n)}_1(G^{(n)}_1)\}.
\end{aligned}$$

Note that outside the above four events the decoder successfully decodes the input graph $G^{(n)}_{1,2}$.

Using Chebyshev's inequality, for some $\kappa > 0$ we have $\mathbb{P}(\mathcal{E}^{(n)}_1) \leq \kappa n^{-1/3}$, which converges to zero as n goes to infinity. Moreover, using the union bound, we have

$$\mathbb{P}\left(\mathcal{E}^{(n)}_2\right) \leq \frac{|\mathcal{G}^{(n)}_{\vec{p},\vec{q}}|}{L^{(n)}_1 L^{(n)}_2}. \tag{5.17}$$

Note that, for each graph $H^{(n)}_{1,2} \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}$, the mark count vectors $\vec{m}_{H^{(n)}_{1,2}}$ and $\vec{u}_{H^{(n)}_{1,2}}$ are in the sets $\mathcal{M}^{(n)}$ and $\mathcal{U}^{(n)}$ respectively. Additionally, we have $|\mathcal{M}^{(n)}| \leq (2n^{2/3}+1)^{|\Xi_{1,2}|}$ and $|\mathcal{U}^{(n)}| \leq (2n^{2/3}+1)^{|\Theta_{1,2}|}$. Therefore,

$$|\mathcal{G}^{(n)}_{\vec{p},\vec{q}}| \leq (2n^{2/3}+1)^{(|\Xi_{1,2}|+|\Theta_{1,2}|)} \max_{\substack{\vec{m} \in \mathcal{M}^{(n)} \\ \vec{u} \in \mathcal{U}^{(n)}}} A_1(\vec{m},\vec{u}), \tag{5.18}$$

where

$$A_1(\vec{m},\vec{u}) := \binom{n}{\{u(\theta)\}_{\theta \in \Theta_{1,2}}} \binom{\binom{n}{2}}{\{m(x)\}_{x \in \Xi_{1,2}}}.$$

Now, let $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ be sequences in $\mathcal{M}^{(n)}$ and $\mathcal{U}^{(n)}$, respectively. Then, for all $x \in \Xi_{1,2}$ and $\theta \in \Theta_{1,2}$, we have $m^{(n)}(x)/n \to p_x/2$ and $u^{(n)}(\theta)/n \to q_\theta$. Thereby, using Lemma 5.1, we have

$$\lim_{n \to \infty} \frac{\log A_1(\vec{m}^{(n)},\vec{u}^{(n)}) - (\sum_{x \in \Xi_{1,2}} m^{(n)}(x)) \log n}{n}$$

$$= H(\vec{q}) + \sum_{x \in \Xi_{1,2}} s(p_x) = \Sigma(\mu^{\mathrm{ER}}_{1,2}).$$

Substituting this into (5.18) and using the fact that $\sum |m^{(n)}(x) - np_x/2| \le n^{2/3}$, we have

$$\limsup_{n\to\infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{p},\vec{q}}| - n\frac{d^{\mathrm{ER}}_{1,2}}{2}\log n}{n} \le \Sigma(\mu^{\mathrm{ER}}_{1,2}). \tag{5.19}$$

Substituting this into (5.17), we have

$$\begin{aligned}
&\limsup \frac{1}{n} \log \mathbb{P}\left(\mathcal{E}^{(n)}_2\right) \\
&\le \limsup \frac{\log |\mathcal{G}^{(n)}_{\vec{p},\vec{q}}| - n\frac{d^{\mathrm{ER}}_{1,2}}{2}\log n - n\Sigma(\mu^{\mathrm{ER}}_{1,2})}{n} \\
&\quad + \limsup \frac{n(\frac{d^{\mathrm{ER}}_{1,2}}{2} - \alpha_1 - \alpha_2)\log n + n(\Sigma(\mu^{\mathrm{ER}}_{1,2}) - R_1 - R_2)}{n} \\
&\quad + \limsup \frac{n(\alpha_1+\alpha_2)\log n + n(R_1+R_2) - \log L^{(n)}_1 L^{(n)}_2}{n}.
\end{aligned}$$

The first term is nonpositive due to (5.19), the second term is strictly negative due to the assumption (5.16c), and the third term is nonpositive due to our choice of $L^{(n)}_1$ and $L^{(n)}_2$. Consequently, the RHS is strictly negative, which implies that $\mathbb{P}(\mathcal{E}^{(n)}_2) \to 0$.

Now, we show that $\mathbb{P}(\mathcal{E}^{(n)}_3 \setminus \mathcal{E}^{(n)}_1)$ vanishes. In order to do so, for $H^{(n)}_1 \in \mathcal{G}^{(n)}_1$, define $S^{(n)}_2(H^{(n)}_1) := \{H^{(n)}_2 \in \mathcal{G}^{(n)}_2 : H^{(n)}_1 \oplus H^{(n)}_2 \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}\}$. Using the union bound, we have

$$\begin{aligned}
\mathbb{P}\left(\mathcal{E}^{(n)}_3 \setminus \mathcal{E}^{(n)}_1\right) &\le \sum_{H^{(n)}_{1,2} \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}} \mathbb{P}(G^{(n)}_{1,2} = H^{(n)}_{1,2}) \frac{|S^{(n)}_2(H^{(n)}_1)|}{L^{(n)}_2} \\
&\le \frac{1}{L^{(n)}_2} \max_{H^{(n)}_{1,2} \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}} |S^{(n)}_2(H^{(n)}_1)|.
\end{aligned} \tag{5.20}$$

It can be shown that (See Appendix D.4)

$$\limsup_{n\to\infty} \frac{\max_{H^{(n)}_{1,2} \in \mathcal{G}^{(n)}_{\vec{p},\vec{q}}} \log |S^{(n)}_2(H^{(n)}_1)| - n\frac{d^{\mathrm{ER}}_{1,2} - d^{\mathrm{ER}}_1}{2}\log n}{n} \le \Sigma(\mu^{\mathrm{ER}}_2 | \mu^{\mathrm{ER}}_1), \tag{5.21}$$

where $H^{(n)}_1$ is the first marginal of $H^{(n)}_{1,2}$. Substituting this in (5.20), we get

$$\begin{aligned}
\limsup \frac{1}{n} \log \mathbb{P}\left(\mathcal{E}^{(n)}_3 \setminus \mathcal{E}^{(n)}_1\right) &\le \limsup \frac{n\frac{d^{\mathrm{ER}}_{1,2}-d^{\mathrm{ER}}_1}{2}\log n + n\Sigma(\mu^{\mathrm{ER}}_2|\mu^{\mathrm{ER}}_1) - \log L^{(n)}_2}{n} \\
&\le \limsup \frac{n(\frac{d^{\mathrm{ER}}_{1,2}-d^{\mathrm{ER}}_1}{2} - \alpha_2)\log n + n(\Sigma(\mu^{\mathrm{ER}}_2|\mu^{\mathrm{ER}}_1) - R_2)}{n} \\
&\quad + \limsup \frac{n\alpha_2 \log n + nR_2 - \log L^{(n)}_2}{n}.
\end{aligned} \tag{5.22}$$

Note that the first term is strictly negative due to the assumption (5.16b), while the second term is nonpositive due to our way of choosing $L_2^{(n)}$. This means that $\mathbb{P}(\mathcal{E}_3^{(n)} \setminus \mathcal{E}_1^{(n)})$ goes to zero as n goes to infinity. Similarly, $\mathbb{P}(\mathcal{E}_4^{(n)} \setminus \mathcal{E}_1^{(n)})$ converges to zero as $n \to \infty$. This means that there exists a sequence of deterministic codebooks with vanishing probability of error, which completes the proof of achievability.

5.3.2 Proof of Achievability for the Configuration model

Our achievability proof for this case is very similar in nature to that for the Erdős–Rényi case, with the modifications discussed below.

Let $\mathcal{D}^{(n)}$ be the set of degree sequences $\vec{d}$ with entries bounded by Δ such that $c_k(\vec{d}) = c_k(\vec{d}^{(n)})$ for all $0 \le k \le \Delta$. Moreover, redefine $\mathcal{M}^{(n)}$ to be the set of mark count vectors $\vec{m}$ such that $\sum_{x \in \Xi_{1,2}} m(x) = m_n$ and $\sum_{x \in \Xi_{1,2}} |m(x) - m_n \gamma_x| \le n^{2/3}$, where we recall that $m_n = (\sum_{i=1}^n d^{(n)}(i))/2$. We use the same definition for $\mathcal{U}^{(n)}$ as in the previous section, i.e. the set of vertex mark count vectors $\vec{u}$ such that $\sum_{\theta \in \Theta_{1,2}} |u(\theta) - n q_\theta| \le n^{2/3}$.

In what follows, let X be a random variable with law $\vec{r}$, X_1 and X_2 defined as in (5.13), and $\Gamma = (\Gamma_1, \Gamma_2)$ a random variable with law $\vec{\gamma}$.

We define $\mathcal{W}^{(n)}$ to be the set of graphs $H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$ such that: (i) $\overrightarrow{\mathrm{dg}}_{H_{1,2}^{(n)}} \in \mathcal{D}^{(n)}$, (ii) $\vec{m}_{H_{1,2}^{(n)}} \in \mathcal{M}^{(n)}$, (iii) $\vec{u}_{H_{1,2}^{(n)}} \in \mathcal{U}^{(n)}$, (iv) for all $0 \le l \le k \le \Delta$, recalling the notation in (5.2), we have

$$|c_{k,l}(\overrightarrow{\mathrm{dg}}_{H_{1,2}^{(n)}}, \overrightarrow{\mathrm{dg}}_{H_1^{(n)}}) - n\mathbb{P}\left(X = k, X_1 = l\right)| \le n^{2/3}, \tag{5.23}$$

and (v) for all $0 \le l \le k \le \Delta$ we have

$$|c_{k,l}(\overrightarrow{\mathrm{dg}}_{H_{1,2}^{(n)}}, \overrightarrow{\mathrm{dg}}_{H_2^{(n)}}) - n\mathbb{P}\left(X = k, X_2 = l\right)| \le n^{2/3}. \tag{5.24}$$

We employ a similar random binning framework as in Section 5.3.1. For decoding, upon receiving a pair (i, j), we form the set of graphs $H_{1,2}^{(n)} \in \mathcal{W}^{(n)}$ such that $f_1^{(n)}(H_1^{(n)}) = i$ and $f_2^{(n)}(H_2^{(n)}) = j$. If this set has only one element, we output it as the source graph; otherwise, we output an indication of error. In order to prove the achievability, we consider the four error events $\mathcal{E}_i^{(n)}$, $1 \le i \le 4$, defined exactly like those in the previous section, with $\mathcal{G}_{\vec{p},\vec{q}}^{(n)}$ being replaced with $\mathcal{W}^{(n)}$.

It can be shown that if $G_{1,2}^{(n)} \sim \mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$, the probability of $G_{1,2}^{(n)} \in \mathcal{W}^{(n)}$ goes to one as n goes to infinity (see Lemma D.1 in Appendix D.3). Therefore, $\mathbb{P}(\mathcal{E}_1^{(n)})$ goes to zero as $n \to \infty$.

To show that $\mathbb{P}(\mathcal{E}_2^{(n)})$ vanishes, similar to the analysis in Section 5.3.1, we find an asymptotic upper bound for $\log |\mathcal{W}^{(n)}|$. By only considering the conditions (i), (ii) and (iii) in the definition of $\mathcal{W}^{(n)}$, we have

$$\begin{aligned}\log|\mathcal{W}^{(n)}| \le \log\binom{n}{\{c_k(\vec{d}^{(n)})\}_{k=0}^{\Delta}} + \log|\mathcal{G}^{(n)}_{\vec{d}^{(n)}}| \\ + \log\left((2n^{2/3}+1)^{|\Xi_{1,2}|}\max_{\vec{m}\in\mathcal{M}^{(n)}}\binom{m_n}{\{m(x)\}_{x\in\Xi_{1,2}}}\right) \\ + \log\left((2n^{2/3}+1)^{|\Theta_{1,2}|}\max_{\vec{u}\in\mathcal{U}^{(n)}}\binom{n}{\{u(\theta)\}_{\theta\in\Theta_{1,2}}}\right).\end{aligned} \tag{5.25}$$

By assumption, we have $r_0 < 1$, hence $d^{\mathrm{CM}}_{1,2} > 0$. The condition (5.10) together with Lemma 5.3 in Appendix D.3 then implies that

$$\begin{aligned}\lim_{n\to\infty}\frac{\log|\mathcal{G}^{(n)}_{\vec{d}^{(n)}}| - n\frac{d^{\mathrm{CM}}_{1,2}}{2}\log n}{n} &= \lim_{n\to\infty}\frac{\log|\mathcal{G}^{(n)}_{\vec{d}^{(n)}}| - m_n\log n}{n} + \lim_{n\to\infty}\frac{(m_n - nd^{\mathrm{CM}}_{1,2}/2)\log n}{n} \\ &= -s(d^{\mathrm{CM}}_{1,2}) - \mathbb{E}\left[\log X!\right],\end{aligned} \tag{5.26}$$

where on the second line we have used the bound $|m_n - nd^{\mathrm{CM}}_{1,2}/2| \le K\Delta n^{1/2}$ which is implied by (5.10). Using this together with Lemma 5.1 for the other terms in (5.25), we have

$$\begin{aligned}\limsup_{n\to\infty}\frac{\log|\mathcal{W}^{(n)}| - n\frac{d^{\mathrm{CM}}_{1,2}}{2}\log n}{n} \le -s(d^{\mathrm{CM}}_{1,2}) + H(X) \\ + \frac{d^{\mathrm{CM}}_{1,2}}{2}H(\Gamma) + H(Q) - \mathbb{E}\left[\log X!\right] = \Sigma(\mu^{\mathrm{CM}}_{1,2}),\end{aligned}$$

where Γ and Q are random variables with law $\vec{\gamma}$ and $\vec{q}$, respectively.

Now, in order to show that $\mathbb{P}(\mathcal{E}^{(n)}_3 \setminus \mathcal{E}^{(n)}_1)$ vanishes, we prove a counterpart for (5.21). For $H^{(n)}_1 \in \mathcal{G}^{(n)}_1$, we define $S^{(n)}_2(H^{(n)}_1)$ to be the set of graphs $H^{(n)}_2 \in \mathcal{G}^{(n)}_2$ such that $H^{(n)}_1 \oplus H^{(n)}_2 \in \mathcal{W}^{(n)}$. Then, it can be shown (see Appendix D.5) that

$$\limsup_{n\to\infty}\frac{\max_{H^{(n)}_{1,2}\in\mathcal{W}^{(n)}}\log|S^{(n)}_2(H^{(n)}_1)| - n\frac{d^{\mathrm{CM}}_{1,2} - d^{\mathrm{CM}}_1}{2}\log n}{n} \le \Sigma(\mu^{\mathrm{CM}}_2|\mu^{\mathrm{CM}}_1). \tag{5.27}$$

Then, similar to (5.22), this shows that $\mathbb{P}(\mathcal{E}^{(n)}_3 \setminus \mathcal{E}^{(n)}_1)$ vanishes as $n \to \infty$. Similarly, $\mathbb{P}(\mathcal{E}^{(n)}_4 \setminus \mathcal{E}^{(n)}_1)$ vanishes as $n \to \infty$. This completes the proof of achievability.

5.3.3 Proof of the Converse for the Erdős–Rényi case

In this section, we show that every rate tuple $(\alpha_1, R_1, \alpha_2, R_2) \in \mathcal{R}$ for the Erdős–Rényi scenario must satisfy the conditions (5.15a)–(5.15c). By definition, for a rate tuple $(\alpha_1, R_1, \alpha_2, R_2) \in \mathcal{R}$, there exist sequences $R^{(m)}_1$ and $R^{(m)}_2$ such that for each m, $(\alpha_1, R^{(m)}_1, \alpha_2, R^{(m)}_2)$ is achievable and, besides, we have $R^{(m)}_1 \to R_1$ and $R^{(m)}_2 \to R_2$. If we show that $(\alpha_1, R^{(m)}_1, \alpha_2, R^{(m)}_2)$

satisfies (5.15a)–(5.15c) for each m, it is easy to see that $(\alpha_1, R_1, \alpha_2, R_2)$ must also satisfy the same inequalities. Therefore, it suffices to show that any achievable rate tuple satisfies (5.15a)–(5.15c).

For this, take an achievable rate tuple $(\alpha_1, R_1, \alpha_2, R_2)$ together with a corresponding sequence of $\langle n, L_1^{(n)}, L_2^{(n)} \rangle$ codes $(f_1^{(n)}, f_2^{(n)}, g^{(n)})$. By definition, we have

$$\limsup_{n\to\infty} \frac{\log L_i^{(n)} - (\alpha_i n \log n + R_i n)}{n} \leq 0 \qquad i \in \{1,2\}, \tag{5.28}$$

and also the error probability $P_e^{(n)}$ goes to zero as n goes to infinity. Now, we define the set $\mathcal{A}^{(n)} \subseteq \mathcal{G}_{1,2}^{(n)}$ as

$$\mathcal{A}^{(n)} := \mathcal{G}_{\vec{p},\vec{q}}^{(n)} \cap \{H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)} : g^{(n)}(f_1^{(n)}(H_1^{(n)}), f_2^{(n)}(H_2^{(n)})) = H_{1,2}^{(n)}\}, \tag{5.29}$$

where $\mathcal{G}_{\vec{p},\vec{q}}^{(n)}$ was defined in Section 5.3.1. In fact, $\mathcal{A}^{(n)}$ is the set of "typical" graphs with respect to the Erdős–Rényi model that are successfully decoded by the code $(f_1^{(n)}, f_2^{(n)}, g^{(n)})$. In the following, let $G_{1,2}^{(n)} \sim \mathcal{G}^{(n)}(n; \vec{p}, \vec{q})$ be distributed according to the Erdős–Rényi model. Moreover, let $P_{\text{ER}}^{(n)}$ be the law of $G_{1,2}^{(n)}$, i.e. for $H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$, $P_{\text{ER}}^{(n)}(H_{1,2}^{(n)}) := \mathbb{P}(G_{1,2}^{(n)} = H_{1,2}^{(n)})$. With this, we define a random variable $\widetilde{G}_{1,2}^{(n)}$ whose distribution is the conditional distribution of $G_{1,2}^{(n)}$, conditioned on lying in $\mathcal{A}^{(n)}$, i.e.

$$\mathbb{P}\left(\widetilde{G}_{1,2}^{(n)} = H_{1,2}^{(n)}\right) = \begin{cases} P_{\text{ER}}^{(n)}(H_{1,2}^{(n)})/\pi_n & H_{1,2}^{(n)} \in \mathcal{A}^{(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.30}$$

where $\pi_n := \mathbb{P}\left(G_{1,2}^{(n)} \in \mathcal{A}^{(n)}\right)$ is the normalizing factor. Note that, since $P_e^{(n)} \to 0$ as $n \to \infty$ and $P(G_{1,2}^{(n)} \in \mathcal{A}^{(n)}) \to 1$ as $n \to \infty$, we have $\pi_n > 0$ for all sufficiently large n, and in fact $\pi_n \to 1$ as $n \to \infty$. Additionally, let $\tilde{P}_{\text{ER}}^{(n)}$ be the law of $\widetilde{G}_{1,2}^{(n)}$. If, for $i \in \{1,2\}$, $\tilde{M}_i^{(n)}$ denotes $f_i^{(n)}(\widetilde{G}_i^{(n)})$, we have

$$\begin{aligned} \log L_1^{(n)} + \log L_2^{(n)} &\geq H(\tilde{M}_1^{(n)}) + H(\tilde{M}_2^{(n)}) \geq H(\tilde{M}_1^{(n)}, \tilde{M}_2^{(n)}) \\ &= H(\widetilde{G}_{1,2}^{(n)}), \end{aligned} \tag{5.31}$$

where the last equality follows from the fact that, by definition, $\widetilde{G}_{1,2}^{(n)}$ takes values among the graphs that are successfully decoded, and hence is uniquely identified given $\tilde{M}_1^{(n)}$ and $\tilde{M}_2^{(n)}$.

Now, we find a lower bound for $H(\widetilde{G}_{1,2}^{(n)})$. For doing so, note that for $H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$ and n

large enough, we have

$$-\log P_{\text{ER}}^{(n)}(H_{1,2}^{(n)}) = -\sum_{x\in\Xi_{1,2}} m_{H_{1,2}^{(n)}}(x)\log\frac{p_x}{n} - \left[\binom{n}{2} - \sum_{x\in\Xi_{1,2}} m_{H_{1,2}^{(n)}}(x)\right]\log\left(1-\frac{\sum_{x\in\Xi_{1,2}} p_x}{n}\right) - \sum_{\theta\in\Theta_{1,2}} u_{H_{1,2}^{(n)}}(\theta)\log q_\theta. \tag{5.32}$$

On the other hand, due to the definition of $\mathcal{G}_{\vec{p},\vec{q}}^{(n)}$, if $H_{1,2}^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)}$ then, for all $x \in \Xi_{1,2}$ and $\theta \in \Theta_{1,2}$, we have

$$n\frac{p_x}{2} - n^{2/3} \le m_{H_{1,2}^{(n)}}(x) \le n\frac{p_x}{2} + n^{2/3}, \text{ and}$$
$$nq_\theta - n^{2/3} \le u_{H_{1,2}^{(n)}}(\theta) \le nq_\theta + n^{2/3}.$$

Substituting these in (5.32) and using the inequality $\log(1-x) \le -x$ which holds for $x \in (0,1)$, for n large enough, we have

$$\begin{aligned}-\log P_{\text{ER}}^{(n)}(H_{1,2}^{(n)}) \ge &\sum_{x\in\Xi_{1,2}}\left(n\frac{p_x}{2} - n^{2/3}\right)(\log n - \log p_x)\\ &+ \left[\binom{n}{2} - \sum_{x\in\Xi_{1,2}}\left(n\frac{p_x}{2}+n^{2/3}\right)\right]\frac{\sum_{x\in\Xi_{1,2}} p_x}{n}\\ &- \sum_{\theta\in\Theta_{1,2}}(nq_\theta - n^{2/3})\log q_\theta.\end{aligned}$$

Using $\sum_{x\in\Xi_{1,2}} p_x = d_{1,2}^{\text{ER}}$ and simplifying the above, we realize that there exists a constant $c > 0$ that does not depend on n or $H_{1,2}^{(n)}$, such that, for all $H_{1,2}^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)}$ and thus, in particular, for all $H_{1,2}^{(n)} \in \mathcal{A}^{(n)}$, we have

$$\begin{aligned}-\log P_{\text{ER}}^{(n)}(H_{1,2}^{(n)}) &\ge n\frac{d_{1,2}^{\text{ER}}}{2}\log n - n\sum_{x\in\Xi_{1,2}}\frac{p_x}{2}\log p_x + n\sum_{x\in\Xi_{1,2}}\frac{p_x}{2} - n\sum_{\theta\in\Theta_{1,2}} q_\theta\log q_\theta - cn^{2/3}\log n\\ &= n\frac{d_{1,2}^{\text{ER}}}{2}\log n + n\Sigma(\mu_{1,2}^{\text{ER}}) - cn^{2/3}\log n.\end{aligned} \tag{5.33}$$

Now, if $\widetilde{G}_{1,2}^{(n)}$ is the random variable defined in (5.30), we have

$$\begin{aligned}H(\widetilde{G}_{1,2}^{(n)}) &= -\sum_{H_{1,2}^{(n)}\in\mathcal{A}^{(n)}} \tilde{P}_{\text{ER}}^{(n)}(H_{1,2}^{(n)})\log\tilde{P}_{\text{ER}}^{(n)}(H_{1,2}^{(n)})\\ &= \log\pi_n - \frac{1}{\pi_n}\sum_{H_{1,2}^{(n)}\in\mathcal{A}^{(n)}} P_{\text{ER}}^{(n)}(H_{1,2}^{(n)})\log P_{\text{ER}}^{(n)}(H_{1,2}^{(n)}).\end{aligned}$$

Note that since the probability of error of the above code vanishes, i.e. $P_e^{(n)} \to 0$, and $\mathbb{P}\left(G_{1,2}^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)}\right) \to 1$, we have $\pi_n \to 1$ as $n \to \infty$. On the other hand, with probability one, we have $\widetilde{G}_{1,2}^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)}$. Also, by the definition of π_n, we have $\sum_{H_{1,2}^{(n)} \in \mathcal{A}^{(n)}} P_{\text{ER}}^{(n)}(H_{1,2}^{(n)}) = \pi_n$. Thereby, employing the bound (5.33), we have

$$\liminf_{n\to\infty} \frac{H(\widetilde{G}_{1,2}^{(n)}) - n\frac{d_{1,2}^{\text{ER}}}{2}\log n}{n} \geq \Sigma(\mu_{1,2}^{\text{ER}}). \tag{5.34}$$

Now, using the assumption (5.28) together with the bound (5.31), we have

$$\begin{aligned} 0 &\geq \limsup_{n\to\infty} \frac{\log L_1^{(n)} + \log L_2^{(n)} - (\alpha_1+\alpha_2)n\log n - n(R_1+R_2)}{n} \\ &\geq \liminf_{n\to\infty} \frac{H(\widetilde{G}_{1,2}^{(n)}) - n\frac{d_{1,2}^{\text{ER}}}{2}\log n - n\Sigma(\mu_{1,2}^{\text{ER}})}{n} \\ &\quad + \liminf_{n\to\infty} \frac{n\frac{d_{1,2}^{\text{ER}}}{2}\log n + n\Sigma(\mu_{1,2}^{\text{ER}}) - (\alpha_1+\alpha_2)n\log n - n(R_1+R_2)}{n}. \end{aligned} \tag{5.35}$$

The first term is nonnegative due to (5.34). Consequently,

$$0 \geq \liminf_{n\to\infty} \frac{n\left(\frac{d_{1,2}^{\text{ER}}}{2} - \alpha_1 - \alpha_2\right)\log n + n(\Sigma(\mu_{1,2}^{\text{ER}}) - R_1 - R_2)}{n}. \tag{5.36}$$

Note that this is impossible unless $\alpha_1 + \alpha_2 \geq d_{1,2}^{\text{ER}}/2$. Furthermore, if $\alpha_1 + \alpha_2 = d_{1,2}^{\text{ER}}$, it must be the case that $R_1 + R_2 \geq \Sigma(\mu_{1,2}^{\text{ER}})$. But this is precisely (5.15c) for $\mu_{1,2} = \mu_{1,2}^{\text{ER}}$.

Now, we turn to showing (5.15a). We have

$$\begin{aligned} \log L_1^{(n)} &\geq H(\tilde{M}_1^{(n)}) \geq H(\tilde{M}_1^{(n)}|\tilde{M}_2^{(n)}) \\ &= H(\widetilde{G}_1^{(n)}, \tilde{M}_1^{(n)}|\tilde{M}_2^{(n)}) - H(\widetilde{G}_1^{(n)}|\tilde{M}_1^{(n)}, \tilde{M}_2^{(n)}) \\ &\overset{(a)}{=} H(\widetilde{G}_1^{(n)}|\tilde{M}_2^{(n)}) \\ &\overset{(b)}{\geq} H(\widetilde{G}_1^{(n)}|\widetilde{G}_2^{(n)}) \\ &= H(\widetilde{G}_{1,2}^{(n)}) - H(\widetilde{G}_2^{(n)}), \end{aligned} \tag{5.37}$$

where (a) uses the facts that $\tilde{M}_1^{(n)}$ is a function of $\widetilde{G}_1^{(n)}$ and also, since $\widetilde{G}_{1,2}^{(n)} \in \mathcal{A}^{(n)}$, given $\tilde{M}_1^{(n)}$ and $\tilde{M}_2^{(n)}$ we can unambiguously determine $\widetilde{G}_{1,2}^{(n)}$ and hence $\widetilde{G}_1^{(n)}$. Also, (b) uses data processing inequality. Now, we find an upper bound for $H(\widetilde{G}_2^{(n)})$. Note that since $\widetilde{G}_{1,2}^{(n)} \in \mathcal{A}^{(n)}$ with probability one, we have

$$H(\widetilde{G}_2^{(n)}) \leq \log |\mathcal{A}_2^{(n)}|, \tag{5.38}$$

where

$$\mathcal{A}_2^{(n)} := \{H_2^{(n)} \in \mathcal{G}_2^{(n)} : H_1^{(n)} \oplus H_2^{(n)} \in \mathcal{A}^{(n)} \text{ for some } H_1^{(n)} \in \mathcal{G}_1^{(n)}\}.$$

Now, take $H_2^{(n)} \in \mathcal{A}_2^{(n)}$ and let $H_1^{(n)} \in \mathcal{G}_1^{(n)}$ be such that $H_{1,2}^{(n)} := H_1^{(n)} \oplus H_2^{(n)} \in \mathcal{A}^{(n)}$. Since $\mathcal{A}^{(n)} \subseteq \mathcal{G}_{\vec{p},\vec{q}}^{(n)}$, by definition we have that, for all $x \in \Xi_{1,2}$ and all $\theta \in \Theta_{1,2}$,

$$\sum_{x \in \Xi_{1,2}} |m_{H_{1,2}^{(n)}}(x) - np_x/2| \leq n^{2/3} \text{ and } \sum_{\theta \in \Theta_{1,2}} |u_{H_{1,2}^{(n)}}(\theta) - nq_\theta| \leq n^{2/3}.$$

Moreover, for $x_2 \in \Xi_2$ and $\theta_2 \in \Theta_2$ we have $m_{H_2^{(n)}}(x_2) = \sum_{x_1 \in \Xi_1 \cup \{\circ_1\}} m_{H_{1,2}^{(n)}}((x_1, x_2))$ and $u_{H_2^{(n)}}(\theta_2) = \sum_{\theta_1 \in \Theta_1} m_{H_{1,2}^{(n)}}((\theta_1, \theta_2))$. Using this in the above and using the triangle inequality, we realize that for $H_2^{(n)} \in \mathcal{A}_2^{(n)}$ we have $\vec{m}_{H_2^{(n)}} \in \mathcal{M}_2^{(n)}$ and $\vec{u}_{H_2^{(n)}} \in \mathcal{U}_2^{(n)}$, where $\mathcal{M}_2^{(n)}$ is the set of edge mark count vectors $\vec{m}$ such that $\sum_{x_2 \in \Xi_2} |m(x_2) - np_{x_2}/2| \leq n^{2/3}$ and $\mathcal{U}_2^{(n)}$ is the set of vertex mark count vectors $\vec{u}$ such that $\sum_{\theta_2 \in \Theta_2} |u(\theta_2) - nq_{\theta_2}| \leq n^{2/3}$. Consequently, we have

$$|\mathcal{A}_2^{(n)}| \leq (2n^{2/3} + 1)^{(|\Xi_2| + |\Theta_2|)} \left(\max_{\vec{m} \in \mathcal{M}_2^{(n)}} \binom{\binom{n}{2}}{\{m(x_2)\}_{x_2 \in \Xi_2}} \right) \left(\max_{\vec{u} \in \mathcal{U}_2^{(n)}} \binom{n}{\{u(\theta_2)\}_{\theta_2 \in \Theta_2}} \right).$$

Using Lemma 5.1 and the definition of $\mathcal{M}_2^{(n)}$ and $\mathcal{U}_2^{(n)}$ above, with $Q = (Q_1, Q_2) \sim \vec{q}$, an argument similar to the one that was used to establish (5.19) implies that

$$\limsup_{n \to \infty} \frac{\log |\mathcal{A}_2^{(n)}| - n\frac{d_2^{\text{ER}}}{2} \log n}{n} \leq H(Q_2) + \sum_{x_2 \in \Xi_2} s(p_{x_2}) = \Sigma(\mu_2^{\text{ER}}).$$

Substituting this into (5.38), we get

$$\limsup_{n \to \infty} \frac{\log H(\widetilde{G}_2^{(n)}) - n\frac{d_2^{\text{ER}}}{2} \log n}{n} \leq \Sigma(\mu_2^{\text{ER}}).$$

Using this together with (5.34) and substituting into (5.37) we get

$$\liminf_{n \to \infty} \frac{\log L_1^{(n)} - n\frac{d_{1,2}^{\text{ER}} - d_2^{\text{ER}}}{2} \log n}{n} \geq \Sigma(\mu_{1,2}^{\text{ER}}) - \Sigma(\mu_2^{\text{ER}}) = \Sigma(\mu_1^{\text{ER}} | \mu_2^{\text{ER}}).$$

Using a similar method as in (5.35) and (5.36), this implies (5.15a). The proof of (5.15b) is similar. This completes the proof of the converse for the Erdős–Rényi case.

5.3.4 Proof of the Converse for the Configuration Model

The proof of the converse for the configuration model is similar to that for the Erdős–Rényi model presented in the previous section. Take an achievable rate tuple $(\alpha_1, R_1, \alpha_2, R_2)$ together with a sequence of $\langle n, L_1^{(n)}, L_2^{(n)} \rangle$ codes $(f_1^{(n)}, f_2^{(n)}, g^{(n)})$ achieving this rate tuple. Moreover, redefine the set $\mathcal{A}^{(n)}$ to be

$$\mathcal{A}^{(n)} := \mathcal{W}^{(n)} \cap \{H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)} : g^{(n)}(f_1^{(n)}(H_1^{(n)}), f_2^{(n)}(H_2^{(n)})) = H_{1,2}^{(n)}\}, \tag{5.39}$$

where the set $\mathcal{W}^{(n)}$ was defined in Section 5.3.2. Now, let $G_{1,2}^{(n)} \sim \mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$ be distributed according to the configuration model ensemble, and let $\widetilde{G}_{1,2}^{(n)} \in \mathcal{A}^{(n)}$ have the distribution obtained from that of $G_{1,2}^{(n)}$ by conditioning on it lying in the set $\mathcal{A}^{(n)}$. Note that the normalizing constant $\pi_n := \mathbb{P}(G_{1,2}^{(n)} \in \mathcal{A}^{(n)})$ goes to 1 as $n \to \infty$ since $\mathbb{P}(G_{1,2}^{(n)} \in \mathcal{W}^{(n)}) \to 1$ and the error probability of the code, $P_e^{(n)}$, vanishes. Moreover, let $P_{\text{CM}}^{(n)}$ and $\tilde{P}_{\text{CM}}^{(n)}$ be the laws of $G_{1,2}^{(n)}$ and $\widetilde{G}_{1,2}^{(n)}$, respectively. In the following, we show that

$$\liminf_{n\to\infty} \frac{H(\widetilde{G}_{1,2}^{(n)}) - n\frac{d_{1,2}^{\text{CM}}}{2}\log n}{n} \geq \Sigma(\mu_{1,2}^{\text{CM}}), \tag{5.40}$$

and

$$\limsup_{n\to\infty} \frac{H(\widetilde{G}_{2}^{(n)}) - n\frac{d_{2}^{\text{CM}}}{2}\log n}{n} \leq \Sigma(\mu_{2}^{\text{CM}}). \tag{5.41}$$

The rest of the proof is then identical to that of the previous section, so we only focus on proving the statements in (5.40) and (5.41).

For (5.40), note that for $H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$ such that $\vec{\text{dg}}_{H_{1,2}^{(n)}} \in \mathcal{D}^{(n)}$, where $\mathcal{D}^{(n)}$ was defined in Section 5.3.2, we have

$$-\log P_{\text{CM}}^{(n)}(H_{1,2}^{(n)}) = \log \binom{n}{\{c_k(\vec{d}^{(n)})\}_{k=0}^{\Delta}} + \log|\mathcal{G}_{\vec{d}^{(n)}}^{(n)}| - \sum_{x\in\Xi_{1,2}} m_{H_{1,2}^{(n)}}(x)\log\gamma_x - \sum_{\theta\in\Theta_{1,2}} u_{H_{1,2}^{(n)}}(\theta)\log q_\theta.$$

Now, if $H_{1,2}^{(n)} \in \mathcal{W}^{(n)}$, using the definition of $\mathcal{W}^{(n)}$ we realize that there exists a constant $c > 0$ such that

$$\begin{aligned} -\log P_{\text{CM}}^{(n)}(H_{1,2}^{(n)}) \geq \log \binom{n}{\{c_k(\vec{d}^{(n)})\}_{k=0}^{\Delta}} + \log|\mathcal{G}_{\vec{d}^{(n)}}^{(n)}| - \sum_{x\in\Xi_{1,2}} m_n\gamma_x\log\gamma_x \\ - \sum_{\theta\in\Theta_{1,2}} nq_\theta\log q_\theta - cn^{2/3} =: K_n. \end{aligned}$$

Note that the right hand side is a constant independent of $H_{1,2}^{(n)}$ and is denoted by K_n. Since $\widetilde{G}_{1,2}^{(n)}$ falls in $\mathcal{W}^{(n)}$ with probability one, this means that $H(\widetilde{G}_{1,2}^{(n)}) \geq \log\pi_n + K_n$. But $\pi_n \to 1$ as $n \to \infty$. Therefore, using the assumption (5.10) together with (5.26) from Section 5.3.2 and also the fact that $m_n/n \to d_{1,2}^{\text{CM}}/2$, we realize that

$$\liminf_{n\to\infty} \frac{H(\widetilde{G}_{1,2}^{(n)}) - n\frac{d_{1,2}^{\text{CM}}}{2}\log n}{n} \geq H(X) - s(d_{1,2}^{\text{CM}}) - \mathbb{E}\left[\log X!\right] + \frac{d_{1,2}^{\text{CM}}}{2}H(\Gamma) + H(Q),$$

where $X \sim \vec{r}$, $\Gamma \sim \vec{\gamma}$ and $Q \sim \vec{q}$. Note that the right hand side is precisely $\Sigma(\mu_{1,2}^{\text{CM}})$. Hence we have proved (5.40).

In order to show (5.41), note that $H(\widetilde{G}_2^{(n)}) \le \log |\mathcal{A}_2^{(n)}|$ where $\mathcal{A}_2^{(n)}$ consists of graphs $H_2^{(n)} \in \mathcal{G}_2^{(n)}$ such that, for some $H_1^{(n)} \in \mathcal{G}_1^{(n)}$, we have $H_1^{(n)} \oplus H_2^{(n)} \in \mathcal{A}^{(n)}$. Since $\mathcal{A}^{(n)} \subseteq \mathcal{W}^{(n)}$, we have for all $H_2^{(n)} \in \mathcal{A}_2^{(n)}$ that

$$\sum_{x_2 \in \Xi_2} |m_{H_2^{(n)}}(x_2) - m_n \gamma_{x_2}| \le n^{2/3} \text{ and } \sum_{\theta_2 \in \Theta_2} |u_{H_2^{(n)}}(\theta_2) - n q_{\theta_2}| \le n^{2/3}. \tag{5.42}$$

On the other hand, the condition (5.24) implies that $\overrightarrow{\mathrm{dg}}_{H_2^{(n)}} \in \mathcal{D}_2^{(n)}$ where $\mathcal{D}_2^{(n)}$ denotes the set of degree sequences $\vec{d}$ of size n with elements bounded by Δ such that

$$|c_k(\vec{d}) - n\mathbb{P}\left(X_2 = k\right)| \le (\Delta + 1)n^{2/3}, \qquad \forall 0 \le k \le \Delta, \tag{5.43}$$

where X_2 is the random variable defined in (5.13). Consequently, we have

$$\begin{aligned} \log |\mathcal{A}_2^{(n)}| \le \log |\mathcal{D}_2^{(n)}| + \max_{\vec{d} \in \mathcal{D}_2^{(n)}} \log |\mathcal{G}_{\vec{d}}^{(n)}| + \max_{H_2^{(n)} \in \mathcal{A}_2^{(n)}} \log \binom{\sum_{x_2 \in \Xi_2} m_{H_2^{(n)}}(x_2)}{\{m_{H_2^{(n)}}(x_2)\}_{x_2 \in \Xi_2}} \\ + \max_{H_2^{(n)} \in \mathcal{A}_2^{(n)}} \log \binom{n}{\{u_{H_2^{(n)}}(\theta_2)\}_{\theta_2 \in \Theta_2}}. \end{aligned} \tag{5.44}$$

Note that (5.43) implies that $|\mathcal{D}_2^{(n)}| \le (2(\Delta+1)n^{2/3}+1)^{\Delta+1} \max_{\vec{d} \in \mathcal{D}_2^{(n)}} \binom{n}{\{c_k(\vec{d})\}_{k=0}^{\Delta}}$. Therefore, Lemma 5.1 implies that

$$\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{D}_2^{(n)}| \le H(X_2). \tag{5.45}$$

On the other hand, the assumptions $r_0 < 1$ and (5.8) imply that $d_2^{\mathrm{CM}} > 0$. Hence, using Lemma 5.3, we have

$$\limsup_{n \to \infty} \frac{\max_{\vec{d} \in \mathcal{D}_2^{(n)}} \log |\mathcal{G}_{\vec{d}}^{(n)}| - n \frac{d_2^{\mathrm{CM}}}{2} \log n}{n} \le -s(d_2^{\mathrm{CM}}) - \mathbb{E}\left[\log X_2!\right]. \tag{5.46}$$

Moreover, if $H_2^{(n)}$ is a sequence in $\mathcal{A}_2^{(n)}$, from (5.42), for all $x_2 \in \Xi_2$, we have

$$\lim_{n \to \infty} \frac{m_{H_2^{(n)}}(x_2)}{\sum_{x_2' \in \Xi_2} m_{H_2^{(n)}}(x_2')} = \frac{\gamma_{x_2}}{\sum_{x_2' \in \Xi_2} \gamma_{x_2'}} = \mathbb{P}\left(\Gamma_2 = x_2 | \Gamma_2 \neq \circ_2\right),$$

where $\Gamma = (\Gamma_1, \Gamma_2)$ has law $\vec{\gamma}$. Additionally, we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{x_2 \in \Xi_2} m_{H_2^{(n)}}(x_2) = \frac{d_2^{\mathrm{CM}}}{2}.$$

Thereby, from Lemma 5.1, we have

$$\limsup_{n \to \infty} \frac{1}{n} \max_{H_2^{(n)} \in \mathcal{A}_2^{(n)}} \log \binom{\sum_{x_2 \in \Xi_2} m_{H_2^{(n)}}(x_2)}{\{m_{H_2^{(n)}}(x_2)\}_{x_2 \in \Xi_2}} \le \frac{d_2^{\mathrm{CM}}}{2} H(\Gamma_2 | \Gamma_2 \neq \circ_2). \tag{5.47}$$

Finally, as we have $u_{H_2^{(n)}}(\theta_2)/n \to q_{\theta_2}$ for all $\theta_2 \in \Theta_2$, another usage of Lemma 5.1 implies that

$$\limsup_{n\to\infty} \frac{1}{n} \max_{H_2^{(n)} \in \mathcal{A}_2^{(n)}} \log \binom{n}{\{u_{H_2^{(n)}}(\theta_2)\}_{\theta_2\in\Theta_2}} \le H(Q_2), \tag{5.48}$$

where $Q = (Q_1, Q_2)$ has law $\vec{q}$. Now, combining (5.45), (5.46), (5.47) and (5.48) and substituting into (5.44), and also using the bound $H(\widetilde{G}_2^{(n)}) \le \log |\mathcal{A}_2^{(n)}|$, we realize that

$$\limsup_{n\to\infty} \frac{H(\widetilde{G}_2^{(n)}) - n\frac{d_2^{\mathrm{CM}}}{2}\log n}{n} \le H(X_2) - s(d_2^{\mathrm{CM}}) - \mathbb{E}\left[\log X_2!\right] + \frac{d_2^{\mathrm{CM}}}{2} H(\Gamma_2|\Gamma_2 \neq \circ_2) + H(Q_2).$$

But the right hand side is precisely $\Sigma(\mu_2^{\mathrm{CM}})$. This completes the proof of (5.41). As was mentioned before, the rest of the proof is identical to that in the previous section.

5.3.5 Generalization to more than two sources

Assume we have $k \ge 2$ sources of graphical data. For $1 \le i \le k$, let Θ_i and Ξ_i denote the vertex and edge mark sets for the ith domain. For $i \in [k]$ and $n \in \mathbb{N}$, $\mathcal{G}_i^{(n)}$ denotes the set of marked graphs on the vertex set $[n]$ with vertex and edge marks coming from Θ_i and Ξ_i, respectively. Given $A \subseteq [k]$ nonempty and for $G_i \in \mathcal{G}_i^{(n)}$, $i \in A$, we define $\bigoplus_{i\in A} G_i$ to be the superposition of graphs in A, which is a simple marked graph on the vertex set $[n]$ such that a vertex $v \in [n]$ carries a vertex mark $(\theta_i : i \in A) \in \Theta_A := \prod_{i\in A} \Theta_i$ such that θ_i is the mark of v in G_i. Moreover, an edge between vertices v and w exists in $\bigoplus_{i\in A} G_i$ if such an edge exists in at least one of the graphs G_i, $i \in A$. If this is the case, the mark of this edge is defined to be $(x_i : i \in A)$, where for $i \in A$, x_i is the mark of the edge (v, w) in G_i if such an edge exists in G_i. Otherwise, we set $x_i = \circ_i$, where $\circ_i$ for $i \in [k]$ is an auxiliary mark not present in Ξ_i. For nonempty $A \subseteq [k]$, we denote $(\circ_i : i \in A)$ by $\circ_A$. Note that with $\Xi_A := (\prod_{i\in A}(\Xi_i \cup \{\circ_i\})) \setminus \{\circ_A\}$, $\bigoplus_{i\in A} G_i$ is a marked graph with vertex and edge mark sets Θ_A and Ξ_A, respectively. Let $\mathcal{G}_A^{(n)}$ denote the set of marked graphs in domain A, which is the set of marked graphs on the vertex set $[n]$ together with vertex and edge mark sets Θ_A and Ξ_A, respectively. Given $G \in \mathcal{G}_{[k]}^{(n)}$ and $A \subset [k]$, we can naturally define the projection of G onto domain A by projecting all vertex and edge marks onto Θ_A and Ξ_A, respectively, followed by removing edges with mark $\circ_A$. It can be checked that the resulting graph, denoted by G_A, lies in domain A, i.e. $G_A \in \mathcal{G}_A^{(n)}$.

A sequence of $\langle n, L_i^{(n)} : i \in [k]\rangle$ codes is defined as a sequence of tuples $((f_i^{(n)} : i \in [k]), g^{(n)})$ such that $f_i^{(n)} : \mathcal{G}_i^{(n)} \to [L_i^{(n)}]$ for $i \in A$ are encoding functions, and $g^{(n)} : \prod_{i\in[k]}[L_i^{(n)}] \to \mathcal{G}_{[k]}^{(n)}$ is the corresponding decoding function. Given a sequence of ensembles $G_{[k]}^{(n)}$ on $\mathcal{G}_{[k]}^{(n)}$, the probability of error $P_e^{(n)}$ is defined to be the probability that $g^{(n)}((f_i^{(n)}(G_i^{(n)}) : i \in [k])) \neq G_{[k]}^{(n)}$.

We say that a rate tuple $((\alpha_i, R_i) : i \in [k])$ is achievable for the distributed compression of the sequence of random graphs $G_{[k]}^{(n)} \in \mathcal{G}_{[k]}^{(n)}$ if there is a sequence of $\langle n, L_i^{(n)} : i \in [k]\rangle$ codes such that for $i \in [k]$,

$$\limsup_{n\to\infty} \frac{\log L_i^{(n)} - (\alpha_i n \log n + R_i n)}{n} \leq 0,$$

and also $P_e^{(n)} \to 0$. We say that $((\alpha_i, R_i) : i \in [k])$ lies in the rate region $\mathcal{R}$ if there exist sequences $R_i^{(m)}$ for $i \in [k]$ such that $R_i^{(m)} \to R_i$ as $m \to \infty$ and, for each m, $((\alpha_i, R_i^{(m)}) : i \in [k])$ is achievable.

We can naturally generalize the Erdős–Rényi and the configuration model ensembles of Section 5.1 to the above setting.

A sequence of Erdős–Rényi ensembles: Given a sequence of nonnegative real numbers $\vec{p} = \{p_x\}_{x\in\Xi_{[k]}}$ and a probability distribution $\vec{q} = \{q_\theta\}_{\theta\in\Theta_{[k]}}$, assume that for all $i \in [k]$ and $x_i \in \Xi_i$ we have

$$\sum_{(x'_j : j\in[k])\in\Xi_{[k]} : x'_i = x_i} p_{(x'_j : j\in[k])} > 0. \tag{5.49}$$

Moreover, assume that for all $i \in [k]$ and all $\theta_i \in \Theta_i$ we have

$$\sum_{(\theta'_i : i\in[k])\in\Theta_{[k]} : \theta'_i = \theta_i} q_{(\theta'_i : i\in[k])} > 0. \tag{5.50}$$

For $n \in \mathbb{N}$ large enough, the probability distribution $\mathcal{G}(n; \vec{p}, \vec{q})$ on $\mathcal{G}_{[k]}^{(n)}$ is defined as follows: for each pair of vertices $1 \leq i < j \leq n$, the edge (i, j) exists and has a mark $x \in \Xi_{[k]}$ with probability p_x/n, and is not present with probability $1 - \sum_{x\in\Xi_{[k]}} p_x/n$. Moreover, each vertex is independently given a mark $\theta \in \Theta_{[k]}$ with probability q_θ. The choices of edge and vertex marks are done independently.

The conditions in (5.49) and (5.50) are required only to ensure that the sets of vertex marks and edge marks are chosen to be as small as possible, and these conditions could be relaxed if desired.

A sequence of configuration model ensembles: Similar to the configuration model ensemble for two sources as we defined in Section 5.1, assume that $\Delta \in \mathbb{N}$ and a probability distribution $\vec{r} = \{r_i\}_{i=0}^{\Delta}$ is given such that $r_0 < 1$. Moreover, for each n, the degree sequence $\vec{d}^{(n)} = \{d^{(n)}(1), \dots, d^{(n)}(n)\}$ is given such that for $i \in [n]$, $d^{(n)}(i) \leq \Delta$, $\sum_{i=1}^n d^{(n)}(i)$ is even, and (5.10) is satisfied. Additionally, assume that probability distributions $\vec{\gamma} = \{\gamma_x\}_{x\in\Xi_{[k]}}$ and $\vec{q} = \{q_\theta\}_{\theta\in\Theta_{[k]}}$ are given such that for all $i \in [k]$ and $x_i \in \Xi_i$ we have

$$\sum_{(x'_j : j\in[k])\in\Xi_{[k]} : x'_i = x_i} \gamma_{(x'_j : j\in[k])} > 0, \tag{5.51}$$

and for all $A \subset [k]$ nonempty, $A \neq [k]$, we have

$$\sum_{(x'_j : j\in[k])\in\Xi_{[k]} : (x'_i : i\in A) = \circ_A} \gamma_{(x'_j : j\in[k])} > 0. \tag{5.52}$$

We also assume that for all $i \in [k]$ and $\theta_i \in \Theta_i$ we have

$$\sum_{(\theta'_j : j \in [k]) \in \Theta_{[k]} : \theta'_i = \theta_i} q_{(\theta'_j : j \in [k])} > 0. \tag{5.53}$$

With these, for n large enough, we define the probability distribution $\mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$ on $\mathcal{G}_{[k]}^{(n)}$ as follows. Similar to the ensemble for two sources, we pick an unmarked graph on the vertex set $[n]$ uniformly at random among the set of graphs with maximum degree Δ such that for $0 \leq k \leq \Delta$, $c_k(\overrightarrow{\text{dg}}_G) = c_k(\vec{d}^{(n)})$. Then, we assign i.i.d. marks with law $\vec{\gamma}$ on the edges and i.i.d. marks with law $\vec{q}$ on the vertices.

The conditions in (5.51) and (5.53) are required only to ensure that the sets of vertex marks and edge marks are chosen to be as small as possible, and these conditions could be relaxed if desired. However, the conditions in (5.52) are essential, since they ensure that for all $A \subset [k]$ nonempty, $A \neq [k]$, the underlying unmarked graph of the projection of the overall graph onto domain A is not a subgraph of the underlying unmarked graph of the projection onto domain A^c.

Similar to our discussion in Section 5.2, it can be seen that the local weak limit of the sequence of Erdős–Rényi ensembles above is a marked Poisson Galton–Watson tree, which we denote by $\mu_{[k]}^{\text{ER}}$. Likewise, the local weak limit of the sequence of configuration model ensembles above is a marked Galton–Watson tree with degree distribution $\vec{r}$, which we denote by $\mu_{[k]}^{\text{CM}}$. For $A \subseteq [k]$ nonempty, we denote the projection of $\mu_{[k]}^{\text{ER}}$ and $\mu_{[k]}^{\text{CM}}$ to domain A by μ_A^{ER} and μ_A^{CM}, respectively. For nonempty $A \subset [k]$, $A \neq [k]$, we define $\Sigma(\mu_A^{\text{ER}} | \mu_{A^c}^{\text{ER}})$ to be $\Sigma(\mu_{[k]}^{\text{ER}}) - \Sigma(\mu_{A^c}^{\text{ER}})$. We similarly define $\Sigma(\mu_A^{\text{CM}} | \mu_{A^c}^{\text{CM}})$.

We are now ready to characterize the rate region for the multi-source scenarios above in the following Theorem 5.2. This is a generalization of Theorem 5.1, and its proof is similar to that of Theorem 5.1. We highlight the proof of Theorem 5.2 in Appendix D.6.

Theorem 5.2. *Assume $\mu_{[k]}$ is either of the two distributions $\mu_{[k]}^{ER}$ or $\mu_{[k]}^{CM}$ defined above. Then, if $\mathcal{R}$ is the rate region for the sequence of ensembles corresponding to $\mu_{[k]}$, as defined above, a rate tuple $((\alpha_i, R_i) : i \in [k]) \in \mathcal{R}$ if and only if for every nonempty $A \subset [k]$, $A \neq [k]$, we have*

$$\left(\sum_{i \in A} \alpha_i, \sum_{i \in A} R_i \right) \succeq ((d_{[k]} - d_{A^c})/2, \Sigma(\mu_A | \mu_{A^c})),$$

and

$$\left(\sum_{i \in [k]} \alpha_i, \sum_{i \in [k]} R_i \right) \succeq (d_{[k]}/2, \Sigma(\mu_{[k]})),$$

where $d_{[k]} = \deg(\mu_{[k]})$ and $d_{A^c} = \deg(\mu_{A^c})$.

5.4 Conclusion

We gave a counterpart of the Slepian–Wolf Theorem for distributed compression of graphical data, employing the framework of local weak convergence. We derived the rate region for two families of sequences of graph ensembles, namely sequences of Erdős–Rényi ensembles having a local weak limit and sequences of configuration model ensembles having a local weak limit. Furthermore, we gave a generalization of this result for Erdős–Rényi and configuration model ensembles with more than two sources.

Part III

Load Balancing

Chapter 6

Asymptotic Behavior of Load Balancing in Hypergrphas

So far, in Chapters 4 and 5, we have studied compression for sparse graphical data. We did so by viewing the local weak convergence framework as a counterpart of the notion of stochastic processes for sparse graphical data. This viewpoint suggests that the applicability of this framework should not be considered as being limited to the problem of graphical data compression. In fact, this framework can potentially be employed in any context involving sparse graphical data. In particular, in this chapter, we employ this framework to study the problem of load balancing. A load balancing network consists of a set of tasks and a set of servers. Each task has an amount of load which can be distributed among a subset of the servers that are accessible to that task. In order to ensure the efficiency of the network, it is important to balance the load among the servers. When the problem size is large, it may be expensive to compute the detailed characteristics of an optimal or sufficiently good allocation of the load. Instead, it is interesting to focus on the statistical characteristics of the allocation, such as the empirical distribution of the load faced by a typical server in the network. This chapter is concerned with developing such a viewpoint in the context of a specific kind of the load balancing problem which has broad applicability. Specifically, we build upon the notion of load balanced introduce by Hajek [Haj90].

The structure of this chapter is as follow. In Section 6.1, we present our model and discuss the prior work. In particular, we model the load balancing network as a hypergraph. In Section 6.2, we set up our notation, and discuss the extension of the local weak convergence framework to hypergraphs. We then state our main results in Section 6.3, and discuss the proof techniques and details in the subsequent sections.

6.1 Model and Prior Work

We model the load balancing problem by a bipartite graph in which every node on the right represents a task and every node on the left represents a server. Each server is accessible to a certain subset of the tasks. Equivalently each task has access to only a certain subset of the servers. We view the bipartite graph as a hypergraph, with each vertex of the hypergraph representing a server and each hyperedge representing a task. The vertices of a hyperedge are then the servers that are accessible to it. Let $\{v_1, \dots, v_n\}$ and $\{e_1, \dots, e_m\}$ denote the set of servers and tasks, or equivalently vertices and hyperedges, respectively. Therefore, $v_i \in e_j$ means that server v_i can be used to contribute to the performance of task e_j. See Figure 6.1 for an example. In general, we might want to consider the scenario where task e_j has an amount of load equal to l_j, which could be arbitrarily allocated among the servers $v_i \in e_j$. For simplicity, we will consider in this chapter only the case where all the l_j equal 1, but we leave the discussion general for the moment. Let θ be an allocation of the load of tasks among the servers, i.e. $\theta(e_j, v_i)$ is the amount of load coming from task e_j assigned to server v_i. Hence, $\theta(e_j, v_i) \geq 0$ and $\sum_{v_i \in e_j} \theta(e_j, v_i) = l_j$. For a server v_i, let $\partial\theta(v_i)$ be the total amount of load assigned to v_i, i.e. $\partial\theta(v_i) = \sum_{e_j : v_i \in e_j} \theta(e_j, v_i)$.

This formulation of load balancing was studied by Hajek [Haj90] who, in particular, formulated the notion of a *balanced allocation*. It is natural to expect that a task would be happier to use servers that are currently handling less load, if available, as opposed to those handling more load. An allocation θ is said to be balanced if no task desires to change the allocation of its load. For finite load balancing problems, this turns out to be equivalent to the statement that the allocation minimizes $\sum_i f(\partial\theta(v_i))$ for any given fixed strictly convex function f. One can think of $\sum_i f(\partial\theta(v_i))$ as the aggregate cost we need to pay to process all the tasks. Hajek showed the existence of balanced allocations and uniqueness of the total load at nodes under any balanced allocation, and suggested algorithms to find a balanced allocation. It is particularly remarkable that the notion of a balanced allocation does not depend on the specific choice of the strictly convex cost function f.

With the aim of understanding the statistical characteristics of balanced allocations in large load balancing problems, Hajek assumed that each task could be performed by only two servers – hence the underlying hypergraph reduces to a graph – and he assumed that each edge in this graph carries one unit of load. He then studied such a load balancing problem in large random graphs [Haj90]. In particular, Hajek considered the sparse Erdős–Rényi model to generate these graphs, where αn edges are distributed among n vertices uniformly at random, with α being a fixed parameter. Recall from Chapter 2 that the asymptotic structure of the local neighborhood of a typical vertex in a sparse Erdős–Rényi model is given by a Poisson Galton–Watson tree. This suggests that the behavior of balanced allocations in Galton–Watson trees might be a good proxy for the load distribution in large Erdős–Rényi graphs. Hajek conjectured that the recursive nature of a Galton–Watson process helps one analyze the distribution of balanced allocations by studying fixed point equations. He was even able to suggest the form of the fixed point equation for the Poisson Galton–Watson tree. However, it turns out that this approach is more subtle than it looks. For one thing, Hajek realized that the notion of balanced allocation in an infinite graph as a proxy for large graphs is not well defined [Haj96]. See Figure 6.2 for an example.

Hajek's conjecture for the graph regime (i.e. when each task could only be distributed

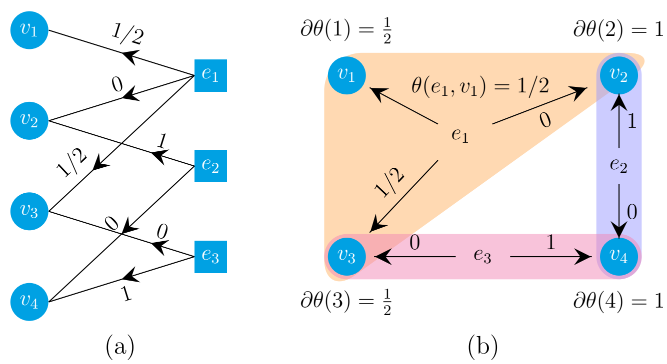

Figure 6.1: Load balancing with 3 tasks and 4 servers. (a) illustrates the bipartite representation together with an allocation. While the load of e_1 could be served by nodes in $\{v_1, v_2, v_3\}$, half of its load is being assigned to v_1 and the other half to v_3, i.e. $\theta(e_1, v_1) = 1/2$, $\theta(e_1, v_2) = 0$ and $\theta(e_1, v_3) = 1/2$. (b) shows the hypergraph representation. The total load at a node i is denoted by $\partial\theta(i)$. For instance, $\partial\theta(v_3) = \theta(e_1, v_3) + \theta(e_3, v_3) = 1/2$.

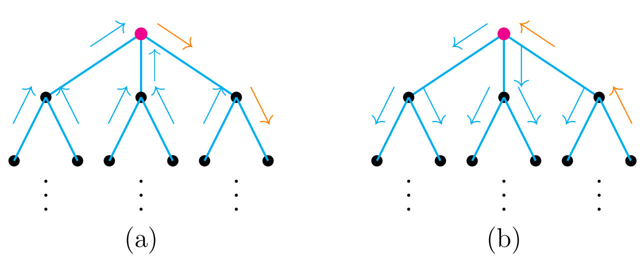

Figure 6.2: Hajek's example to show non–uniqueness of the load for infinite graphs. Consider the rooted 3–regular graph with infinite depth as shown. We send all of the unit load corresponding to each edge in the direction of the shown arrows. The red path goes to infinity in both pictures. The allocation in (a) makes the total load at every vertex equal to 2 while that in (b) makes the total load at all vertices equal to 1. Therefore, both are balanced.

among two servers) was settled by Anantharam and Salez [AS16]. They achieved this by employing the framework of local weak convergence that we discussed in Chapter 2. In particular, they settled Hajek's conjecture by first defining a notion of balancedness for unimodular random rooted graphs. Moreover, they showed that if a sequence of finite graphs G_n converges to a random rooted graph in the above local weak sense, the total load associated to a balanced allocation at a vertex chosen uniformly at random in G_n converges in distribution to the total load associated to the balanced allocation at the root of the limit. Additionally, they managed to express a certain functional of the distribution of the load at the root of the Galton–Watson tree in terms of a fixed point distributional equation, settling Hajek's conjecture in the graph regime. Beyond this, they also proved the convergence of the maximum load for a sequence of finite graphs resulting from a certain configuration model to that of their local weak limit, under some additional conditions.

6.1.1 Our Contributions

We study the above load balancing problem in the more general regime where each task could have access to more than two servers, i.e. the underlying network is a hypergraph instead of a graph.

Our machinery for deriving results analogous to those in the graph regime will be a generalized method of local weak convergence on hypergraphs. One novelty of our development is to introduce a notion analogous to unimodularity for processes on random rooted hypergraphs. We believe that this generalized framework could be of independent interest in a variety of problems in which the underlying model is best expressed in terms of hypergraphs rather than graphs.

In particular, we prove that for any unimodular probability distribution on the set of rooted hypergraphs with finite expected degree, there exists a balanced allocation which is consistent with the local weak limit theory, i.e. the load distribution of a sequence of hypergraphs converges to that of the limit. For a special class of branching process on rooted hypertrees which is a generalization of Galton–Watson processes, we show that the distribution of the load at the root can be specified via a fixed point distributional equation. Finally, we study the convergence of the maximum load for a sequence of random hypergraphs generated from a configuration model to that of the limit, under some additional conditions.

6.2 Prerequisites and Notation

In this section, we set up our notation, and discuss our extension of the local weak convergence for hypergraphs. Throughout this chapter, $\mathbb{R}_{\geq 0}$ denotes the set of nonnegative real numbers. Moreover, $\mathbb{Q}$ denotes the set of rational numbers. $\mathbb{N}$ denotes the set of positive integers and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For a real number $x \in \mathbb{R}$, we denote $\max\{x, 0\}$ by x^+, and we denote $\min\{x^+, 1\}$ by $[x]_0^1$.

6.2.1 Hypergraphs

We work with simple hypergraphs defined on a countable vertex set, where each edge is a finite subset of the vertex set. For a hypergraph H, the sets of vertices and edges are denoted by $V(H)$ and $E(H)$, respectively. We write H as $\langle V, E \rangle$, where $V = V(H)$ and $E = E(H)$. We say a hypergraph is simple if $E(H) \subset 2^{V(H)}$. This means that in any edge each vertex can show up at most once, and that each subset of vertices can show up at most once as an edge. All hypergraphs appearing in this chapter will be simple, unless otherwise stated. For a vertex $i \in V(H)$, denote its degree by $\deg_H(i) := |\{e \in E(H) : e \ni i\}|$. For a given hypergraph H, let

$$\Psi(H) := \{(e, i) : e \in E(H), i \in e\} \tag{6.1}$$

denote the set of all edge-vertex pairs in the hypergraph.

Definition 6.1. *A hypergraph H is said to be locally finite if $\deg_H(i)$ is finite for all $i \in V(H)$ and e is finite for all $e \in E(H)$.*

Note that the above definition does not imply that there is a uniform bound on edge sizes or vertex degrees. Hence, a locally finite hypergraph can have arbitrarily large edges or vertex degrees. Throughout this chapter, we assume that all the hypergraphs are locally finite, unless otherwise stated. Thus, by default, the term hypergraph in this chapter means a simple, locally finite hypergraph on a countable vertex set.

For technical reasons, it is sometimes easier to work with bounded hypergraphs in proofs and then relax the boundedness condition.

Definition 6.2. *A hypergraph H is said to be bounded if $\deg_H(i) \leq \Delta$ for all $i \in V(H)$ and also $|e| \leq L$ for all $e \in E(H)$, for finite constants Δ and L.*

A path from node i to node j is an alternating vertex–edge sequence $i_0, e_1, i_1, e_2, i_2, \ldots, e_n, i_n$ with $i_0 = i$ and $i_n = j$, and $i_k \in e_{k+1}$ for $0 \leq k < n$. The length of such a path is defined to be n. The distance between vertices i and j, denoted by $d_H(i, j)$, is defined to be the length of the shortest path between i and j if $i \neq j$, and 0 when $i = j$. A path is called closed if $i_0 = i_n$.

Definition 6.3. *A hypergraph H is called a hypertree if there is no closed path*

$$i_0, e_1, i_1, \ldots, e_{n-1}, i_{n-1}, e_n, i_n$$

with $n \geq 2$ such that $i_j \neq i_l$ and $e_j \neq e_l$ for $1 \leq j \neq l \leq n$.

Remark 6.1. *Note that a hypertree need not be connected. It is straightforward to prove that if there is a path between vertices i and j in a hypertree, then the shortest path between these vertices is unique.*

Definition 6.4. *For a hypergraph H and a subset $W \subset V(H)$, define $E_H(W) := \{e \in E(H) : e \subset W\}$. $E_H(W)$ is comprised of the edges of H with all endpoints in the set W. For $i \in V(H)$ and $d \geq 0$, define $V_{i,d}^H := \{j \in V(H) : d_H(i,j) \leq d\}$ and $D_{i,d}^H := \{j \in V(H) : d_H(i,j) = d\}$. In particular, $V_{i,0}^H = D_{i,0}^H = \{i\}$.*

Definition 6.5. *A vertex rooted hypergraph is a hypergraph H with a distinguished vertex $i \in V(H)$. We denote this by (H,i). An edge-vertex rooted hypergraph is a hypergraph with a distinguished edge $e \in E(H)$ and a distinguished vertex $i \in V(H)$ such that $i \in e$. This is denoted by (H,e,i) .*

Definition 6.6. *We say that two hypergraphs H and H' are isomorphic and write $H \equiv H'$ when there is a bijection $\phi : V(H) \to V(H')$ such that $e \in E(H)$ if and only if $\phi(e) := \{\phi(j) : j \in e\} \in E(H')$. Also we say two vertex rooted hypergraphs (H,i) and (H',i') are isomorphic and write $(H,i) \equiv (H',i')$ if the above bijection ϕ exists and we have $\phi(i) = i'$ as well. Furthermore, we say two edge-vertex rooted hypergraphs (H,e,i) and (H',e',i') are isomorphic and write $(H,e,i) \equiv (H',e',i')$ if the above bijection exists, and we have $\phi(i) = i'$ and $\phi(e) = e'$.*

Instead of working with global isomorphisms as above, we can consider local isomorphisms, i.e. comparing two rooted hypergraphs up to some given depth.

Definition 6.7. *We say two vertex rooted hypergraphs (H,i) and (H',i') are isomorphic up to depth d and write $(H,i) \equiv_d (H',i')$ if their truncations up to depth d are isomorphic, i.e. $\langle V_{i,d}^H, E_H(V_{i,d}^H)\rangle \equiv \langle V_{i',d}^{H'}, E_{H'}(V_{i',d}^{H'})\rangle$ and also $\phi(i) = i'$, where $\phi : V_{i,d}^H \to V_{i',d}^{H'}$ is the vertex bijection establishing this isomorphism. Also, for $d \geq 1$, we say two edge-vertex rooted hypergraphs (H,e,i) and (H',e',i') are isomorphic up to depth d and write $(H,e,i) \equiv_d (H',e',i')$ if $\langle V_{i,d}^H, E_H(V_{i,d}^H)\rangle \equiv \langle V_{i',d}^{H'}, E_{H'}(V_{i',d}^{H'})\rangle$, $\phi(i) = i'$, and $\phi(e) = e'$, where $\phi(e) := \{\phi(j) : j \in e\}$. Here $\phi : V_{i,d}^H \to V_{i',d}^{H'}$ is the vertex bijection establishing this isomorphism.*

Definition 6.8. *Given two hypergraphs H and H', for $i \in V(H)$ and $i' \in V(H')$ we say that (H,i) has a local embedding up to depth $d \geq 1$ into (H',i') and write $(H,i) \hookrightarrow_d (H',i')$ if there is an injective mapping $\phi : V_{i,d}^H \hookrightarrow V_{i',d}^{H'}$ such that:*

1. *$\phi(i) = i'$, and*
2. *for all $e \in E_H(V_{i,d}^H)$, we have $\phi(e) \in E(H')$ where $\phi(e) := \{\phi(j) : j \in e\}$.*

Definition 6.9. *Given a hypergraph H, a node $i \in V(H)$ and $d \geq 1$, let $(H,i)_d$ denote the vertex rooted hypergraph $(\langle V_{i,d}^H, E_H(V_{i,d}^H)\rangle, i)$. In fact, $(H,i)_d$ is the d–neighborhood of vertex i.*

6.2.2 Balanced allocations on a hypergraph

Definition 6.10. *An allocation on hypergraph $H = \langle V, E\rangle$ is a mapping $\theta : \Psi(H) \to [0,1]$ such that $\theta(e,i)$ with $i \in e \in E$ tells us how much load from resource e is being given to node i. More formally, it is characterized by the properties:*

$$\theta(e,i) \geq 0 \ , \qquad \forall e \in E(H), i \in e, \ \textit{and}$$
$$\sum_{j \in e} \theta(e,j) = 1 \ , \qquad \forall e \in E(H) \ .$$

In any allocation, a given vertex $i \in V(H)$ receives a portion $\theta(e,i)$ of the total unit load of resource e. The total load at the vertex is then the sum of portions it receives from resources $e \ni i$. The following definition establishes the notation to discuss this load.

Definition 6.11. *Given an allocation θ on a hypergraph $H = \langle V, E\rangle$, define the function $\partial\theta : V(H) \to \mathbb{R}_{\geq 0}$ by*

$$\partial\theta(i) := \sum_{e: i \in e} \theta(e,i) \ , \qquad \textit{for all } i \in V(H).$$

Definition 6.12. *For a hypergraph $H = \langle V, E\rangle$, an allocation θ is called balanced if for all $e \in E$ and $i, j \in e$ we have*

$$\partial\theta(i) > \partial\theta(j) \quad \Rightarrow \quad \theta(e,i) = 0.$$

Much of the chapter is concerned with understanding the structure of balanced allocations on hypergraphs, and of the load resulting from such allocations. As we will soon see via examples, balanced allocations can exhibit phenomena analogous to phase transitions in statistical mechanics models. This is because the hard constraint defining balancedness can be thought of as analogous to a zero temperature limit. Following this analogy further, it is therefore convenient to deal with what might be called a positive temperature notion of balancedness, and then to send the temperature to zero. This is captured in the concept of ϵ–balance.

Definition 6.13. *For a hypergraph $H = \langle V, E\rangle$, an allocation θ is called ϵ–balanced, if for all $e \in E$ and $i \in e$ we have*

$$\theta(e,i) = \frac{\exp(-\partial\theta(i)/\epsilon)}{\sum_{j \in e} \exp(-\partial\theta(j)/\epsilon)}.$$

Remark 6.2. *Let θ be an ϵ-balanced allocation on a hypergraph H. Note that if $e \in E$ and $i, j \in e$ are such that $\partial\theta(i) > \partial\theta(j)$ then*

$$\theta(e,i) = \frac{\exp(-\partial\theta(i)/\epsilon)}{\sum_j \exp(-\partial\theta(j)/\epsilon)} \leq \frac{1}{1 + \exp\left(\frac{\partial\theta(i) - \partial\theta(j)}{\epsilon}\right)}.$$

Roughly speaking, if $\partial\theta(i) > \partial\theta(j)$ and ϵ is small, then $\theta(e,i) \approx 0$ and hence θ is approximately balanced. Also, roughly speaking, the smaller ϵ is, the more balanced an ϵ-balanced allocation is.

In the above, we defined balancedness when all the loads come from the edges of the hypergraph. We can generalize this to the case where, in addition to the internal load imposed by the edges, we have external load as well. External load is modeled by a function $b : V(H) \to \mathbb{R}$, called the *baseload function*. For a vertex $i \in V(H)$, $b(i)$ denotes the external load applied to node i. Throughout this chapter, we assume that each baseload function is bounded, but we do not assume a uniform bound on all baseload functions. More precisely, for each baseload function b on a given hypergraph H, we assume that there exists $M < \infty$ such that $|b(i)| < M$ for all $i \in V(H)$, where the constant M may depend on H and b. The concept of balancedness can be extended to the scenario with baseloads as follows.

Definition 6.14. *For a hypergraph $H = \langle V, E\rangle$, together with a baseload function $b : V(H) \to \mathbb{R}$, an allocation $\theta : \Psi(H) \to [0,1]$ is called balanced with respect to the baseload b, if for all $e \in E$ and $i, j \in e$ we have*

$$\partial\theta(i) + b(i) > \partial\theta(j) + b(j) \quad \Rightarrow \quad \theta(e,i) = 0.$$

Note that $\partial\theta(i) + b(i)$ is the total load at node i where $\partial\theta(i)$ is the internal load and $b(i)$ is the contribution from the external load. We use the notation $\partial_b\theta$ as a shorthand for $\partial\theta + b$.

The concept of an ϵ–balanced allocations can be similarly extended to the scenario with baseloads.

Definition 6.15. *For a given hypergraph $H = \langle V, E\rangle$, together with a baseload function $b : V(H) \to \mathbb{R}$, we say an allocation $\theta : \Psi(H) \to [0,1]$ is ϵ–balanced with respect to the baseload b, if for all $e \in E(H)$ and $i \in e$ we have*

$$\theta(e,i) = \frac{\exp\left(-\partial_b\theta(i)/\epsilon\right)}{\sum_{j\in e}\exp\left(-\partial_b\theta(j)/\epsilon\right)}. \tag{6.2}$$

It is known that if the hypergraph is finite, then balanced allocations exist with respect to any baseload and the resulting load vector is the same for all balanced allocations for the given baseload (see Theorem 2 and Corollary 5 in [Haj90]). This result is stated in the following proposition.

Proposition 6.1. *If $H = \langle V, E\rangle$ is a finite hypergraph, and $b : V(H) \to \mathbb{R}$ is a given baseload function, then there exists at least one balanced allocation θ on H with respect to the baseload b. Moreover, if θ and θ' are two balanced allocations on H with respect to b, then $\partial_b\theta(i) = \partial_b\theta'(i)$ for all $i \in V(H)$.*

Figure 6.3: A graph with 3 vertices and three edges and two different balanced allocations with zero baseload. Note that the total load at each vertex is the same for the two allocations and is equal to 1 at each vertex.

Later, in Section 6.4, we will study ϵ–balanced allocations with baseload for hypergraphs that are not necessarily finite. In particular, we will show in Corollary 6.4 therein that for bounded hypergraphs, for any baseload function, the total load at any vertex corresponding to any ϵ–balanced allocation is uniquely defined.

Note that for finite hypergraphs, although the balanced allocations with respect to a given baseload might not be uniquely defined, the total loads at the vertices resulting from any two balanced allocations necessarily have to be the same. See Figure 6.3 for an example. The case for infinite hypergraphs is more complicated though; in this case, the loads at the vertices also may not be unique, see Figure 6.2. See [Haj96] for more discussion on this. However, we can state a weak uniqueness result in this case. The proof is given in Appendix E.1.

Proposition 6.2. *Given the hypergraph $H = \langle V, E\rangle$ with the baseload function $b : V(H) \to \mathbb{R}$, suppose θ and θ' are two balanced allocations on H with respect to the baseload b. If $\sum_{i\in V(H)} |\partial_b\theta(i) - \partial_b\theta'(i)| < \infty$ then $\partial_b\theta(i) = \partial_b\theta'(i)$ for all $i \in V(H)$.*

6.2.3 $\mathcal{H}_*$ and $\mathcal{H}_{**}$

It is easy to check that the isomorphism between vertex rooted hypergraphs defined in Definition 6.6 is an equivalence relation. For a vertex rooted hypergraph (H, i), let $[H, i]$ denote the equivalence class corresponding to (H, i). Also, the isomorphism between edge-vertex rooted hypergraphs defined in Definition 6.6 is an equivalence relation. Let $[H, e, i]$ be the equivalence class corresponding to (H, e, i).

Definition 6.16. *Let $\mathcal{H}_*$ be the set of all equivalence classes of connected vertex rooted hypergraphs and $\mathcal{H}_{**}$ the set of all equivalence classes of connected edge-vertex rooted hypergraphs. Hence, each element of $\mathcal{H}_*$ is of the form $[H, i]$, where $[H, i]$ denotes the equivalence class of (H, i), where $i \in V(H)$ for some connected hypergraph $H = \langle V, E\rangle$. Similarly, each element of $\mathcal{H}_{**}$ is of the form $[H, e, i]$, where $e \in E(H)$ and $i \in V(H)$ such that $i \in e$ for some connected hypergraph $H = \langle V, E\rangle$.*

Definition 6.17. *For two vertex rooted hypergraphs (H, i) and (H', i'), define*

$$d_*((H,i),(H',i')) := \frac{1}{1+m^*} \ ,$$

where $m^* := \sup\{m \geq 1 : (H,i) \equiv_m (H',i')\}$, and $m^* := 0$ if there is no $m \geq 1$ satisfying this. For two equivalence classes $[H,i] \in \mathcal{H}_*$ and $[H',i'] \in \mathcal{H}_*$, define $d_{\mathcal{H}_*}([H,i],[H',i'])$ to be $d_*((H,i),(H',i'))$ where (H,i) and (H',i') are arbitrary members of $[H,i]$ and $[H',i']$, respectively. For two edge-vertex rooted hypergraphs (H,e,i) and (H',e',i'), define

$$d_{**}((H,e,i),(H',e',i')) := \frac{1}{1+m^*},$$

where $m^* := \sup\{m \geq 1 : (H,e,i) \equiv_m (H',e',i')\}$, and $m^* := 0$ if there is no $m \geq 1$ satisfying this. For two equivalence classes $[H,e,i] \in \mathcal{H}_{**}$ and $[H',e',i'] \in \mathcal{H}_{**}$, define $d_{\mathcal{H}_{**}}([H,e,i],[H',e',i'])$ to be $d_{**}((H,e,i),(H',e',i'))$ where (H,e,i) and (H',e',i') are arbitrary members of $[H,e,i]$ and $[H',e',i']$, respectively.

Since all members of $[H,i]$ are isomorphic, it is not difficult to see that $d_{\mathcal{H}_*}$ is well–defined. Note that $(H,i) \equiv_m (H',i')$ and $(H',i') \equiv_{m'} (H'',i'')$ implies $(H,i) \equiv_{\min\{m,m'\}} (H'',i'')$. Hence $d_{\mathcal{H}_*}$ satisfies the triangle inequality. Moreover, $d_{\mathcal{H}_*}([H,i],[H',i']) = 0$ iff $(H,i) \equiv_m (H',i')$ for all m, i.e. $[H,i] = [H',i']$. Hence $d_{\mathcal{H}_*}$ defines a metric on $\mathcal{H}_*$. We will show in Appendix E.2 that $\mathcal{H}_*$ with the metric $d_{\mathcal{H}_*}$ is a Polish space (see Corollary E.1). Similarly, $d_{\mathcal{H}_{**}}$ is well defined and gives a metric on $\mathcal{H}_{**}$. In Appendix E.2 we will also show that $\mathcal{H}_{**}$ with $d_{\mathcal{H}_{**}}$ is a Polish space.

Remark 6.3. *One can think of a function f on $\mathcal{H}_*$ as a function on vertex rooted hypergraphs which is loyal to the isomorphism relation, i.e. $f((H,i)) = f((H',i'))$ whenever $(H,i) \equiv (H',i')$. This allows us to abuse notation and write $f(H,i)$ instead of $f([H,i])$. We will follow a similar convention for functions on $\mathcal{H}_{**}$.*

Definition 6.18. *By abuse of notation, let $\mathcal{T}_*$ and $\mathcal{T}_{**}$ denote the set of equivalence classes of connected vertex rooted hypertrees and connected edge–vertex rooted hypertrees, respectively. It can be checked that $\mathcal{T}_*$ (respectively $\mathcal{T}_{**}$) is a closed subset of $\mathcal{H}_*$ (respectively $\mathcal{H}_{**}$).*

The set $\mathcal{T}_*$ of the above definition should not be confused with the set of rooted trees from Chapter 2. Throughout this chapter, and throughout Appendix E, we use $\mathcal{T}_*$ to denote the set of equivalence classes of connected vertex rooted hypertrees.

6.2.4 Operators for total load and average load: ∂ and ∇

Definition 6.19. *For a function $f : \mathcal{H}_{**} \to \mathbb{R}$, define $\partial f : \mathcal{H}_* \to \mathbb{R}$ as follows. For an equivalence class $[H,i] \in \mathcal{H}_*$, pick an arbitrary $(H',i') \in [H,i]$ and define*

$$\partial f([H,i]) = \sum_{e' \in E(H'), e' \ni i'} f([H',e',i'])$$

Remark 6.4. *Note that in the above definition, $[H', e', i']$ denotes the equivalence class of edge–vertex rooted hypergraph (H', e', i'). Also, since all the representatives of $[H, i]$ are isomorphic, it is easy to check that the above expression is not dependent on the specific choice of (H', i'). More precisely, if (H_1, i_1) and (H_2, i_2) are both members of $[H, i]$, then if $\phi : V(H_1) \to V(H_2)$ is the function establishing the isomorphism, ϕ gives a one to one mapping between the set $\{e \in E(H_1), e \ni i_1\}$ and $\{e \in E(H_2), e \ni i_2\}$ and also $[H_1, e, i_1] \equiv [H_2, \phi(e), i_2]$. Hence*

$$\sum_{e \in E(H_1), e \ni i_1} f([H_1, e, i_1]) = \sum_{e \in E(H_2), e \ni i_2} f([H_2, e, i_2]),$$

which shows that ∂f is well–defined.

Remark 6.5. *By the above discussion, we may write $\partial f(H, i) = \sum_{e \ni i} f(H, e, i)$, where by (H, i) we mean any arbitrary member of $[H, i]$.*

Remark 6.6. *By abuse of notation, we can think of ∂f as a function on $\mathcal{H}_{**}$ by identifying $\partial f(H, e, i) := \partial f(H, i)$. This will be helpful when we have functions both on $\mathcal{H}_*$ and $\mathcal{H}_{**}$ and want to unify the domain.*

Definition 6.20. *For a function $f : \mathcal{H}_{**} \to \mathbb{R}$, define the function $\nabla f : \mathcal{H}_{**} \to \mathbb{R}$ as follows. Given $[H, e, i] \in \mathcal{H}_{**}$, take an arbitrary representative $(H', e', i') \in [H, e, i]$ and define*

$$\nabla f([H, e, i]) := \frac{1}{|e'|} \sum_{j' \in e'} f([H', e', j']).$$

Remark 6.7. *As in our discussion in Remark 6.4, it can be easily checked that the above expression does not depend on the specific choice (H', e', i'). We can therefore abuse notation and write $\nabla f(H, e, i) = \frac{1}{|e|} \sum_{j \in e} f(H, e, j)$.*

Definition 6.21. *For a distribution $\mu \in \mathcal{P}(\mathcal{H}_*)$, define*

$$\deg(\mu) := \int_{\mathcal{H}_*} \deg_H(i) d\mu.$$

6.2.5 From $\mu \in \mathcal{P}(\mathcal{H}_*)$ to its directed version $\vec{\mu} \in \mathcal{M}(\mathcal{H}_{**})$

Definition 6.22. *Given $\mu \in \mathcal{P}(\mathcal{H}_*)$ with $\deg(\mu) < \infty$, define the measure $\vec{\mu} \in \mathcal{M}(\mathcal{H}_{**})$ as the one with the property that for any Borel function $f : \mathcal{H}_{**} \to [0, \infty)$, we have*

$$\int_{\mathcal{H}_{**}} f d\vec{\mu} = \int_{\mathcal{H}_*} \partial f d\mu.$$

Note that $\deg(\mu) = \int_{\mathcal{H}_{**}} 1 d\vec{\mu} = \vec{\mu}(\mathcal{H}_{**})$ is the total mass of $\vec{\mu}$. Hence the assumption $\deg(\mu) < \infty$ guarantees $\vec{\mu}(\mathcal{H}_{**}) < \infty$ and so $\vec{\mu} \in \mathcal{M}(\mathcal{H}_{**})$.

This following useful lemma is proved in Appendix E.3.

Lemma 6.1. *(i) Assume $A \subseteq \mathcal{H}_*$ happens μ–almost surely Then $\tilde{A} \subseteq \mathcal{H}_{**}$ defined as*

$$\tilde{A} := \{[H, e, i] \in \mathcal{H}_{**} : [H, i] \in A\},$$

happens $\vec{\mu}$–almost everywhere, i.e. $\vec{\mu}(\tilde{A}) = \vec{\mu}(\mathcal{H}_{**})$.

(ii) Assume $B \subseteq \mathcal{H}_{**}$ happens $\vec{\mu}$–almost everywhere. Then, $\tilde{B} := \{[H, i] \in \mathcal{H}_* : [H, e, i] \in B \ \forall\, e \ni i\}$ happens μ–almost surely.

The following fact relating the convergence of a sequence of functions on $\mathcal{H}_*$ and that of their counterparts on $\mathcal{H}_{**}$ will be useful later. This is proved in Appendix E.3.

Lemma 6.2. *Let $f_k : \mathcal{H}_* \to \mathbb{R}, k \geq 0$, be measurable functions such that we have $\lim_{k\to\infty} f_k = f_0$, μ–almost surely. Then, if we define their $\mathcal{H}_{**}$ counterparts $\tilde{f}_k : \mathcal{H}_{**} \to \mathbb{R}$ via*

$$\tilde{f}_k(H, e, i) := f_k(H, i),$$

then we have $\tilde{f}_k \to \tilde{f}_0$, $\vec{\mu}$–almost everywhere.

Another useful lemma is the following one, which relates the convergence of a sequence of functions on $\mathcal{H}_{**}$ to that of their ∂ on $\mathcal{H}_*$. This lemma is also proved in Appendix E.3.

Lemma 6.3. *Given $\mu \in \mathcal{P}(\mathcal{H}_*)$ and a sequence of functions $f_k : \mathcal{H}_{**} \to \mathbb{R}$, $k \geq 0$, assume that we have $f_k \to f_0$ $\vec{\mu}$–almost everywhere. Then we have $\partial f_k \to \partial f_0$ μ–almost surely.*

6.2.6 Local Weak Convergence

For a finite hypergraph H, define $u_H \in \mathcal{P}(\mathcal{H}_*)$ by choosing a vertex uniformly at random in H as the root. More precisely, if i is a vertex in H and $H(i)$ denotes the connected component of i, define

$$u_H := \frac{1}{|V(H)|} \sum_{i \in V(H)} \delta_{[H(i), i]}.$$

The reason why we take the connected component of H is that $\mathcal{H}_*$ is the space of equivalence classes of connected vertex rooted hypergraphs.

If, for a sequence of finite hypergraphs $\{H_n\}$, u_{H_n} converges weakly to some measure $\mu \in \mathcal{P}(\mathcal{H}_*)$, we say that μ is the "local weak limit" of the sequence H_n. The following lemma is useful in checking when local weak convergence occurs. See Appendix E.4 for a proof.

Lemma 6.4. *Given a sequence $\{\mu_n\}_{n\geq 1}$ in $\mathcal{P}(\mathcal{H}_*)$, and $\mu \in \mathcal{P}(\mathcal{H}_*)$ such that $supp(\mu) \subseteq \mathcal{T}_*$, $\mu_n \Rightarrow \mu$ iff the following condition is satisfied: for all $d \geq 1$ and for all rooted hypertrees (T, i) with depth at most d, if*

$$A_{(T,i)} := \{[H, j] \in \mathcal{H}_* : (H, j)_d \equiv (T, i)\},$$

then $\mu_n(A_{(T,i)}) \to \mu(A_{(T,i)})$.

The way to understand $u_{H_n} \Rightarrow \mu$ is that the local structure around a uniformly chosen vertex in H_n, where local means up to a fixed depth, looks more and more similar to that corresponding to μ, hence the term "local" weak convergence. In particular, Lemma 6.4 says that if we have a sequence of finite hypergraphs H_n, then $u_{H_n} \Rightarrow \mu$, where $\mu \in \mathcal{P}(\mathcal{T}_*)$, if and only if for each $d \geq 1$ and all rooted hypertrees (T, i), if we choose a vertex v in H_n uniformly at random, the probability that the local structure of H_n rooted at v, i.e. $(H_n, v)_d$, is isomorphic to (T, i) of depth at most d converges to the probability that the rooted tree with law μ up to depth d is isomorphic to T. See [BS01], [AS04], [AL07], [Bor14] for a review of the notion of local weak convergence in the graph regime.

6.2.7 Balanced allocations with respect to a distribution on $\mathcal{H}_*$

Definition 6.23. *A function $\Theta : \mathcal{H}_{**} \to [0,1]$ is called a Borel allocation, or just an allocation, if Θ is a Borel function and also*

$$\nabla\Theta(H, e, i) = \frac{1}{|e|}.$$

From Definition 6.20 and using our simplified notational conventions, we can equivalently say that Θ is an allocation precisely when it is a Borel function and $\sum_{i \in e} \Theta(H, e, i) = 1$ for the edge–vertex rooted hypergraph (H, e, i). It may seem strange that the condition is required only at the root edge. This will become more clear when we discuss unimodular measures in the next subsection, see especially Proposition 6.3.

Definition 6.24. *Assume $\mu \in \mathcal{P}(\mathcal{H}_*)$. A Borel allocation $\Theta : \mathcal{H}_{**} \to [0,1]$ is called balanced with respect to μ if for $\vec{\mu}$ almost all $[H, e, i] \in \mathcal{H}_{**}$ and any $(H', e', i') \in [H, e, i]$ we have:*

$$\forall j_1, j_2 \in e', \quad \partial\Theta([H', e', j_1]) > \partial\Theta([H', e', j_2]) \qquad \Rightarrow \qquad \Theta([H', e', j_1]) = 0.$$

Remark 6.8. *As in our discussion in Remarks 6.4 and 6.7, the above predicate does not depend on the specific choice of $(H', e', i') \in [H, e, i]$. Hence, by abuse of notation, we may write the above predicate simply as "for $\vec{\mu}$ almost all (H, e, i) and $j_1, j_2 \in e$, $\partial\Theta(H, j_1) > \partial\Theta(H, j_2)$ implies $\Theta(H, e, j_1) = 0$".*

Similar to our notion of ϵ–balanced allocation for a specific hypergraph, we can define ϵ–balanced allocations with respect to a measure $\mu \in \mathcal{P}(\mathcal{H}_*)$. As in the discussion in Remark 6.8, we use a simplified language in describing the definition.

Definition 6.25. *Assume $\mu \in \mathcal{P}(\mathcal{H}_*)$. A Borel allocation $\Theta_\epsilon : \mathcal{H}_{**} \to [0,1]$ is called ϵ–balanced with respect to μ if for $\vec{\mu}$ almost all $[H, e, i] \in \mathcal{H}_{**}$ we have*

$$\Theta_\epsilon(H, e, i) = \frac{\exp\left(-\partial\Theta_\epsilon(H, i)/\epsilon\right)}{\sum_{j \in e} \exp\left(-\partial\Theta_\epsilon(H, j)/\epsilon\right)}.$$

6.2.8 Unimodularity

Definition 6.26. *A probability measure $\mu \in \mathcal{P}(\mathcal{H}_*)$ is called unimodular if for every Borel function $f : \mathcal{H}_{**} \to [0, \infty)$ we have*

$$\int f d\vec{\mu} = \int \nabla f d\vec{\mu}.$$

We denote the set of unimodular measures on $\mathcal{H}_*$ by $\mathcal{P}_u(\mathcal{H}_*)$.

See [AL07] for a definition of unimodularity for graphs. It can be easily checked that our definition of unimodularity reduces to the definition in [AL07] when we restrict to graphs, i.e. when we restrict μ to have support on hypergraphs with all edges having size two.

It can be shown that for a finite hypergraph, u_H defined in Section 6.2.6 is unimodular. Moreover, if a sequence of finite hypergraphs has a local weak limit μ, then μ is unimodular[1]. See Appendix E.5 for a proof. Roughly speaking, unimodular measures are extensions of the kinds of measures on equivalence classes of vertex rooted hypergraphs that arise from choosing the root uniformly at random in finite hypergraphs.

The following property of unimodular measures is crucial in our analysis. It essentially says that "everything shows at the root" of a unimodular measure. See Appendix E.5 for its proof.

Proposition 6.3. *Assume $\tau : \mathcal{H}_{**} \to \mathbb{R}$ and $\mu \in \mathcal{P}_u(\mathcal{H}_*)$ is a unimodular probability measure such that $\tau = 1$ $\vec{\mu}$–almost everywhere. Then there exists some $A \subset \mathcal{H}_{**}$ such that $\vec{\mu}(A^c) = 0$ and*

$$\forall [H, e, i] \in A \quad \tau([H, e', i']) = 1, \quad \forall e' \in E(H), i' \in e', \forall (H, e, i) \in [H, e, i] \ .$$

Note that this statement is consistent with our intuition regarding unimodular measures: when some property holds at the root, since the root is chosen "uniformly" and so all vertices have the same "weight", that property should hold everywhere. See Lemma 2.3 in [AL07] for a version of the above statement for graphs.

6.2.9 Unimodular Galton–Watson hypertrees

In this section, we introduce an analogue of Galton Watson processes on graphs for hypertrees. In the graph regime, a Galton–Watson process is defined by generating the degree of the root at random from a given distribution and then, iteratively, the degree of each child is generated at random and so forth. In order to generalize the notion of a Galton–Watson process to hypertrees, since there might exist edges of different sizes, one needs to make sense of the "degree" of edges of each possible size at each node.

To this end, we introduce the notion of *type* as a generalization of the notion of degree. The type of each node is a vector of integers specifying how many edges of each size a node

[1] Whether the converse is true is an open question, even in the graph regime

is connected to. More precisely, since we want all the hypergraphs to be locally finite, we define the set of types, denoted by Λ, as

$$\Lambda := \{\gamma \in \mathbb{N}_0^{\{2,3,\dots\}} : \gamma(k) = 0, k > k_0 \text{ for some } k_0 \geq 2\}, \tag{6.3}$$

where $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For a type $\gamma \in \Lambda$, $\gamma(k)$ determines the number of edges of size k a node is connected to. For instance $(2, 1, 0, 1)$ means a node is connected to 2 edges of size 2, 1 edge of size 3 and 1 edge of size 5 (we haven't shown the rest of the sequence, which is zero).

For $\gamma \in \Lambda$ define

$$\|\gamma\|_1 := \sum_{k \geq 2} \gamma(k), \tag{6.4}$$

and

$$h(\gamma) := \max\{k \geq 2 : \gamma(k) > 0\}. \tag{6.5}$$

For $k \geq 2$, define $\mathbb{e}_k \in \Lambda$ to be the vector with value 1 at coordinate k and zero elsewhere.

Assume $P \in \mathcal{P}(\Lambda)$ is a probability distribution over the set of types, such that $\mathbb{E}[\Gamma(m)] < \infty$ for $m \geq 2$, where Γ is a random variable with law P. For $m \geq 2$ such that $\mathbb{E}[\Gamma(m)] > 0$, we define the size biased distributions $\hat{P}_m$ as

$$\hat{P}_m(\gamma) = \frac{(\gamma(m) + 1) P(\gamma + \mathbb{e}_m)}{\mathbb{E}[\Gamma(m)]}, \tag{6.6}$$

where it is easy to check that the normalizing term makes $\hat{P}_m$ a probability distribution. In case $\mathbb{E}[\Gamma(m)] = 0$, or equivalently $\Gamma(m) = 0$ with probability one, we define $\hat{P}_m$ to be an arbitrary distribution, e.g. $\hat{P}_m(\gamma) = 1$ when γ is the type with all coordinates being zero.

Let $\emptyset$ denote the root of the Galton–Watson hypertree and let $\mathbb{N}_{\text{vertex}}$ denote the set of vertices, so $\emptyset \in \mathbb{N}_{\text{vertex}}$. Let $\mathbb{N}_{\text{edge}}$ denote the set of hyperedges of the Galton–Watson hypertree. Each non-root element of $\mathbb{N}_{\text{vertex}}$ will be of the form $(s_1, e_1, i_1, \dots, s_k, e_k, i_k)$ where $s_j \geq 2$, $e_j \geq 1$, and $1 \leq i_j \leq s_j - 1$ for all $1 \leq j \leq k$. The semantics of (s_1, e_1, i_1) is that it is the vertex numbered i_1 of the e_1-th copy of a hyperedge of size s_1 attached to the root, and so on. Thus, for example $(3, 5, 2, 5, 8, 3)$ represents the vertex numbered 3 of the eighth hyperedge of size 5 that attaches to the vertex labeled 2 of the fifth hyperedge of size 3 that attaches to the root. The elements of $\mathbb{N}_{\text{edge}}$ are thus of the form $(s_1, e_1, i_1, \dots, s_k, e_k)$, where $s_j \geq 2$, $e_j \geq 1$, and $1 \leq i_j \leq s_j - 1$ for all $1 \leq j \leq k - 1$.

Given a sequence of types $\{\gamma_a\}_{a \in \mathbb{N}_{\text{vertex}}}$, we can construct a hypertree with vertex set and edge set $\mathbb{N}_{\text{vertex}}$ and $\mathbb{N}_{\text{edge}}$, respectively, where for $a \in \mathbb{N}_{\text{vertex}}$, γ_a determines the type of the node a in the subtree below node a. See Figure 6.4 for an example.

Definition 6.27. *Let $P \in \mathcal{P}(\Lambda)$ such that $\mathbb{E}[\Gamma(m)] < \infty$ for $m \geq 2$, where Γ has the distribution P. Construct a random rooted tree $(T, \emptyset)$ by generating $(\Gamma_a, a \in \mathbb{N}_{vertex})$ independently such that $\Gamma_\emptyset$ has law P and for any non–root node $a = (s_1, e_1, i_1, \dots, s_k, e_k, i_k)$, Γ_a has law $\hat{P}_{s_k}$. Then $\mathsf{UGWHT}(P)$ is the law of $[((T, \emptyset), \emptyset)]$ where $[((T, \emptyset), \emptyset)]$ denotes the equivalence class of $((T, \emptyset), \emptyset)$ in $\mathcal{H}_*$, with $(T, \emptyset)$ being the connected component of $\emptyset$ in T.*

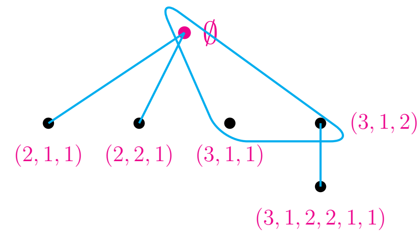

Figure 6.4: The tree rooted at $\emptyset$ generated by the sequence $\{\gamma_a\}_{a \in \mathbb{N}_{\text{vertex}}}$ where $\gamma_\emptyset = (2, 1)$. Furthermore, $\gamma_{(3,1,2)}$, which determines the type of $(3, 1, 2)$ in the subtree below $(3, 1, 2)$, is equal to (2) (which results in the single edge of size 2 below $(3, 1, 2)$), and γ_a is the zero vector for all other nodes a.

One important observation is that $\mathsf{UGWHT}(P)$, as in Definition 6.27, is unimodular. See Appendix E.6 for the proof.

Proposition 6.4. *Assume $P \in \mathcal{P}(\Lambda)$ is a distribution over types such that $\mathbb{E}[\Gamma(k)] < \infty$ for $k \geq 2$, then $\mathsf{UGWHT}(P)$ is unimodular.*

Note that if $(T, \emptyset)$ is generated as in Definition 6.27 and (s_1, e_1, i_1) is a child of the root present in T, the subtree rooted at i_1 has a similar distribution to $(T, \emptyset)$ except that the type of its root has law $\hat{P}_{s_1}$. It is useful for our subsequent discussion to define a notation for this distribution.

Definition 6.28. *Let $P \in \mathcal{P}(\Lambda)$ such that $\mathbb{E}[\Gamma(l)] < \infty$ for $l \geq 2$ and fix some $k \geq 2$. Construct a random rooted tree $(T, \emptyset)$ by generating $(\Gamma_a, a \in \mathbb{N}_{vertex})$ independently such that $\Gamma_\emptyset$ has law $\hat{P}_k$ and for any non–root node $a = (s_1, e_1, i_1, \ldots, s_r, e_r, i_r)$, Γ_a has law $\hat{P}_{s_r}$. Then $GWT_k(P)$ denotes the law of $[(T, \emptyset), \emptyset]$ where $[(T, \emptyset), \emptyset]$ denotes the equivalence class of $((T, \emptyset), \emptyset)$ in $\mathcal{H}_*$.*

6.2.10 Equivalence classes of marked hypergraphs: $\bar{\mathcal{H}}_*(\Xi)$ and $\bar{\mathcal{H}}_{**}(\Xi)$

Recall that $\mathcal{H}_*$ and $\mathcal{H}_{**}$ are Polish spaces of isomorphism classes of vertex rooted hypergraphs and edge–vertex rooted hypergraphs respectively. We can extend the procedure by which these spaces were created to hypergraphs with marks on their edges. Hypergraphs with marks on their edges would be called “hypernetworks”, following the terminology in Aldous and Lyons [AL07]. However, we prefer to call them *marked hypergraphs*.

Definition 6.29. *Assume H is a locally finite simple hypergraph on a countable vertex set and Ξ is a complete separable metric space. A Ξ–valued edge mark on H is a function*

$$\xi : \Psi(H) \to \Xi,$$

where we recall that $\Psi(H)$ denotes the set of edge-vertex pairs of H. A hypergraph carrying such a mark is called a marked hypergraph, and is denoted (H, ζ).

A vertex rooted marked hypergraph is a marked hypergraph (H, ξ) together with a distinguished vertex $i \in V(H)$, and is denoted $((H, \xi), i)$. An edge-vertex rooted marked hypergraph is a marked hypergraph (H, ξ) together with a distinguished edge $e \in E(H)$, and a distinguished vertex $i \in e$. It is denoted $((H, \xi), e, i)$.

Remark 6.9. *To simplify the notation, we employ the notation $\bar{H}$ to denote a marked hypergraph, where its mark function is denoted by $\xi_{\bar{H}}$, and its underlying unmarked hypergraph is denoted by H. Moreover, we may use $V(\bar{H})$, $E(\bar{H})$ and $\Psi(\bar{H})$ instead of $V(H)$, $E(H)$ and $\Psi(H)$.*

Definition 6.30. *We call two vertex rooted marked hypergraphs $(\bar{H}_1, i_1)$ and $(\bar{H}_2, i_2)$ isomorphic, and write $(\bar{H}_1, i_1) \equiv (\bar{H}_2, i_2)$ if there is a bijection $\phi : V(\bar{H}_1) \to V(\bar{H}_2)$, such that $\phi(i_1) = i_2$, and $e \in E(\bar{H}_1)$ iff $\phi(e) \in E(\bar{H}_2)$, where $\phi(e) := \{\phi(i) : i \in e\}$. Moreover, for $(e, i) \in \Psi(\bar{H}_1)$, we require that $\xi_{\bar{H}_1}(e, i) = \xi_{\bar{H}_2}(\phi(e), \phi(i))$.*

We say two edge-vertex rooted marked hypergraphs $(\bar{H}_1, e_1, i_1)$ and $(\bar{H}_2, e_2, i_2)$ are isomorphic, and write $(\bar{H}_1, e_1, i_1) \equiv (\bar{H}_2, e_2, i_2)$, if there is a bijection $\phi : V(\bar{H}_1) \to V(\bar{H}_2)$, such that $\phi(i_1) = i_2$, $\phi(e_1) = e_2$, and such that $e \in E(\bar{H}_1)$ iff $\phi(e) \in E(\bar{H}_2)$. Moreover, we must have $\xi_{\bar{H}_1}(e, i) = \xi_{\bar{H}_2}(\phi(e), \phi(i))$ for $(e, i) \in \Psi(\bar{H}_1)$.

Definition 6.31. *Let $\bar{\mathcal{H}}_*(\Xi)$ be the space of equivalence classes of connected vertex rooted marked hypergraphs with marks taking values in Ξ, with the equivalence class of $(\bar{H}, i)$ being denoted $[\bar{H}, i]$. Similarly, let $\bar{\mathcal{H}}_{**}(\Xi)$ be the space of equivalence classes of edge-vertex rooted marked hypergraphs with marks taking values in Ξ, with the equivalence class of $(\bar{H}, e, i)$ being denoted $[\bar{H}, e, i]$.*

We endow $\bar{\mathcal{H}}_*(\Xi)$ with the metric $\bar{d}_*$ where the distance between $[\bar{H}_1, i_1]$ and $[\bar{H}_2, i_2]$ is defined in the following way: take arbitrary representatives $(\bar{H}'_1, i'_1) \in [\bar{H}_1, i_1]$ and $(\bar{H}'_2, i'_2) \in [\bar{H}_2, i_2]$, then let m^* be the supremum over all m such that $(H'_1, i'_1) \equiv_m (H'_2, i'_2)$, and the Ξ–distance between the corresponding marks up to level m is at most $1/m$, i.e. if ϕ is the level m isomorphism, then

$$d_\Xi(\xi_{\bar{H}_1}(\tilde{e}, \tilde{i}), \xi_{\bar{H}_2}(\phi(\tilde{e}), \phi(\tilde{i}))) \le \frac{1}{m} , \qquad \forall \tilde{e} \in E_{H'_1}\left(V^{H'_1}_{i'_1, m}\right),$$

where d_Ξ denotes the metric on Ξ. If there is no m satisfying the above conditions, we set m^* to be 0. Then, $\bar{d}_*([\bar{H}_1, i_1], [\bar{H}_2, i_2])$ is defined to be $1/(1 + m^*)$. Since all the members in $[\bar{H}, i]$ are isomorphic as vertex rooted marked hypergraphs, $\bar{d}_*$ can be easily checked to be well defined. One can also check that it is a metric; in particular it satisfies the triangle inequality.

Similarly, we endow $\bar{\mathcal{H}}_{**}(\Xi)$ with the metric $\bar{d}_{**}$ where the distance between $[\bar{H}_1, e_1, i_1]$ and $[\bar{H}_2, e_2, i_2]$ is defined in the following way: take arbitrary representatives $(\bar{H}'_1, e'_1, i'_1) \in [\bar{H}_1, e_1, i_1]$ and $(\bar{H}'_2, e'_2, i'_2) \in [\bar{H}_2, e_2, i_2]$, then let m^{**} be the supremum over all m such that

$(H'_1, e'_1, i'_1) \equiv_m (H'_2, i'_2, e'_2)$ and the Ξ–distance between the corresponding marks up to level m is at most $1/m$. If there is no m satisfying these conditions, we set m^* to be 0. Finally, define $\bar{d}_{**}([\bar{H}_1, e_1, i_1], [\bar{H}_2, e_2, i_2])$ to be $1/(1 + m^{**})$.

In Appendix E.2 we prove that $\bar{\mathcal{H}}_*(\Xi)$ and $\bar{\mathcal{H}}_{**}(\Xi)$ with their respective metrics are Polish spaces, see Proposition E.1.

Similar to what we did for $\mathcal{H}_*$ and $\mathcal{H}_{**}$, we can define ∂ and ∇ operators as follows, where we use a simplified notation, whose validity can be justified as in Remarks 6.5 and 6.7:

$$\partial f(\bar{H}, i) := \sum_{e \in E(\bar{H}), e \ni i} f(\bar{H}, e, i),$$

and

$$\nabla f(\bar{H}, e, i) := \frac{1}{|e|} \sum_{j \in e} f(\bar{H}, e, j).$$

Remark 6.10. *The preceding notation, strictly speaking, applies only to real valued functions on $\mathcal{H}_{**}$. However, if we consider the function $f : \bar{\mathcal{H}}_{**}(\Xi) \to \Xi$ defined as $f(\bar{H}, e, i) = \xi_{\bar{H}}(e, i)$, then, when Ξ has an additive structure, we may define $\partial f : \bar{\mathcal{H}}_*(\Xi) \to \Xi$ as*

$$\partial f(\bar{H}, i) := \sum_{e \in E(H), e \ni i} f(\bar{H}, e, i) = \sum_{e \in E(\bar{H}), e \ni i} \xi_{\bar{H}}(e, i).$$

By abuse of notation, we may use the notation $\partial \xi_{\bar{H}}(i)$ instead of $\partial f(\bar{H}, i)$ with f defined above, and similarly for $\nabla \xi_{\bar{H}}(i)$. We will use such notation in this document because the marks we are interested in will be real valued.

For a probability measure $\mu \in \mathcal{P}(\bar{\mathcal{H}}_*(\Xi))$, we define $\vec{\mu} \in \mathcal{M}(\bar{\mathcal{H}}_{**}(\Xi))$ in a manner similar to what was done in Section 6.2.5. Namely, $\vec{\mu} \in \mathcal{M}(\bar{\mathcal{H}}_{**}(\Xi))$ is defined by requiring that for every nonnegative Borel function $f : \bar{\mathcal{H}}_{**}(\Xi) \to [0, \infty)$ we have

$$\int f d\vec{\mu} = \int \partial f d\mu.$$

A probability measure $\mu \in \mathcal{P}(\bar{\mathcal{H}}_*(\Xi))$ is called unimodular if for every nonnegative Borel function $f : \bar{\mathcal{H}}_{**}(\Xi) \to [0, \infty)$ we have

$$\int f d\vec{\mu} = \int \nabla f d\vec{\mu}.$$

By removing marks, we get a natural projection $\mathsf{Proj}_{\bar{\mathcal{H}}_*(\Xi) \to \mathcal{H}_*} : \bar{\mathcal{H}}_*(\Xi) \to \mathcal{H}_*$ defined as

$$\mathsf{Proj}_{\bar{\mathcal{H}}_*(\Xi) \to \mathcal{H}_*}([\bar{H}, i]) = [H, i] \ . \tag{6.7}$$

This can easily be checked to be continuous.

As was done in Section 6.2.8, choosing a vertex uniformly at random from a finite marked hypergraph results in a unimodular measure. More precisely, if H is a finite hypergraph together with a mark ξ taking values in Ξ,

$$u_{\bar{H}} := \frac{1}{|V(\bar{H})|} \sum_{i \in V(\bar{H})} \delta_{[\bar{H}(i),i]},$$

is unimodular. Here, $\bar{H}(i)$ denotes the connected component of i in $\bar{H}$. Moreover, the local weak limit of finite marked hypergraphs, i.e. the weak limit of the measures u_{H_n}, is unimodular, if it exists. See Appendix E.5 for a proof.

6.3 Main Results

We now summarize the main results of the chapter. We first prove some properties of balanced Borel allocations, as defined in Definition 6.24.

Theorem 6.1. *Let $\mu \in \mathcal{P}(\mathcal{H}_*)$ be a unimodular probability measure such that $\deg(\mu) < \infty$. The following are true.*

1. *There exists a Borel allocation $\Theta : \mathcal{H}_{**} \to [0,1]$ which is balanced with respect to μ.*
2. *Let Θ be a balanced Borel allocation with respect to μ. Then we have the following variational characterization of the mean excess load under Θ above the load level t. For any $t \in \mathbb{R}$:*
$$\int (\partial\Theta - t)^+ d\mu = \max_{\substack{f:\mathcal{H}_* \to [0,1] \\ \text{Borel}}} \int \tilde{f}_{min} d\vec{\mu} - t \int f d\mu,$$
where $\tilde{f}_{min}$ is defined as
$$\tilde{f}_{min}(H,e,i) := \frac{1}{|e|} \min_{j \in e} f(H,j).$$
3. *The following are equivalent for a Borel allocation $\Theta : \mathcal{H}_{**} \to [0,1]$:*
   - *a) Θ is balanced with respect to μ.*
   - *b) Θ minimizes $\int f \circ \partial\Theta d\mu$ among all Borel allocations, for some strictly convex function $f : [0,\infty) \to [0,\infty)$.*
   - *c) Θ minimizes $\int f \circ \partial\Theta d\mu$ among all Borel allocations for every convex function $f : [0,\infty) \to [0,\infty)$.*
4. *Assume Θ_0 is a balanced allocation with respect to μ and Θ is any other allocation on $\mathcal{H}_{**}$. Then Θ is balanced if and only if $\partial\Theta_0 = \partial\Theta$, μ–almost surely.*

5. *Let* $\{H_n\}_{n\geq 1}$ *be a sequence of finite hypergraphs with local weak limit* μ. *Let* $\mathcal{L}_{H_n}$ *denote the distribution of the total load at a vertex in* H_n *chosen uniformly at random. Namely,* $\mathcal{L}_{H_n} = \frac{1}{|V(H_n)|}\sum_{i\in V(H_n)} \delta_{\partial\theta_n(i)}$, *where* $(\partial\theta_n(i), i \in V(H_n))$ *denotes the load vector corresponding to any balanced allocation* θ_n, *which we recall exists and is unique, due to Proposition 6.1. Let* $\mathcal{L}$ *denote the law of the total load at the root of the balanced allocation on* μ, *i.e. the pushforward of* μ *under the mapping* $\partial\Theta$, *which we have just shown is well defined and unique, due to parts 1 and 4 of this theorem. Then* $\mathcal{L}_{H_n}$ *converges weakly to* $\mathcal{L}$.

The proof of the above theorem is given in Section 6.6. Before that, we first investigate, in Section 6.4, the properties of ϵ–balanced allocations for a specific hypergraph, as introduced in Definition 6.13. We then investigate, in Section 6.5, the properties of ϵ–balanced allocations on $\mathcal{H}_{**}$ for a given $\mu \in \mathcal{P}(\mathcal{H}_*)$, as defined in Definition 6.25. Then, by sending ϵ to zero, we prove part 1 of the above theorem in Section 6.6.1. The proofs of other parts of the theorem are given in Sections 6.6.2 through 6.6.5 respectively.

Note that, as a result of part 4 of the theorem, for a given $t \in \mathbb{R}$, the value of the integral $\int(\partial\Theta - t)^+ d\mu$ for a balanced allocation Θ does not depend on the particular choice of Θ. It only depends on μ and t, so it can be written as

$$\Phi_\mu(t) := \int (\partial\Theta - t)^+ d\mu. \tag{6.8}$$

The function Φ_μ is called the mean–excess function. Knowledge of Φ_μ is equivalent to determining the distribution $\mathcal{L}$ of the load at the root associated to a balanced allocation Θ. We now describe Φ_μ for the class of unimodular Galton–Watson process defined in Section 6.2.9.

Recall the notation Λ defined in (6.3). Assume $P \in \mathcal{P}(\Lambda)$ and $t \in \mathbb{R}$ are fixed. For a sequence of Borel probability measures $(Q_l, l \geq 2)$ on real numbers, let $F^{(k)}_{P,t}(\{Q_l\}_{l\geq 2})$ be the distribution of the random variable

$$t - \sum_{k'\geq 2}\sum_{i=1}^{\Gamma(k')} \left[1 - X^+_{k',i,1} - \cdots - X^+_{k',i,k'-1}\right]_0^1, \tag{6.9}$$

where Γ has law $\hat{P}_k$, and $X_{k',i,j}$ are random variables which are mutually independent and independent of Γ, with $X_{k',i,j}$ having law $Q_{k'}$. Note that $\hat{P}_k$ is the size biased version of P defined in (6.6). Also note that the first sum on the right hand side of (6.9) is a finite sum, because Γ has finite support, pointwise.

Let $\mathcal{Q}$ be the set of sequences $\{Q_l\}_{l\geq 2}$ such that, for all $k \geq 2$, we have:

$$Q_k = F^{(k)}_{P,t}(\{Q_l\}_{l\geq 2}). \tag{6.10}$$

Now, we are ready to provide a characterization of the mean excess function. We give the proof of the following result in Section 6.8.

Theorem 6.2. *Let P be a distribution on Λ such that $\mathbb{E}\left[\|\Gamma\|_1\right] < \infty$ where Γ has law P. Then, with $\mu := \mathsf{UGWHT}(P)$, for any $t \in \mathbb{R}$, we have*

$$\Phi_\mu(t) = \max_{\{Q_k\}_{k\geq 2}\in\mathcal{Q}} \left(\sum_{k=2}^{\infty} \frac{\mathbb{E}\left[\Gamma(k)\right]}{k} \mathbb{P}\left(\sum_{i=1}^{k} X_{k,i}^+ < 1 \right) \right) - t\mathbb{P}\left(\sum_{k=2}^{h(\Gamma)} \sum_{i=1}^{\Gamma(k)} Y_{k,i} > t \right), \tag{6.11}$$

where, in the first expression, Γ is a random variable on Λ with law P and $\{X_{k,i}\}_{k,i}$ are i.i.d. such that $X_{k,i}$ has law Q_k. Also, in the second expression, Γ has law P and $\{Y_{k,i}\}_{k,i}$ are independent from each other and from Γ, with $Y_{k,i}$ having the law of the random variable $[1 - (Z_1^+ + \cdots + Z_{k-1}^+)]_0^1$, where Z_j are i.i.d. with law Q_k.

For a finite hypergraph H, we define $\varrho(H)$ to be the maximum load corresponding to a balanced allocation on H, i.e. if θ is a balanced allocation on H,

$$\varrho(H) := \max_{v\in V(H)} \partial\theta(v), \tag{6.12}$$

which is well defined due to Proposition 6.1. From [Haj90, Corollary 7] we know that there is a duality between this parameter and the subgraph of maximum edge density, i.e.

$$\varrho(H) = \max_{S\subseteq V(H), S\neq\emptyset} \frac{|E_H(S)|}{|S|}, \tag{6.13}$$

where $E_H(S)$ denotes the set of edges of H with all endpoints in S.

For a unimodular probability distribution μ on $\mathcal{H}_*$ with finite $\deg(\mu)$, we define

$$\varrho(\mu) := \sup\{t \in \mathbb{R} : \Phi_\mu(t) > 0\}, \tag{6.14}$$

where $\Phi_\mu(.)$ is the mean excess function defined above. In other words, if Θ is the balanced allocation corresponding to μ introduced in Theorem 6.1 and $\mathcal{L}_\mu$ is the law of $\partial\Theta$ under μ, then

$$\varrho(\mu) = \sup\{t \in \mathbb{R} : \mathcal{L}_\mu([t, \infty)) > 0\}.$$

One question is whether local weak convergence implies convergence of maximum load, i.e., if H_n is a sequence of graphs with local weak limit μ, does $\varrho(H_n)$ converge to $\varrho(\mu)$? Similar to the graph case, this is not true in general, since we can always add an arbitrary but bounded clique to boost $\varrho(H_n)$ without changing the local weak limit. We prove convergence, under some conditions, for the special case where the limit μ is the UGWT model defined in Section 6.2.9 and, for each n, H_n is a random hypergraph obtained from a generalized hypergraph configuration model defined in Section 6.9.1.

Theorem 6.3. *Let P be a probability distribution on Λ such that, if Γ is a random variable with law P, $P(\Gamma(k) > 0) > 0$ for finitely many k and $\mathbb{E}\left[\Gamma(k)\right] < \infty$ for all $k \geq 2$. Moreover, let $\mu := \mathsf{UGWHT}(P)$. Then, if $\{H_n\}_{n=1}^\infty$ is a sequence of random hypergraphs obtained from a configuration model, under some conditions stated in Proposition 6.18, $\varrho(H_n)$ converges in probability to $\varrho(\mu)$.*

This theorem is proved in Section 6.9.

6.4 ϵ–balancing with baseloads

In this section, we analyze the properties of ϵ–balanced allocations with respect to a baseload, which were introduced in Definition 6.15. Note that throughout this section, we are dealing with a given hypergraph, not a distribution on $\mathcal{H}_*$. By setting the baseload function b to zero, our results here reduce to those for ϵ–balanced allocations as introduced in Definition 6.13.

6.4.1 Existence

The existence of ϵ–balanced allocations with respect to a baseload b on hypergraphs is a consequence of the Schauder–Tychonoff fixed point theorem (see, for instance, [AMO09]). Here we give the details. Fix a hypergraph $H = \langle V, E \rangle$ and define the topological vector space W to be:

$$W := \{\theta : \Psi(H) \to \mathbb{R}\} = \mathbb{R}^{\Psi(H)},$$

with the product topology of $\mathbb{R}$. Here, we recall that $\Psi(H)$ is the set of all edge–vertex pairs of the hypergraph (6.1), and that this is a countable set. Define the following convex subset of functions with values in $[0, 1]$:

$$A := \{\theta : \Psi(H) \to [0, 1]\}.$$

Since we have employed the product topology, Tychonoff's theorem tells us that A is a compact set (see, for instance, [Mun00]). Define the mapping $T : A \to A$ via:

$$(T\theta)(e, i) := \frac{\exp\left(-\frac{\partial_b\theta(i)}{\epsilon}\right)}{\sum_{j\in e}\exp\left(-\frac{\partial_b\theta(j)}{\epsilon}\right)}.$$

We want to show that T has a fixed point. In order to do so, we need to show that T is continuous. Since we have employed the product topology, we need to show that, for all $(e, i) \in \Psi(H)$, the projected version $T_{e,i}$ defined as:

$$T_{e,i}(\theta) := (T\theta)(e, i),$$

which is a mapping from A to $[0, 1]$, is continuous. In order to show this, note that $T_{e,i}$ is the concatenation of a projection $\mathsf{Proj}_e : A \to \mathbb{R}^{|U_e|}$, where

$$U_e := \{(e', j) : e' \in E, e' \cap e \neq \emptyset, j \in e \cap e'\},$$

and an addition function from $\mathbb{R}^{|U_e|}$ to $\mathbb{R}^{|e|}$, which gives us the vector $[\partial\theta(j)]_{j\in e}$, and then a function $f : \mathbb{R}^{|e|} \to \mathbb{R}$ defined as:

$$f([x_j]_{j\in e}) := \frac{e^{-(x_i+b(i))/\epsilon}}{\sum e^{-(x_j+b(j))/\epsilon}}.$$

Since all these three functions are continuous (note that U_e is a finite set since we have assumed that the graph is locally finite and all edges have finite size), T is also continuous. Therefore, since W is Hausdorff and locally convex, A is compact, and T is continuous, the Schauder–Tychonoff fixed point theorem implies that T has a fixed point (see, for instance, [AMO09, Theorem 8.2]). Note that $\theta' := T(\theta)$ satisfies $\sum_{i\in e} \theta'(e,i) = 1$ for any $\theta \in W$. Therefore this fixed point is an allocation in the sense of Definition 6.10, and is also ϵ–balanced.

6.4.2 Monotony and Uniqueness

Intuitively, we expect that when we add more edges to a hypergraph and increase baseloads, the total load for an ϵ–balanced allocation would increase. We also expect that the effect of an increase in baseload at any vertex tends to dissipate as one moves away from the vertex, when comparing the respective balanced allocations. Lemmas 6.5 and 6.7 below quantify these phenomena. They are formulated in the language of vertex rooted hypergraph embedding from Definition 6.8.

Lemma 6.5 (depth 1 local contraction). *Assume the vertex rooted hypergraph (H, i) can be embedded up to depth 1 into the vertex rooted hypergraph (H', i'), i.e. $(H, i) \hookrightarrow_1 (H', i')$, with embedding $\phi : V^H_{i,1} \hookrightarrow V^{H'}_{i',1}$. Let θ_ϵ and θ'_ϵ be ϵ–balanced allocations on H and H' respectively, with respective baseload functions b and b', with $b(i) \le b'(i')$. If*

$$M := \max_{j:d_H(i,j)=1} \partial_b\theta_\epsilon(j) - \partial_{b'}\theta'_\epsilon(\phi(j)),$$

then we have

$$\partial_{b'}\theta'_\epsilon(i') \ge \partial_b\theta_\epsilon(i) - \frac{|D^H_{i,1}|}{|D^H_{i,1}| + 4\epsilon} M^+,$$

where $D^H_{i,1}$ is the set of nodes at distance one from node i as was defined in Definition 6.4.

Note that, in this lemma, θ_ϵ and θ'_ϵ are two arbitrary ϵ–balanced allocations on H and H' respectively, with respective baseload functions b and b'. We know from Section 6.4.1 that such allocations exist, but they might a priori not be unique. We will later prove uniqueness for the special case of bounded hypergraphs, which were introduced in Definition 6.2.

Before proving this result, we need the following tool, whose proof is given after the proof of Lemma 6.5.

Lemma 6.6. *Assume that for $\epsilon > 0$, the function $f_\epsilon : \mathbb{R}^k \to \mathbb{R}$ is defined in the following way:*

$$f_\epsilon(x_1, \ldots, x_k) = \frac{1}{1 + \sum_{i=1}^k e^{-\frac{x_i}{\epsilon}}}.$$

Then, for arbitrary real valued sequences $(x_1, \ldots, x_k)$ and $(x'_1, \ldots, x'_k)$, we have,

$$f_\epsilon(x_1, \ldots, x_k) - f_\epsilon(x'_1, \ldots, x'_k) \le \frac{1}{4\epsilon} \sum_{i=1}^k [x_i - x'_i]^+.$$

Proof of Lemma 6.5. Since $\phi(e) \in E(H')$ for $e \ni i$, and θ'_ϵ is a nonnegative function, we have

$$\sum_{e' \ni i'} \theta'_\epsilon(e', i') \geq \sum_{e \ni i} \theta'_\epsilon(\phi(e), i'). \tag{6.15}$$

On the other hand, we have

$$\begin{aligned}
\partial_b \theta_\epsilon(i) - \partial_{b'} \theta'_\epsilon(i') &\overset{(a)}{\leq} \sum_{e \ni i} \theta_\epsilon(e, i) - \sum_{e' \ni i'} \theta'_\epsilon(e', i') \\
&\overset{(b)}{\leq} \sum_{e \ni i} \theta_\epsilon(e, i) - \theta'_\epsilon(\phi(e), i') \\
&\overset{(c)}{=} \sum_{e \ni i} \left(\frac{1}{1 + \sum_{\substack{j \in e \\ j \neq i}} \exp\left(-\frac{\partial_b \theta_\epsilon(j) - \partial_b \theta_\epsilon(i)}{\epsilon}\right)} \right. \\
&\qquad \left. - \frac{1}{1 + \sum_{\substack{j \in e \\ j \neq i}} \exp\left(-\frac{\partial_{b'} \theta'_\epsilon(\phi(j)) - \partial_{b'} \theta'_\epsilon(i')}{\epsilon}\right)} \right) \\
&\overset{(d)}{\leq} \frac{1}{4\epsilon} \sum_{e \ni i} \sum_{\substack{j \in e \\ j \neq i}} \left[(\partial_b \theta_\epsilon(j) - \partial_b \theta_\epsilon(i)) - (\partial_{b'} \theta'_\epsilon(\phi(j)) - \partial_{b'} \theta'_\epsilon(i')) \right]^+,
\end{aligned} \tag{6.16}$$

where (a) results from $b(i) \leq b'(i')$, (b) uses (6.15), (c) is a substitution from Definition 6.14, and (d) uses Lemma 6.6.

Now, let

$$I := \{(e, j) : e \ni i, j \in e, j \neq i, \partial_b \theta_\epsilon(j) - \partial_{b'} \theta'_\epsilon(\phi(j)) \geq \partial_b \theta_\epsilon(i) - \partial_{b'} \theta'_\epsilon(i')\}.$$

Then the inequality in (6.16) together with the definition of M implies

$$\partial_b \theta_\epsilon(i) - \partial_{b'} \theta'_\epsilon(i') \leq \frac{1}{4\epsilon} |I| (M - (\partial_b \theta_\epsilon(i) - \partial_{b'} \theta'_\epsilon(i'))).$$

Rearranging the terms, we get

$$\partial_b \theta_\epsilon(i) - \partial_{b'} \theta'_\epsilon(i') \leq \frac{|I|}{|I| + 4\epsilon} M \leq \frac{|D^H_{i,1}|}{|D^H_{i,1}| + 4\epsilon} M^+,$$

which is exactly what we wanted to prove. □

Proof of Lemma 6.6. First we prove the statement for $k = 1$. In this case, the function $f_\epsilon(x) = \frac{1}{1 + e^{-\frac{x}{\epsilon}}}$ is $\frac{1}{4\epsilon}$–Lipschitz and increasing in x, hence the statement holds for $k = 1$.

Now assume $k > 1$ is arbitrary. We have

$$f_\epsilon(x_1, \ldots, x_k) - f_\epsilon(x'_1, \ldots, x'_k) = \sum_{i=1}^{k} \frac{e^{-\frac{x'_i}{\epsilon}} - e^{-\frac{x_i}{\epsilon}}}{\left(1 + \sum_{r=1}^{k} e^{-\frac{x_r}{\epsilon}}\right) \left(1 + \sum_{s=1}^{k} e^{-\frac{x'_s}{\epsilon}}\right)}. \tag{6.17}$$

Now, for each $1 \le i \le k$, if $x'_i \ge x_i$ then

$$\frac{e^{-\frac{x'_i}{\epsilon}} - e^{-\frac{x_i}{\epsilon}}}{\left(1+\sum_{r=1}^k e^{-\frac{x_r}{\epsilon}}\right)\left(1+\sum_{s=1}^k e^{-\frac{x'_s}{\epsilon}}\right)} \le 0 = \frac{1}{4\epsilon}[x_i - x'_i]^+.$$

On the other hand, if $x'_i < x_i$, then we have

$$\begin{aligned}\frac{e^{-\frac{x'_i}{\epsilon}} - e^{-\frac{x_i}{\epsilon}}}{\left(1+\sum_{r=1}^k e^{-\frac{x_r}{\epsilon}}\right)\left(1+\sum_{s=1}^k e^{-\frac{x'_s}{\epsilon}}\right)} &\le \frac{e^{-\frac{x'_i}{\epsilon}} - e^{-\frac{x_i}{\epsilon}}}{\left(1+e^{-\frac{x_i}{\epsilon}}\right)\left(1+e^{-\frac{x'_i}{\epsilon}}\right)} \\ &= \frac{1}{1+e^{-\frac{x_i}{\epsilon}}} - \frac{1}{1+e^{-\frac{x'_i}{\epsilon}}} \\ &\le \frac{1}{4\epsilon}[x_i - x'_i]^+,\end{aligned}$$

where the last step uses the statement for $k = 1$. Therefore, in either case, we have proved that for all $1 \le i \le k$ we have:

$$\frac{e^{-\frac{x'_i}{\epsilon}} - e^{-\frac{x_i}{\epsilon}}}{\left(1+\sum_{r=1}^k e^{-\frac{x_r}{\epsilon}}\right)\left(1+\sum_{s=1}^k e^{-\frac{x'_s}{\epsilon}}\right)} \le \frac{1}{4\epsilon}[x_i - x'_i]^+.$$

Substituting this into (6.17) we get the desired result. □

Now we generalize Lemma 6.5 to depth d local embeddings.

Lemma 6.7 (depth d local contraction)**.** *Assume $(H, i) \hookrightarrow_d (H', i')$ with embedding $\phi : V^H_{i,d} \hookrightarrow V^{H'}_{i',d}$. Also let θ_ϵ and θ'_ϵ be ϵ–balanced allocations on H and H' respectively, with respective baseload functions b and b', where $b(j) \le b'(\phi(j))$ for all j such that $d_H(i, j) \le d-1$. Then we have:*

$$\partial_{b'}\theta'_\epsilon(i') \ge \partial_b\theta_\epsilon(i) - \left(\frac{L\Delta}{4\epsilon + L\Delta}\right)^d M_d^+,$$

where

$$M_d := \max_{j \in D^H_{i,d}} \partial_b\theta_\epsilon(j) - \partial_{b'}\theta'_\epsilon(\phi(j)),$$

and

$$L := \max_{e \in E_H(V^H_{i,d})} |e| \ ,$$

$$\Delta := \max_{j \in V^H_{i,d-1}} \deg_H(j) \ .$$

Proof. Take some $j \in V(H)$ with $d_H(i,j) = k$. If $j' \in V(H)$ is such that $d_H(j,j') = 1$, the triangle inequality implies that

$$|d_H(i,j') - d_H(i,j)| \le d_H(j,j') = 1.$$

Thus

$$D^H_{j,1} \subset D^H_{i,k-1} \cup D^H_{i,k} \cup D^H_{i,k+1}. \tag{6.18}$$

Hence, if $d_H(i,j) \le d-1$, using the same embedding map ϕ we have $(H,j) \hookrightarrow_1 (H',\phi(j))$. Hence, if we define

$$M_k := \max_{j:d_H(i,j)=k} \partial_b\theta_\epsilon(j) - \partial_{b'}\theta'_\epsilon(\phi(j)) \qquad 0 \le k \le d,$$

and

$$\alpha := \frac{L\Delta}{4\epsilon + L\Delta},$$

then using Lemma 6.5 and (6.18) we have

$$M_k \le \alpha(M^+_{k-1} \vee M^+_k \vee M^+_{k+1}) \qquad 1 \le k \le d-1, \tag{6.19}$$

and

$$M_0 \le \alpha M^+_1. \tag{6.20}$$

We show by induction that $M^+_k \le \alpha M^+_{k+1}$ for $0 \le k \le d-1$. For $k=0$, this follows from (6.20) and the fact that $x \mapsto x^+$ is increasing. For $k \ge 1$ we have from (6.19) that

$$M^+_k \le \alpha(M^+_{k-1} \vee M^+_k \vee M^+_{k+1}) \le \alpha((\alpha M^+_k \vee M^+_k) \vee M^+_{k+1}) = \alpha(M^+_k \vee M^+_{k+1}).$$

We claim that $M^+_k \le \alpha M^+_{k+1}$. The above inequality means that either $M^+_k \le \alpha M^+_k$ or $M^+_k \le \alpha M^+_{k+1}$. The latter case is precisely what we have claimed, while in the former case, as $\alpha < 1$, we have $M_k = 0$ and the inequality $M^+_k \le \alpha M^+_{k+1}$ is automatic. This implies that $M_0 \le \alpha^d M^+_d$ and completes the proof. □

Using the above local results, we now show that if we add edges to a hypergraph and/or increase the baseload, the total load should increase.

Proposition 6.5. *Let H and H' be two hypergraphs defined on the same vertex set $V = V(H) = V(H')$, such that $E(H) \subset E(H')$, and H is bounded (but H' is not necessarily bounded). Suppose two bounded baseload functions $b, b' : V \to \mathbb{R}$ are given, such that $b(i) \le b'(i)$ for all $i \in V$. If θ_ϵ and θ'_ϵ are two ϵ–balanced allocations on H and H', respectively, for the respective baseload functions b and b', then we have $\partial_b\theta_\epsilon(i) \le \partial_{b'}\theta'_\epsilon(i)$ for all $i \in V$.*

Proof. Since H is bounded, there are constants Δ and L such that, for all $i \in V$, $\deg_H(i) \le \Delta$ and $|e| \le L$ for all $e \in E(H)$. Also, as b is bounded, for some $K > 0$ and all $i \in V$, we have $|b(i)| \le K$.

Now, fix some $i \in V$. Since $E(H) \subset E(H')$, we have $(H, i) \hookrightarrow_1 (H', i)$ with the identity map as the embedding function. With this, define

$$M := \sup_{j \in V} \partial_b \theta_\epsilon(j) - \partial_{b'} \theta'_\epsilon(j).$$

Note that, since $\partial_b \theta_\epsilon^b$ is bounded to $\Delta + K$, the above quantity is finite and well defined. Now, using Lemma 6.5, we have

$$\partial_b \theta_\epsilon^b(i) - \partial_{b'} \theta_\epsilon^{b'}(i) \leq \frac{\Delta L}{4\epsilon + \Delta L} M^+.$$

Taking the supremum over i on the left hand side, we get

$$M \leq \frac{\Delta L}{4\epsilon + \Delta L} M^+,$$

which, since $\frac{\Delta L}{4\epsilon + \Delta L} < 1$, implies $M \leq 0$ and completes the proof. □

If we have a fixed bounded hypergraph H with a baseload function, and $\theta_\epsilon, \theta'_\epsilon$ are two ϵ–balanced allocations on H, repeating the above proposition twice, we get $\partial \theta_\epsilon \leq \partial \theta'_\epsilon$ and $\partial \theta'_\epsilon \leq \partial \theta_\epsilon$, which implies uniqueness. To sum up, we have:

Corollary 6.1. *If a hypergraph H is bounded, there is a unique ϵ–balanced allocation with respect to any given baseload on it.*

6.4.3 ϵ–balanced allocations for unbounded hypergraphs with respect to a baseload: canonical allocations

We now produce ϵ–balanced allocations on a hypergraph H with respect to a given baseload even when the hypergraph is not necessarily bounded. Note that we do not make any claims about uniqueness.

For a given hypergraph H and $\Delta \in \mathbb{N}$, define H^Δ to be the hypergraph with vertex set $V(H)$ and edge set E^Δ, where

$$E^\Delta := \{e \in E(H) : |e| \leq \Delta, \deg_H(i) \leq \Delta \quad \forall i \in e\}. \tag{6.21}$$

Given the baseload function b, define θ_ϵ^Δ to be the unique ϵ–balanced allocation on H^Δ with respect to the baseload b, where the existence and uniqueness of θ_ϵ^Δ is a consequence of Corollary 6.1 above. Since E^Δ increases to $E(H)$ as Δ increases, Proposition 6.5 implies that $\partial_b \theta_\epsilon^\Delta$ is pointwise increasing in Δ. Since it is also pointwise bounded, it is convergent. For a node $i \in V(H)$, define

$$l_i := \lim_{\Delta \to \infty} \partial_b \theta_\epsilon^\Delta(i),$$

Now, for $e \in E(H)$ and $i \in e$, define

$$\theta_\epsilon(e,i) := \frac{\exp\left(-\frac{l_i}{\epsilon}\right)}{\sum_{j\in e}\exp\left(-\frac{l_j}{\epsilon}\right)}. \tag{6.22}$$

Now we prove that θ_ϵ is an ϵ–balanced allocation with respect to the baseload b. First, observe that, because of the normalizing term in the denominator, $\sum_{i\in e}\theta_\epsilon(e,i) = 1$ for all $e \in E(H)$. We now show that $l_i = \partial_b\theta_\epsilon(i)$. Note that

$$\begin{aligned}\sum_{e\ni i}\theta_\epsilon(e,i) &= \sum_{e\ni i}\frac{\exp\left(-\frac{l_i}{\epsilon}\right)}{\sum_{j\in e}\exp\left(-\frac{l_j}{\epsilon}\right)}\\ &= \lim_{\Delta\to\infty}\sum_{e\ni i}\frac{\exp\left(-\frac{\partial_b\theta^\Delta_\epsilon(i)}{\epsilon}\right)}{\sum_{j\in e}\exp\left(-\frac{\partial_b\theta^\Delta_\epsilon(j)}{\epsilon}\right)}.\end{aligned}$$

Observe that, for $\Delta > \max_{e\ni i}(|e| \vee \max_{j\in e}\deg_H(j))$, we have $e \in E^\Delta$ for all $e \ni i$. Hence, the term inside the summation is $\theta^\Delta_\epsilon(e,i)$, because θ^Δ_ϵ is the unique ϵ-balanced allocation on H^Δ with respect to the baseload b. Since we are taking $\Delta \to \infty$, we can assume it is big enough to get

$$\partial_b\theta_\epsilon(i) = b(i) + \lim_{\Delta\to\infty}\sum_{e\ni i}\theta^\Delta_\epsilon(e,i) = \lim_{\Delta\to\infty}\partial_b\theta^\Delta_\epsilon(i) = l_i.$$

Substituting this into (6.22), we conclude that θ_ϵ is ϵ–balanced, with respect to the baseload b.

Remark 6.11. *Note that the above procedure gives rise to an ϵ–balanced allocation with respect to any given baseload, but we do not know if it is the only possible ϵ–balanced allocation or not. In fact, we have proved uniqueness for bounded hypergraphs only. To emphasize this and avoid confusion, we call the ϵ–balanced allocation resulting from the procedure above the "canonical" allocation with respect to the given baseload. A special case of the procedure yields the canonical ϵ–balanced allocation when there is no baseload.*

Now, we generalize the monotonicity property of Proposition 6.5 to the case of not necessarily bounded hypergraphs, for these canonical allocations.

Proposition 6.6. *Given hypergraphs H and H' on the same vertex set V, with $E(H) \subseteq E(H')$, and baseload functions $b, b' : V \to \mathbb{R}$ such that $b(i) \le b'(i)$ for all $i \in V$, if θ_ϵ and θ'_ϵ are the canonical ϵ–balanced allocations on H and H', with respect to baseloads b and b', respectively, then $\partial_b\theta_\epsilon(i) \le \partial_{b'}\theta'_\epsilon(i)$ for all $i \in V$.*

Proof. Set $E := E(H)$ and $E' := E(H')$. Note that since $E \subseteq E'$ and the vertex sets are the same for H and H', $E^\Delta \subseteq E'$ for any Δ. Thus, if θ^Δ_ϵ is the unique ϵ–balanced allocation on $\langle V, E^\Delta\rangle$ with respect to the baseload b and θ'_ϵ is the canonical ϵ–balanced allocation on

$\langle V, E' \rangle$ with respect to the baseload b', Proposition 6.5 implies that $\partial_b \theta_\epsilon^\Delta(i) \leq \partial_{b'} \theta'_\epsilon(i)$ for all $i \in V$ (note that in Proposition 6.5 only the smaller hypergraph needs to be bounded, hence we only need to truncate E). By sending Δ to infinity, we get the desired result. □

6.4.4 Nonexpansivity

Proposition 6.7. *Let H be a given hypergraph and b, b' be two baseload functions. Let θ and θ' be the canonical ϵ–balanced allocations on H with respect to b and b', respectively. Then, we have*

$$\|\partial_b \theta_\epsilon - \partial_{b'} \theta'_\epsilon\|_{l^1(V(H))} \leq \|b - b'\|_{l^1(V(H))} .$$

Proof. To start with, assume that H is bounded. Later, we will relax this assumption.

Consider first the special case $b(i) \geq b'(i)$ for all $i \in V(H)$. Then, using the monotonicity property of Proposition 6.5, we have $\partial_b \theta_\epsilon(i) \geq \partial_{b'} \theta'_\epsilon(i)$ for all $i \in V(H)$. In particular,

$$\|\partial_b \theta_\epsilon - \partial_{b'} \theta'_\epsilon\|_{l^1(V(H))} = \sum_{i \in V(H)} \partial_b \theta_\epsilon(i) - \partial_{b'} \theta'_\epsilon(i).$$

Now, let $\{V_n\}$ be a nested sequence of finite subsets of $V(H)$ converging to $V(H)$, i.e. $V_n \uparrow V(H)$. Let $\theta_{n,\epsilon}$ and $\theta'_{n,\epsilon}$ be the ϵ–balanced allocations on $\langle V_n, E_H(V_n) \rangle$ with respect to the restrictions of b and b' to V_n, respectively. The monotonicity property of Proposition 6.5 and an argument similar to the one given in Section 6.4.3 above, implies that $\partial_b \theta_{n,\epsilon}$ and $\partial_{b'} \theta'_{n,\epsilon}$ converge to ϵ–balanced allocations on H and H' with respect to b and b', respectively. Since H is bounded, there is a unique ϵ–balanced allocations with respect to any baseload function. Therefore, we have $\partial_b \theta_{n,\epsilon} \uparrow \partial_b \theta_\epsilon$ and $\partial_{b'} \theta'_{n,\epsilon} \uparrow \partial_{b'} \theta'_\epsilon$ as $n \to \infty$. Using the conservation of mass, we have

$$\sum_{i \in V_n} \partial_b \theta_{n,\epsilon}(i) - \partial_{b'} \theta'_{n,\epsilon}(i) = \sum_{i \in V_n} b(i) - b'(i) \leq \sum_{i \in V(H)} b(i) - b'(i) = \|b - b'\|_{l^1(V(H))} ,$$

where we have used the assumption $b \geq b'$. In fact, for any finite subset of vertices $K \subset V(H)$, we have $K \subset V_n$ for n large enough and so using $\partial_b \theta_{n,\epsilon} \uparrow \partial_b \theta_\epsilon$ and $\partial_{b'} \theta'_{n,\epsilon} \uparrow \partial_{b'} \theta'_\epsilon$ and also the monotonicity property of Proposition 6.5, we have

$$\begin{aligned}
\sum_{i \in K} \partial_b \theta_\epsilon(i) - \partial_{b'} \theta'_\epsilon(i) &= \lim_{n \to \infty} \sum_{i \in K} \partial_b \theta_{n,\epsilon}(i) - \partial_{b'} \theta'_{n,\epsilon}(i) \\
&\leq \lim_{n \to \infty} \sum_{i \in V_n} \partial_b \theta_{n,\epsilon}(i) - \partial_{b'} \theta'_{n,\epsilon}(i) \\
&\leq \|b - b'\|_{l^1(V(H))} .
\end{aligned}$$

Since this holds for every finite subset $K \subset V(H)$, we conclude that

$$\|\partial_b \theta_\epsilon - \partial_{b'} \theta'_\epsilon\|_{l^1(V(H))} \leq \|b - b'\|_{l^1(V(H))} .$$

Now, continuing to assume that H is bounded, suppose the condition $b \geq b'$ does not necessarily hold. Define $b'' := b \wedge b'$ with θ''_ϵ being the unique ϵ–balanced allocation on H with respect to b''. Due to monotonicity property of Proposition 6.5, we have $\partial_b \theta_\epsilon \geq \partial_{b''}\theta''_\epsilon$ and $\partial_{b'}\theta'_\epsilon \geq \partial_{b''}\theta''_\epsilon$. Thus, using the above argument and the triangle inequality, we have

$$\begin{aligned}
\|\partial_b\theta_\epsilon - \partial_{b'}\theta'_\epsilon\|_{l^1(V(H))} &\leq \|\partial_b\theta_\epsilon - \partial_{b''}\theta''_\epsilon\|_{l^1(V(H))} + \|\partial_{b'}\theta'_\epsilon - \partial_{b''}\theta''_\epsilon\|_{l^1(V(H))} \\
&\leq \|b - b''\|_{l^1(V(H))} + \|b' - b''\|_{l^1(V(H))} \\
&= \sum_{i \in V(H)} b(i) - b''(i) + b'(i) - b''(i) \\
&= \sum_{i \in V(H)} |b(i) - b'(i)| = \|b - b'\|_{l^1(V(H))}.
\end{aligned}$$

Finally, we relax the boundedness assumption on H.

We take a not necessarily bounded hypergraph H and let H^Δ be the truncation of H, as defined in Section 6.4.3. Let θ^Δ_ϵ and $\theta'^{,\Delta}_\epsilon$ be the unique ϵ–balanced allocations on H^Δ, with respect to the baseloads b and b', respectively. Since H^Δ is bounded, using the above argument, we have

$$\left\|\partial_b\theta^\Delta_\epsilon - \partial_{b'}\theta'^{,\Delta}_\epsilon\right\|_{l^1(V(H))} \leq \|b - b'\|_{l^1(V(H))}.$$

Hence, for any finite subset $K \subseteq V(H)$, we have

$$\sum_{i \in K} |\partial_b\theta^\Delta_\epsilon(i) - \partial_{b'}\theta'^{,\Delta}_\epsilon(i)| \leq \left\|\partial_b\theta^\Delta_\epsilon - \partial_{b'}\theta'^{,\Delta}_\epsilon\right\|_{l^1(V(H))} \leq \|b - b'\|_{l^1(V(H))}.$$

Sending Δ to infinity and using the facts that $\partial\theta^\Delta_\epsilon$ and $\partial\theta'^{,\Delta}_\epsilon$ converge to $\partial\theta_\epsilon$ and $\partial\theta'_\epsilon$, respectively, and also the fact that K is finite, we have

$$\sum_{i \in K} |\partial_b\theta_\epsilon(i) - \partial_{b'}\theta'_\epsilon(i)| \leq \|b - b'\|_{l^1(V(H))}.$$

Since the above is true for all finite K, we have

$$\|\partial_b\theta_\epsilon - \partial_{b'}\theta'_\epsilon\|_{l^1(V(H))} \leq \|b - b'\|_{l^1(V(H))},$$

and the proof is complete. □

6.4.5 Regularity property for canonical ϵ-balanced allocations with respect to a baseload

In this section, we give a regularity property of canonical allocations which is crucial in our analysis.

Let T be a hypertree. For a node $i \in V(T)$ and $e \in E(T)$, $e \ni i$, define $T_{e \to i}$ to the connected subtree with root i that does not contain the part of T directed from e. To be more precise, the vertex set of $T_{e \to i}$ is the set of vertices $j \in V(T)$ such that the shortest path from i to j does not contain e. The edge set of $T_{e \to i}$ contains all the edges with all their end points in this subset, i.e. $E_T(V(T_{e \to i}))$ in the notation of Section 6.2.1.

Proposition 6.8. *Let T be a hypertree, b a baseload function on T, and θ the canonical ϵ–balanced allocation on T with respect to the baseload b. Let $e \in E(T)$ and $i \in e$, and let $\theta_{T_{e\to i}}$ denote the restriction of θ to $T_{e\to i}$, i.e.*

$$\theta_{T_{e\to i}}(e', i') = \theta(e', i') , \qquad e' \in E(T_{e\to i}), i' \in e'.$$

Then, $\theta_{T_{e\to i}}$ is the canonical ϵ–balanced allocation on $T_{e\to i}$ with respect to the baseload function $\tilde{b}$ defined as

$$\tilde{b}(i) = b(i) + \theta(e, i),$$

and $\tilde{b}(j) = b(j)$ for $j \in V(T_{e\to i}) \setminus \{i\}$.

Proof. It is straightforward to check that $\theta_{T_{e\to i}}$ is an ϵ–balanced allocation on $T_{e\to i}$ with baseload function $\tilde{b}$. Thus, the content of the theorem is the statement about this ϵ–balanced allocation being the canonical ϵ–balanced allocation on $T_{e\to i}$ with the baseload function $\tilde{b}$.

Let $\tilde{\theta}$ denote the canonical ϵ-balanced allocation on $T_{e\to i}$ with respect to the baseload $\tilde{b}$. Moreover, let $\tilde{\theta}^{\Delta}$ denote the unique ϵ-balanced allocation on the bounded tree $(T_{e\to i})^{\Delta}$ with respect to the baseload $\tilde{b}$. Throughout this proof, we assume that $\Delta > \Delta_0$ where

$$\Delta_0 := |e| \vee \max\{\deg_H(j) : j \in e\}.$$

If Δ satisfies this property, then $e \in E(T^{\Delta})$, and one can also check that

$$(T_{e\to i})^{\Delta} \equiv (T^{\Delta})_{e\to i}, \tag{6.23}$$

so we can unambiguously write $T^{\Delta}_{e\to i}$ for this tree. (Note that T^{Δ} need not be connected, but this is not relevant.) If θ^{Δ} denotes the unique ϵ–balanced allocation on T^{Δ} with respect to the baseload b and $\theta^{\Delta}_{T_{e\to i}}$ is its restriction to $T^{\Delta}_{e\to i}$, then, since $\Delta \geq \Delta_0$, $\theta^{\Delta}_{T_{e\to i}}$ is the unique ϵ-balanced allocation on $T^{\Delta}_{e\to i}$ with respect to the baseload b^{Δ} defined as $b^{\Delta}(i) = b(i) + \theta^{\Delta}(e, i)$ and $b^{\Delta}(j) = b(j)$ for $j \neq i$. Now, $\tilde{\theta}^{\Delta}$ and $\theta^{\Delta}_{T_{e\to i}}$ are canonical ϵ-balanced allocations on $T^{\Delta}_{e\to i}$ with respect to the baseloads $\tilde{b}$ and b^{Δ}, respectively. Using Proposition 6.7, we conclude that for an arbitrary vertex $k \in V(T_{e\to i})$ we have

$$|\partial_{b^{\Delta}}\theta^{\Delta}_{T_{e\to i}}(k) - \partial_{\tilde{b}}\tilde{\theta}^{\Delta}(k)| \leq \left\| \partial_{b^{\Delta}}\theta^{\Delta}_{T_{e\to i}} - \partial_{\tilde{b}}\tilde{\theta}^{\Delta} \right\|_{l^1(V(T_{e\to i}))} \leq |\theta^{\Delta}(e, i) - \theta(e, i)|.$$

Now, sending Δ to infinity and noting the fact that $\theta^{\Delta}(e, i) \to \theta(e, i)$, we have $|\partial_{b^{\Delta}}\theta^{\Delta}_{T_{e\to i}}(k) - \partial_{\tilde{b}}\tilde{\theta}^{\Delta}(k)| \to 0$. Since $\theta^{\Delta}_{T_{e\to i}}$ is the restriction to $T^{\Delta}_{e\to i}$ of the ϵ-balanced allocation θ^{Δ} on T^{Δ} with respect to the baseload b, we have $\partial_{b^{\Delta}}\theta^{\Delta}_{T_{e\to i}}(k) = \partial_b\theta^{\Delta}(k)$ for all $k \in V(T^{\Delta}_{e\to i})$. Since $\partial_b\theta^{\Delta}(k) \to \partial_b\theta(k)$ and $\partial_{\tilde{b}}\tilde{\theta}^{\Delta}(k) \to \partial_{\tilde{b}}\tilde{\theta}(k)$ for all $k \in V(T)$ as $\Delta \to \infty$, we conclude that

$$\partial_{\tilde{b}}\tilde{\theta}(k) = \partial_b\theta(k) , \qquad \forall k \in V(T_{e\to i}).$$

But it is straightforward to check that $\partial_b\theta(k) = \partial_{\tilde{b}}\theta_{T_{e\to i}}(k)$ for all $k \in V(T_{e\to i})$. From the definition of ϵ-balanced allocations, we conclude that $\theta_{T_{e\to i}}(e', i')$ equals $\tilde{\theta}(e', i')$ for all $e' \in E(T_{e\to i})$ and all $i' \in e'$, i.e. that $\theta_{T_{e\to i}}$ is the canonical ϵ-balanced allocation on $T_{e\to i}$ with respect to the baseload $\tilde{b}$, which was what was to be shown. ∎

6.5 ϵ–balanced allocations on $\mathcal{H}_{**}$

In this section, we discuss how to find an ϵ–balanced allocation on $\mathcal{H}_{**}$, in the sense of Definition 6.25, with respect to a unimodular measure $\mu \in \mathcal{P}(\mathcal{H}_*)$. Recall that this is a Borel allocation $\Theta_\epsilon : \mathcal{H}_{**} \to [0, 1]$ such that

$$\Theta_\epsilon(H, e, i) = \frac{\exp(-\partial\Theta_\epsilon(H, i)/\epsilon)}{\sum_{j\in e} \exp(-\partial\Theta_\epsilon(H, j)/\epsilon)}, \quad \vec{\mu}\text{–a.e..}$$

What we do here is in fact stronger, in the sense that we introduce a Borel allocation Θ_ϵ such that the above condition is satisfied pointwise, i.e.

$$\Theta_\epsilon(H, e, i) = \frac{\exp(-\partial\Theta_\epsilon(H, i)/\epsilon)}{\sum_{j\in e} \exp(-\partial\Theta_\epsilon(H, j)/\epsilon)} \quad \forall [H, e, i] \in \mathcal{H}_{**}. \tag{6.24}$$

In fact, we can define $\Theta_\epsilon(H, e, i)$ to be $\theta^H_\epsilon(e, i)$ where θ^H_ϵ is the canonical ϵ-balanced allocation for H, as was introduced in Section 6.4.3. Defining Θ_ϵ in this way guarantees that (6.24) is satisfied. Therefore, it remains to show that Θ_ϵ is a Borel allocation. In the following, we construct Θ_ϵ differently, but as we will see later, the $\Theta_\epsilon(H, e, i)$ we construct will be equal to $\theta^H_\epsilon(e, i)$ (see Remark 6.12 below).

For the construction, given $\Delta \in \mathbb{N}$, define $F^\Delta_\epsilon : \mathcal{H}_* \to \mathbb{R}$ via $F^\Delta_\epsilon(H, i) := \partial\theta^{H^\Delta}_\epsilon(i)$, where H^Δ is the truncated hypergraph introduced in Section 6.4.3. The uniqueness property of ϵ–balanced allocations for bounded hypergraphs (Corollary 6.1) implies that the above definition does not depend on the specific choice of (H, i) in the equivalence class. Therefore, F^Δ_ϵ is well defined. We claim that F^Δ_ϵ is a continuous function on $\mathcal{H}_*$. Indeed, if $(H_1, i_1) \equiv_d (H_2, i_2)$, then $(H^\Delta_1, i_1) \equiv_{d-1} (H^\Delta_2, i_2)$, and using Lemma 6.7 we have

$$\left|\partial\theta^{H^\Delta_1}_\epsilon(i_1) - \partial\theta^{H^\Delta_2}_\epsilon(i_2)\right| \le \left(\frac{\Delta^2}{4\epsilon + \Delta^2}\right)^{d-1} \Delta\ .$$

This implies that F^Δ_ϵ is (uniformly) continuous. Moreover, Proposition 6.5 implies that F^Δ_ϵ is pointwise increasing in Δ. On the other hand, $F^\Delta_\epsilon(H, i) \le \deg_H(i)$. Hence, there is a pointwise limit

$$F_\epsilon(H, i) := \lim_{\Delta\to\infty} F^\Delta_\epsilon(H, i). \tag{6.25}$$

We now define $\Theta_\epsilon : \mathcal{H}_{**} \to [0, 1]$ in the following way:

$$\Theta_\epsilon(H, e, i) := \frac{\exp(-F_\epsilon(H, i)/\epsilon)}{\sum_{j\in e} \exp(-F_\epsilon(H, j)/\epsilon)}. \tag{6.26}$$

We want to show that Θ_ϵ is an ϵ–balanced allocation on $\mathcal{H}_{**}$. Since F^Δ_ϵ is continuous, F_ϵ is Borel, and hence Θ_ϵ is a Borel map. Also, the normalization factor in the denominator guarantees that $\sum_{i\in e} \Theta_\epsilon(H, e, i) = 1$, and hence Θ_ϵ is a Borel allocation. Comparing (6.24) with (6.26), it suffices to show that

$$\partial\Theta_\epsilon(H, i) = F_\epsilon(H, i). \tag{6.27}$$

In order to show this, note that

$$\begin{aligned}\partial\Theta_\epsilon(H,i) &= \sum_{e\ni i}\Theta_\epsilon(H,e,i)\\ &= \sum_{e\ni i}\frac{\exp(-F_\epsilon(H,i)/\epsilon)}{\sum_{j\in e}\exp(-F_\epsilon(H,j)/\epsilon)}\\ &= \lim_{\Delta\to\infty}\sum_{e\ni i}\frac{\exp(-F_\epsilon^\Delta(H,i)/\epsilon)}{\sum_{j\in e}\exp(-F_\epsilon^\Delta(H,j)/\epsilon)}.\end{aligned}$$

Now, when $\Delta \geq \max_{e\ni i}|e|$ and also $\Delta \geq \max_{j\in e, e\ni i}\deg_H(j))$, we have $e \in E(H^\Delta)$, and we have

$$\frac{\exp(-F_\epsilon^\Delta(H,i)/\epsilon)}{\sum_{j\in e}\exp(-F_\epsilon^\Delta(H,j)/\epsilon)} = \frac{\exp(-\partial\theta_\epsilon^{H^\Delta}(H,i)/\epsilon)}{\sum_{j\in e}\exp(-\partial\theta_\epsilon^{H^\Delta}(H,j)/\epsilon)} = \theta_\epsilon^{H^\Delta}(H,e,i).$$

Consequently,

$$\begin{aligned}\partial\Theta_\epsilon(H,i) &= \lim_{\Delta\to\infty}\sum_{e\ni i}\theta_\epsilon^{H^\Delta}(H,e,i)\\ &= \lim_{\Delta\to\infty}\partial\theta_\epsilon^{H^\Delta}(H,i)\\ &= \lim_{\Delta\to\infty}F_\epsilon^\Delta(H,i)\\ &= F_\epsilon(H,i).\end{aligned}$$

Therefore, we have shown (6.27), which shows that Θ_ϵ satisfies (6.24) and hence is an ϵ–balanced allocation (both pointwise and $\vec{\mu}$–almost everywhere).

Remark 6.12. *Note that* (6.25) *means that* $F_\epsilon(H,i) = \partial\theta_\epsilon^H(i)$, *where* θ_ϵ^H *is the canonical* ϵ*-balanced allocation in* H*. Hence,* (6.26) *implies that* $\Theta_\epsilon(H,e,i)$ *is pointwise equal to* $\theta_\epsilon^H(e,i)$.

The following Proposition proves an almost sure uniqueness property for Borel ϵ–balanced allocations which is similar in flavor to part 4 of Theorem 6.1. The proof of this statement is given in Appendix E.8.

Proposition 6.9. *Assume* μ *is a unimodular measure on* $\mathcal{H}_*$ *such that* $\deg(\mu) < \infty$*. Given* $\epsilon > 0$*, let* Θ_ϵ *be the* ϵ*–balanced allocation defined in this section, and let* Θ'_ϵ *be any other* ϵ*–balanced allocation, both with respect to* μ*. Then, we have* $\partial\Theta_\epsilon = \partial\Theta'_\epsilon$*,* μ*–a.s. and* $\Theta_\epsilon = \Theta'_\epsilon$*,* $\vec{\mu}$*–a.e..*

6.6 Properties of balanced allocations

6.6.1 Existence

In this section we prove the existence of balanced allocations (part 1 of Theorem 6.1):

Proposition 6.10. *Assume $\mu \in \mathcal{P}(\mathcal{H}_*)$ is unimodular with $\deg(\mu) < \infty$. Then, there is a sequence $\epsilon_k \downarrow 0$ such that Θ_{ϵ_k} converges to a balanced allocation Θ_0, with the convergence being both in $L^2(\vec{\mu})$ and $\vec{\mu}$–almost everywhere.*

Proof. First, we show that Θ_ϵ is Cauchy in $L^2(\vec{\mu})$. To do so, we take $\epsilon, \epsilon' > 0$ and try to bound $\|\Theta_\epsilon - \Theta_{\epsilon'}\|_{L^2(\vec{\mu})}$. For an integer $\Delta > 0$, define the function Θ_ϵ^Δ on $\mathcal{H}_{**}$ as follows:

$$\Theta_\epsilon^\Delta(H, e, i) = \begin{cases} \theta_\epsilon^{H^\Delta}(e, i) & \text{if } e \in E(H^\Delta), \\ 0 & \text{otherwise,} \end{cases} \tag{6.28}$$

where H^Δ is the truncated hypergraph introduced in Section 6.4.3 and $\theta_\epsilon^{H^\Delta}$ is the unique ϵ–balanced allocation associated to it.

Take a locally finite hypergraph H and an edge e in $E(H^\Delta)$. As $\theta_\epsilon^{H^\Delta}$ is ϵ–balanced on H^Δ and $e \in E(H^\Delta)$, for $i, j \in e$ we have

$$\frac{\theta_\epsilon^{H^\Delta}(e, i)}{\theta_\epsilon^{H^\Delta}(e, j)} = \frac{\exp(-\partial\theta_\epsilon^{H^\Delta}(i)/\epsilon)}{\exp(-\partial\theta_\epsilon^{H^\Delta}(j)/\epsilon)}.$$

By the definition of Θ_ϵ^Δ, if $e \in E(H^\Delta)$, then for all $j' \in e$, $\Theta_\epsilon^\Delta(H, e, j') = \theta_\epsilon^{H^\Delta}(e, j')$ and $\partial\Theta_\epsilon^\Delta(H, j') = \partial\theta_\epsilon^{H^\Delta}(j')$. Taking logarithms on both sides of the above equation, we have

$$\partial\Theta_\epsilon^\Delta(H, i) + \epsilon \log \Theta_\epsilon^\Delta(H, e, i) = \partial\Theta_\epsilon^\Delta(H, j) + \epsilon \log \Theta_\epsilon^\Delta(H, e, j), \qquad \forall i, j \in e,$$

with this equation holding pointwise whenever $e \in E(H^\Delta)$. As this equality holds for all $i, j \in e$, any two convex combinations of the values of $\partial\Theta_\epsilon^\Delta + \epsilon \log \Theta_\epsilon^\Delta$ evaluated at nodes in e are equal, whenever $e \in E(H^\Delta)$. In particular, if $\Theta_{\epsilon'}^\Delta$ denotes the ϵ'–balanced allocation on $\mathcal{H}_{**}$ defined similarly to the above, then, whenever $e \in E(H^\Delta)$, we have $\sum_{i \in e} \Theta_\epsilon^\Delta(H, e, i) = \sum_{i \in e} \Theta_{\epsilon'}^\Delta(H, e, i) = 1$. Hence we have:

$$\begin{aligned} \sum_{i \in e} & \Theta_\epsilon^\Delta(H, e, i) \left(\partial\Theta_\epsilon^\Delta(H, i) + \epsilon \log \Theta_\epsilon^\Delta(H, e, i)\right) \\ & = \sum_{i \in e} \Theta_{\epsilon'}^\Delta(H, e, i) \left(\partial\Theta_\epsilon^\Delta(H, i) + \epsilon \log \Theta_\epsilon^\Delta(H, e, i)\right), \end{aligned} \tag{6.29}$$

which holds pointwise, whenever $e \in E(H^\Delta)$. On the other hand, if $e \notin E(H^\Delta)$, then, by definition, $\Theta_\epsilon^\Delta(H, e, j)$ as well as $\Theta_{\epsilon'}^\Delta(H, e, j)$ are zero for all $j \in e$. In this case the above equality again holds, with $0 \log 0$ being interpreted as 0. In other words, pointwise on $\mathcal{H}_{**}$, we have

$$\nabla(\Theta_\epsilon^\Delta(\partial\Theta_\epsilon^\Delta + \epsilon \log \Theta_\epsilon^\Delta)) = \nabla(\Theta_{\epsilon'}^\Delta(\partial\Theta_\epsilon^\Delta + \epsilon \log \Theta_\epsilon^\Delta)),$$

where, by abuse of notation, we have treated $\partial\Theta_\epsilon^\Delta$ as a function on $\mathcal{H}_{**}$ rather than on $\mathcal{H}_*$, via $\partial\Theta_\epsilon^\Delta(H, e, i) := \partial\Theta_\epsilon^\Delta(H, i)$, and likewise for $\partial\Theta_{\epsilon'}^\Delta$. Rewriting the above identity, we have:

$$\nabla\left(\Theta_\epsilon^\Delta \partial\Theta_\epsilon^\Delta + \epsilon\Theta_\epsilon^\Delta \log \Theta_\epsilon^\Delta - \Theta_{\epsilon'}^\Delta \partial\Theta_\epsilon^\Delta\right) = \epsilon\nabla(\Theta_{\epsilon'}^\Delta \log \Theta_\epsilon^\Delta). \tag{6.30}$$

Note that $\partial\Theta^\Delta_\epsilon$ is pointwise bounded by Δ by definition. Moreover, $\deg(\mu) < \infty$, which implies that $\vec{\mu}$ has finite total measure. Hence, all terms in the above equation have finite integral. On the other hand, from the definition of $\vec{\mu}$, we have:

$$\int (\Theta^\Delta_\epsilon - \Theta^\Delta_{\epsilon'})\partial\Theta^\Delta_\epsilon d\vec{\mu} = \int \partial((\Theta^\Delta_\epsilon - \Theta^\Delta_{\epsilon'})\partial\Theta^\Delta_\epsilon)d\mu = \int (\partial\Theta^\Delta_\epsilon - \partial\Theta^\Delta_{\epsilon'})\partial\Theta^\Delta_\epsilon d\mu.$$

Substituting this into (6.30) and using unimodularity, we have

$$\langle \partial\Theta^\Delta_\epsilon - \partial\Theta^\Delta_{\epsilon'}, \partial\Theta^\Delta_\epsilon \rangle + \epsilon \int \Theta^\Delta_\epsilon \log \Theta^\Delta_\epsilon d\vec{\mu} = \epsilon \int \Theta^\Delta_{\epsilon'} \log \Theta^\Delta_\epsilon d\vec{\mu}, \tag{6.31}$$

where $\langle ., . \rangle$ denotes the inner product of two functions in $L^2(\mu)$. Now, changing the order of ϵ and ϵ', we have

$$\langle \partial\Theta^\Delta_{\epsilon'} - \partial\Theta^\Delta_\epsilon, \partial\Theta^\Delta_{\epsilon'} \rangle + \epsilon' \int \Theta^\Delta_{\epsilon'} \log \Theta^\Delta_{\epsilon'} d\vec{\mu} = \epsilon' \int \Theta^\Delta_\epsilon \log \Theta^\Delta_{\epsilon'} d\vec{\mu}. \tag{6.32}$$

Summing up these two equalities, we have:

$$\begin{aligned} \left\|\partial\Theta^\Delta_\epsilon - \partial\Theta^\Delta_{\epsilon'}\right\|_2^2 = \epsilon \int \Theta^\Delta_{\epsilon'} \log \Theta^\Delta_\epsilon d\vec{\mu} - \epsilon \int \Theta^\Delta_\epsilon \log \Theta^\Delta_\epsilon d\vec{\mu} \\ + \epsilon' \int \Theta^\Delta_\epsilon \log \Theta^\Delta_{\epsilon'} d\vec{\mu} - \epsilon' \int \Theta^\Delta_{\epsilon'} \log \Theta^\Delta_{\epsilon'} d\vec{\mu}. \end{aligned} \tag{6.33}$$

We now use the following information theoretic notation. For functions $\Theta, \Theta' : \mathcal{H}_{**} \to [0, 1]$, we define

$$\begin{aligned} H(\Theta) &:= -\int \Theta \log \Theta d\vec{\mu}, \text{ and} \\ D(\Theta\|\Theta') &:= \int \Theta \log \frac{\Theta}{\Theta'} d\vec{\mu}, \end{aligned}$$

where $0 \log 0$ is interpreted as 0, and $\Theta(H, e, i)$ is assumed to be zero whenever $\Theta'(H, e, i)$ is zero (by definition, Θ^Δ_ϵ and $\Theta^\Delta_{\epsilon'}$ have this property). Rearranging the terms in (6.33),

$$\left\|\partial\Theta^\Delta_\epsilon - \partial\Theta^\Delta_{\epsilon'}\right\|_2^2 + \epsilon D(\Theta^\Delta_{\epsilon'}\|\Theta^\Delta_\epsilon) + \epsilon' D(\Theta^\Delta_\epsilon\|\Theta^\Delta_{\epsilon'}) = (\epsilon - \epsilon')(H(\Theta^\Delta_\epsilon) - H(\Theta^\Delta_{\epsilon'})). \tag{6.34}$$

Note that, since $\deg(\mu) < \infty$ and $\partial\Theta^\Delta_\epsilon$ and $\partial\Theta^\Delta_{\epsilon'}$ are pointwise bounded to Δ, all the terms are finite. With, $\mathcal{A} := \{[H, e, i] \in \mathcal{H}_{**} : e \in E(H^\Delta)\}$, we have

$$\begin{aligned} D(\Theta^\Delta_{\epsilon'}\|\Theta^\Delta_\epsilon) &= \int_{\mathcal{A}} \Theta^\Delta_{\epsilon'} \log \frac{\Theta^\Delta_{\epsilon'}}{\Theta^\Delta_\epsilon} d\vec{\mu} \\ &= \int_{\mathcal{A}} \nabla \left(\Theta^\Delta_{\epsilon'} \log \frac{\Theta^\Delta_{\epsilon'}}{\Theta^\Delta_\epsilon} \right) d\vec{\mu} \\ &= \int_{\mathcal{A}} \frac{1}{|e|} D_{\mathrm{KL}}((\Theta^\Delta_{\epsilon'}(H, e, j))_{j \in e} \| (\Theta^\Delta_\epsilon(H, e, j))_{j \in e}) d\vec{\mu}(H, e, i) \\ &\geq 0, \end{aligned} \tag{6.35}$$

where D_{KL} denotes the standard KL divergence, and for $[H,e,i] \in \mathcal{A}$, by definition, $(\Theta^\Delta_{\epsilon'}(H,e,j))_{j\in e}$ and $(\Theta^\Delta_{\epsilon'}(H,e,j))_{j\in e}$ are vectors of nonnegative values summing up to one. Similarly, one can show that $D(\Theta^\Delta_\epsilon \| \Theta^\Delta_{\epsilon'}) \geq 0$. Therefore, all the terms on the left hand side of (6.34) are nonnegative. As a result

$$\epsilon > \epsilon' \qquad \Rightarrow \qquad H(\Theta^\Delta_\epsilon) \geq H(\Theta^\Delta_{\epsilon'}). \tag{6.36}$$

As $H(\Theta^\Delta_\epsilon) \geq 0$ for all $\epsilon > 0$, the above inequality implies that $H(\Theta^\Delta_\epsilon)$ converges to some value as $\epsilon \downarrow 0$. Another result of (6.34) is that for $\epsilon > \epsilon'$,

$$\epsilon D(\Theta^\Delta_{\epsilon'} \| \Theta^\Delta_\epsilon) \leq (\epsilon - \epsilon')(H(\Theta^\Delta_\epsilon) - H(\Theta^\Delta_{\epsilon'})).$$

Dividing by ϵ and using $\epsilon > \epsilon'$ and $H(\Theta^\Delta_\epsilon) \geq H(\Theta^\Delta_{\epsilon'})$, we get

$$D(\Theta^\Delta_{\epsilon'} \| \Theta^\Delta_\epsilon) \leq H(\Theta^\Delta_\epsilon) - H(\Theta^\Delta_{\epsilon'}). \tag{6.37}$$

Using (6.35) and Pinsker's inequality (see, for instance, [CK11]),

$$\begin{aligned}
D(\Theta^\Delta_{\epsilon'} \| \Theta^\Delta_\epsilon) &= \int_{\mathcal{A}} \frac{1}{|e|} D_{\text{KL}}((\Theta^\Delta_{\epsilon'}(H,e,j))_{j\in e} \| (\Theta^\Delta_\epsilon(H,e,j))_{j\in e}) d\vec{\mu} \\
&\stackrel{(a)}{\geq} \int_{\mathcal{A}} \frac{1}{|e|} \frac{1}{2} \sum_{j\in e} |\Theta^\Delta_{\epsilon'}(H,e,j) - \Theta^\Delta_\epsilon(H,e,j)|^2 d\vec{\mu} \\
&\stackrel{(b)}{=} \frac{1}{2} \int_{\mathcal{A}} |\Theta^\Delta_{\epsilon'} - \Theta^\Delta_\epsilon|^2 d\vec{\mu} \\
&= \frac{1}{2} \left\| \Theta^\Delta_\epsilon - \Theta^\Delta_{\epsilon'} \right\|_2^2,
\end{aligned}$$

where (a) uses the Pinsker's inequality, and (b) uses unimodularity of μ. Combining this with (6.37), for $\epsilon, \epsilon' > 0$ we have

$$\left\| \Theta^\Delta_\epsilon - \Theta^\Delta_{\epsilon'} \right\|_2^2 \leq 2|H(\Theta^\Delta_\epsilon) - H(\Theta^\Delta_{\epsilon'})|.$$

Now, as we send Δ to infinity, Θ^Δ_ϵ converges pointwise to the function Θ_ϵ defined in Section 6.5. Moreover, $\deg(\mu) < \infty$ and the function $x \log x$ is bounded for $x \in [0,1]$. Hence, using dominated convergence theorem, we have

$$\|\Theta_\epsilon - \Theta_{\epsilon'}\|_2^2 \leq 2|H(\Theta_\epsilon) - H(\Theta_{\epsilon'})|. \tag{6.38}$$

Similarly, sending $\Delta \to \infty$, (6.36) implies that

$$\epsilon > \epsilon' > 0 \qquad \Rightarrow \qquad H(\Theta_\epsilon) \geq H(\Theta_{\epsilon'}). \tag{6.39}$$

This means that, as $H(\Theta_\epsilon) \geq 0$ for all $\epsilon > 0$, $H(\Theta_\epsilon)$ is convergent as $\epsilon \downarrow 0$. In particular, this together with (6.38) implies that there is a sequence of positive values ϵ_k converging to zero such that Θ_{ϵ_k} converges to some $\Theta_0 : \mathcal{H}_{**} \to [0,1]$ in $L^2(\vec{\mu})$. Therefore, there is a subsequence of this sequence converging to Θ_0 $\vec{\mu}$–almost everywhere. Without loss of generality, we may assume this subsequence is the whole sequence. With this, Θ_{ϵ_k} converges to Θ_0 both in $L^2(\vec{\mu})$ and $\vec{\mu}$–almost everywhere. Using Lemma 6.3, we have $\partial\Theta_{\epsilon_k} \to \partial\Theta_0$ μ–almost surely. Lemma 6.3 then implies that for μ–almost all $[H,i] \in \mathcal{H}_*$, $\partial\Theta_{\epsilon_k}(H,j) \to \partial\Theta_0(H,j)$ for all $j \in e$. Following Remark 6.6, we can treat $\partial\Theta_0$ and the $\partial\Theta_{\epsilon_k}$ as functions on $\mathcal{H}_{**}$ instead of $\mathcal{H}_*$. Then, Lemma 6.2 implies that

$$\partial\Theta_{\epsilon_k}(H,e,j) \to \partial\Theta_0(H,e,j) \ \forall j \in e, \qquad \vec{\mu}\text{–a.e..} \tag{6.40}$$

Now, we are ready to show that Θ_0 is actually $\vec{\mu}$–balanced. To do so, assume that $\partial\Theta_0(H,i) > \partial\Theta_0(H,j)$ for some $i,j \in e$. Equivalently, $\partial\Theta_0(H,e,i) > \partial\Theta_0(H,e,j)$. Then (6.40) implies that, outside a measure zero set, for some fixed δ, and for k large enough,

$$\partial\Theta_{\epsilon_k}(H,e,i) - \partial\Theta_{\epsilon_k}(H,e,j) > \delta \qquad \vec{\mu}\text{–a.e..}$$

On the other hand, using the definition of an ϵ–balanced allocation, we have

$$\Theta_{\epsilon_k}(H,e,i) \le \frac{1}{1+\exp\left(-\frac{\partial\Theta_{\epsilon_k}(H,e,j)-\partial\Theta_{\epsilon_k}(H,e,i)}{\epsilon_k}\right)} \le \frac{1}{1+\exp(\delta/\epsilon_k)}.$$

Sending k to infinity, since δ is fixed, the above inequality implies that $\Theta_{\epsilon_k}(H,e,i)$ converges to zero. Also, we know that $\Theta_{\epsilon_k} \to \Theta_0$ $\vec{\mu}$–almost everywhere. Thus, we have shown that

$$\partial\Theta_0(H,i) > \partial\Theta_0(H,j) \text{ for } i,j \in e \qquad \Rightarrow \qquad \Theta_0(H,e,i) = 0, \qquad \vec{\mu}\text{–a.e.,}$$

which shows that Θ_0 is balanced and the proof is complete. □

6.6.2 Variational characterization

In this section, we prove the variational characterization (part 2) of Theorem 6.1.

Proposition 6.11. *Assume Θ is a Borel allocation on $\mathcal{H}_{**}$ and $\mu \in \mathcal{P}(\mathcal{H}_*)$ is unimodular with $\deg(\mu) < \infty$. Then, for all $t \in \mathbb{R}$ we have*

$$\int (\partial\Theta - t)^+ d\mu \ge \sup_{\substack{f:\mathcal{H}_* \to [0,1] \\ \text{Borel}}} \int \tilde{f}_{min} d\vec{\mu} - t \int f d\mu, \tag{6.41}$$

where $\tilde{f}_{min}$ is defined as

$$\tilde{f}_{min}(H,e,i) = \frac{1}{|e|} \min_{j \in e} f(H,j).$$

Furthermore, equality happens for all $t \in \mathbb{R}$ if and only if Θ is balanced. Moreover, when Θ is balanced, the function $f = \mathbb{1}_{\partial\Theta > t}$ achieves the supremum.

Proof. Note that for any real number x, we have $x^+ \geq xy$ for $y \in [0,1]$. Therefore, for any Borel function $f : \mathcal{H}_* \to [0,1]$, we have

$$\int (\partial\Theta - t)^+ d\mu \geq \int (\partial\Theta - t) f d\mu = \int f \partial\Theta d\mu - t \int f d\mu. \tag{6.42}$$

The assumption that $\deg(\mu) < \infty$ guarantees that both integrals on the RHS are finite, and hence $\int (\partial\theta - t) f d\mu$ exists. It is easy to see that $f\partial\Theta = \partial(\tilde{f}\Theta)$, where $\tilde{f} : \mathcal{H}_{**} \to \mathbb{R}$ is defined as $\tilde{f}(H, e, i) := f(H, i)$. Therefore, using the unimodularity of μ, we have

$$\int f \partial\Theta d\mu = \int \tilde{f}\Theta d\vec{\mu} = \int \nabla(\tilde{f}\Theta) d\vec{\mu}. \tag{6.43}$$

Now, we have

$$\begin{aligned} \nabla(\tilde{f}\Theta)(H, e, i) &= \frac{1}{|e|} \sum_{j \in e} \tilde{f}(H, e, j)\Theta(H, e, j) \\ &= \frac{1}{|e|} \sum_{j \in e} f(H, j)\Theta(H, e, j) \\ &\overset{(a)}{\geq} \frac{1}{|e|} \min_{j \in e} f(H, j) \\ &= \tilde{f}_{\min}(H, e, i), \end{aligned} \tag{6.44}$$

where (a) holds since $\sum_{j \in e} \Theta(H, e, j) = 1$ and $\Theta(H, e, j) \geq 0$ for all $j \in e$. This, together with (6.42) and (6.43), proves (6.41).

Now, we show that equality holds if and only if Θ is balanced. First, assume Θ is balanced. We show that equality holds in (6.41) for all $t \in \mathbb{R}$. Take $f = \mathbb{1}_{\partial\Theta > t}$. Then, (6.42) becomes an equality. Therefore, it remains to show that (6.44) also becomes an equality, $\vec{\mu}$–almost everywhere. Note that if $f(H, j) = 0$ for all $j \in e$ or $f(H, j) = 1$ for all $j \in e$, equality holds in (6.44). Thereby, it suffices to consider the case that for some $j \in e$, $f(H, j) = 0$ and for some other $j' \in e$, $f(H, j') = 1$. This implies $\partial\Theta(H, j') \geq t > \Theta(H, j)$. As Θ is balanced, outside a measure zero set, we can conclude from the above that $\Theta(H, e, j') = 0$, and so its contribution in (6.44) vanishes. Since this is true for any j' with $f(H, j') = 1$, both sides of the inequality (a) in (6.44) become equal to zero. As this argument holds outside a measure zero set, the above discussion shows that (6.41) becomes an equality when Θ is balanced. Moreover, the function $f = \mathbb{1}_{\partial\Theta > t}$ achieves the supremum.

Now, we show that if (6.41) is an equality for all $t \in \mathbb{R}$, then Θ is balanced. Fix some $t \in \mathbb{R}$. First, we claim that the supremum on the RHS of (6.41) is a maximum. To see this, note that we have already shown in Section 6.6.1 that a balanced allocation Θ_0 with respect to μ exists, and the above discussion then implies that $f = \mathbb{1}_{\partial\Theta_0 > t}$ achieves the equality in (6.41) with Θ being replaced with Θ_0. But the RHS of (6.41) has no dependence on Θ. Thereby, the RHS of (6.41) always has $f := \mathbb{1}_{\partial\Theta_0 > t}$ as a maximizer, whenever μ is unimodular with bounded mean. Since $f \in [0,1]$, we have $(\partial\Theta - t)^+ \geq (\partial\Theta - t)f$ pointwise. However, equality holds in (6.41), so by (6.44), we must have

$$(\partial\Theta - t)f = (\partial\Theta - t)^+ \qquad \mu\text{–a.s..} \tag{6.45}$$

Using Lemma 6.1, this means that

$$(\partial\Theta - t)\tilde{f} = (\partial\Theta - t)^+ \qquad \vec{\mu}\text{–a.e.,} \tag{6.46}$$

where, by abuse of notation, we have treated $\partial\Theta$ as being defined on $\mathcal{H}_{**}$ via $\partial\Theta(H,e,i) = \partial\Theta(H,i)$. On the other hand, (6.44) must be an equality and

$$\nabla(\tilde{f}\Theta) = \tilde{f}_{\min} \qquad \vec{\mu}\text{–a.e..} \tag{6.47}$$

Now, we show that Θ must be balanced by checking the conditions in Defintion 6.24. For the above fixed t, if for some $[H,e,i] \in \mathcal{H}_{**}$ and some $j \in e$ we have

$$\partial\Theta(H,e,i) > t > \partial\Theta(H,e,j),$$

then (6.46) implies that outside a measure zero set, we can conclude that $f(H,i) = 1$ and $f(H,j) = 0$. This can be seen by comparing the left hand side and right hand side of (6.46) in each of the two cases, and then recalling that $\tilde{f}(H,e,i) = f(H,i)$.

Hence, by definition, $\tilde{f}_{\min}(H,e,i) = 0$. Therefore, (6.47) implies that outside a measure zero set, we have

$$\begin{aligned} 0 = \nabla(\tilde{f}\Theta)(H,e,i) &= \frac{1}{|e|}\sum_{k\in e} f(H,k)\Theta(H,e,k) \\ &\geq \frac{1}{|e|} f(H,i)\Theta(H,e,i) \\ &= \frac{1}{|e|}\Theta(H,e,i), \end{aligned}$$

which implies that $\Theta(H,e,i) = 0$.

So far we have shown that, for a fixed t, for almost all $[H,e,i] \in \mathcal{H}_{**}$, if for some $j \in e$ we have $\Theta(H,e,i) > t > \Theta(H,e,j)$, then $\Theta(H,e,i) = 0$. Since this holds for all $t \in \mathbb{Q}$, and $\mathbb{Q}$ is countable and dense in $\mathbb{R}$, we can conclude that for $\vec{\mu}$–almost every $[H,e,i] \in \mathcal{H}_{**}$, $\partial\Theta(H,i) > \partial\Theta(H,j)$ for some $j \in e$ implies that $\Theta(H,e,i) = 0$. Using Proposition 6.3, we can conclude that for $\vec{\mu}$–almost all $[H,e,i] \in \mathcal{H}_{**}$, $\partial\Theta(H,e,j_1) \geq \partial\Theta(H,e,j_2)$ for some $j_1, j_2 \in e$ implies that $\partial\Theta(H,j_1) = 0$. Thus, Θ is balanced. □

6.6.3 Optimality

In this section, we prove part 3 of Theorem 6.1. We do this in three steps:

$\underline{(a) \Rightarrow (c)}$ Let Θ be a balanced allocation with respect to μ. Let Θ' be any other allocation, and $f : [0,\infty) \to [0,\infty)$ be convex. We will show that

$$\int f \circ \partial\Theta d\mu \leq \int f \circ \partial\Theta' d\mu.$$

Using Proposition 6.11, since Θ is balanced and Θ' is a Borel allocation we have, for any $t \in \mathbb{R}$,

$$\int (\partial\Theta - t)^+ d\mu = \sup_{\substack{f:\mathcal{H}_* \to [0,1] \\ \text{Borel}}} \int \tilde{f}_{\min} d\vec{\mu} - t \int f d\mu \leq \int (\partial\Theta' - t)^+ d\mu. \tag{6.48}$$

On the other hand,

$$\int \partial\Theta d\mu = \int \Theta d\vec{\mu} = \int \nabla\Theta d\vec{\mu} = \int \frac{1}{|e|} d\vec{\mu}.$$

The same chain of equalities holds for Θ'. Hence, $\int \partial\Theta d\vec{\mu} = \int \partial\Theta' d\vec{\mu}$. Standard results in the convex ordering of random variables (see Theorem 3.A.1 of [SS07] for instance) show that (6.48) implies that $\Theta \preceq_{\text{cx}} \Theta'$, where $\preceq_{\text{cx}}$ denotes the partial order defined by the property that $\int f \circ \Theta d\mu \leq \int f \circ \Theta' d\mu$ for any convex function f.

$\underline{(c) \Rightarrow (b)}$ Just restrict to the given strictly convex function.

$\underline{(b) \Rightarrow (a)}$ Assume Θ_0 is a balanced allocation w.r.t. μ (from Section 6.6.1 we know it exists). As we showed above, $\int f \circ \partial\Theta_0 d\mu \leq \int f \circ \partial\Theta d\mu$ for the given strictly convex function $f : [0,\infty) \to [0,\infty)$. But Θ is a minimizer. Therefore,

$$\int f \circ \partial\Theta d\mu = \int f \circ \partial\Theta_0 d\mu.$$

Now, define $\Theta' : \mathcal{H}_{**} \to [0,1]$ as $\Theta' = (\Theta + \Theta_0)/2$. Note that Θ' is a Borel allocation. Also, using the convexity of f, we have

$$\int f \circ \partial\Theta' d\mu \leq \frac{1}{2}\left(\int f \circ \partial\Theta d\mu + \int f \circ \partial\Theta_0 d\mu\right) = \int f \circ \partial\Theta_0 d\mu. \tag{6.49}$$

On the other hand, since Θ' is a Borel allocation and Θ_0 is balanced, the $(a) \Rightarrow (c)$ part implies that

$$\int f \circ \partial\Theta_0 d\mu \leq \int f \circ \partial\Theta' d\mu.$$

Hence, we have equality in (6.49). As f is nonnegative, it must be the case that

$$f \circ \partial\Theta' = \frac{f \circ \partial\Theta_0 + f \circ \partial\Theta}{2} \qquad \mu\text{–a.s.}.$$

Now, since f is strictly convex, this implies that $\partial\Theta = \partial\Theta_0$, μ–almost surely. Therefore, for any value of t, we have

$$\int (\partial\Theta - t)^+ d\mu = \int (\partial\Theta_0 - t)^+ d\mu.$$

Comparing this with (6.41), we realize that Θ achieves the equality therein, which means, from Proposition 6.11, that Θ is balanced.

6.6.4 Uniqueness of $\partial\Theta$

In this section we prove part 4 of Theorem 6.1. If Θ is balanced, part 3 of the theorem, which was proved in the preceding section, implies that for a strictly convex function f, $\int f \circ \partial\Theta d\mu$ is minimized. Then, in the proof of the $(b) \Rightarrow (a)$ part of Section 6.6.3 it was shown that $\partial\Theta = \partial\Theta_0$ μ–almost surely. This is precisely what we want to show.

For the converse, assume that $\partial\Theta = \partial\Theta_0$, μ-almost surely. Then, for all t, we have

$$\int (\partial\Theta - t)^+ d\mu = \int (\partial\Theta_0 - t)^+ d\mu,$$

so Proposition 6.11 implies that Θ is balanced, by the same logic that was used at the end of the preceding section.

6.6.5 Continuity with respect to the local weak limit

In this section we prove Part 5 of Theorem 6.1. Prior to that, we need some notation and tools.

Recall the notion of marked hypergraphs from Section 6.2.10. Let $\mathsf{Proj}_{\bar{\mathcal{H}}_*(\Xi)\to\mathcal{H}_*} : \bar{\mathcal{H}}_*(\Xi) \to \mathcal{H}_*$ be the function that removes marks, i.e.

$$\mathsf{Proj}_{\bar{\mathcal{H}}_*(\Xi)\to\mathcal{H}_*}([\bar{H}, i]) = [H, i], \tag{6.50}$$

where H is the underlying hypergraph associated to $\bar{H}$. It can be easily checked that $\mathsf{Proj}_{\bar{\mathcal{H}}_*(\Xi)\to\mathcal{H}_*}$ is a continuous map.

Note that allocations could be considered as marks with values in $[0, 1]$. Hence, we can capture the notion of a balanced allocation using the formalism in Section 6.2.10, via the following definition.

Consider the function $f : \bar{\mathcal{H}}_{**}(\Xi) \to \Xi$ defined as $f(\bar{H}, e, i) = \xi_{\bar{H}}(e, i)$. When Ξ has an additive structure (e.g. $\Xi = [0, 1]$) we may consider $\partial f : \bar{\mathcal{H}}_*(\Xi) \to \Xi$, defined as

$$\partial f(\bar{H}, i) = \sum_{e\in E(\bar{H}), e\ni i} f(\bar{H}, e, i) = \sum_{e\in E(\bar{H}), e\ni i} \xi_{\bar{H}}(e, i). \tag{6.51}$$

By abuse of notation, we may write $\partial\xi_{\bar{H}}(i)$ instead of $\partial f(\bar{H}, i)$ with the above f.

Definition 6.32. *A measure $\bar{\mu} \in \mathcal{P}(\bar{\mathcal{H}}_*([0, 1]))$ is called balanced if for $\vec{\bar{\mu}}$–almost every $[\bar{H}, e, i] \in \bar{\mathcal{H}}_*([0, 1])$, we have*

$$\sum_{j\in e} \xi_{\bar{H}}(e, j) = 1, \tag{6.52}$$

and

$$\partial\xi_{\bar{H}}(j) > \partial\xi_{\bar{H}}(j') \quad \text{for some } j, j' \in e \quad \Longrightarrow \quad \xi_{\bar{H}}(e, j) = 0. \tag{6.53}$$

Before proving the result of this section, we state the following lemmas and postpone their proofs until the end of this section.

Lemma 6.8. *If $\bar{\mu}_n$ is a sequence of balanced probability measures on $\bar{\mathcal{H}}_*([0,1])$ with weak limit $\bar{\mu}$, then $\bar{\mu}$ is also balanced.*

Lemma 6.9. *Assume $\theta_1, \dots, \theta_n$ and $\theta'_1, \dots, \theta'_n$ are non–negative real numbers such that $\sum_i \theta_i = \sum_i \theta'_i = 1$. Also, assume that nonnegative real numbers $l_1, \dots, l_n$ and $l'_1, \dots, l'_n$ are given such that for $1 \le i, j \le n$, $l_i > l_j$ implies $\theta_i = 0$. Similarly, assume that $l'_i > l'_j$ implies $\theta'_i = 0$. Then, we have*

$$\sum_{i=1}^{n} (\theta_i - \theta'_i) \mathbb{1}_{l_i > l'_i} \le 0.$$

Lemma 6.10. *Assume K is a compact subset of $\mathcal{H}_*$ and Ξ is a compact metric space. Define $\bar{K} \subset \bar{\mathcal{H}}_*(\Xi)$ as*

$$\bar{K} := \{[\bar{H}, i] \in \bar{\mathcal{H}}_*(\Xi) : [H, i] \in K\}.$$

Then, $\bar{K}$ is compact in $\bar{\mathcal{H}}_*(\Xi)$.

Proposition 6.12. *Let $\{H_n\}_{n \ge 1}$ be a sequence of finite hypergraphs with local weak limit μ. Then, if $\mathcal{L}_{H_n}$ denotes the distribution of balanced load on H_n with a vertex chosen uniformly at random, and $\mathcal{L}$ is the law of $\partial\Theta$ corresponding to the balanced allocation Θ on μ, we have*

$$\mathcal{L}_{H_n} \Rightarrow \mathcal{L}.$$

Proof. Define θ_n to be a balanced allocation on H_n and $\bar{H}_n$ to be the marked hypergraph obtained by adding θ_n to H_n as edge marks, i.e. $\xi_{\bar{H}_n}(e, i) = \theta_n(e, i)$ for $(e, i) \in \Psi(H)$. Note that H_n is a finite hypergraph; hence, θ_n is well defined and $\partial\theta_n(i)$ is unique for $i \in V(H_n)$. Now, define $\bar{\mu}_n \in \mathcal{P}(\bar{\mathcal{H}}_*([0,1]))$ to be the distribution of $[\bar{H}_n, v]$ where $v \in V(H_n)$ is chosen uniformly at random. We claim that $\{\bar{\mu}_n\}_{n=1}^{\infty}$ is a tight sequence, which means that it has a convergent subsequence. Since μ_n converges weakly to μ, Prokhorov's theorem implies that μ_n is tight in $\mathcal{P}(\mathcal{H}_*)$ (see, for instance, [Bil13, Theorems 5.1 and 5.2]). Consequently, for $\epsilon > 0$, there is a compact set $K \subset \mathcal{H}_*$ such that $\mu_n(K^c) \le \epsilon$ for all n. Define

$$\bar{K} := \{[\bar{H}, i] \in \bar{\mathcal{H}}_*([0,1]) : [H, i] \in K\}.$$

From Lemma 6.10, $\bar{K}$ is compact in $\bar{\mathcal{H}}_*([0,1])$. It is easy to see that $\bar{\mu}_n(\bar{K}^c) = \mu_n(K^c)$ which means that $\bar{\mu}_n$ is a tight sequence. Hence, it has a subsequence converging weakly to some $\bar{\mu} \in \mathcal{P}(\bar{\mathcal{H}}_*)$. In order to simplify the notation, assume that this subsequence is the whole sequence.

With the projection map $\mathsf{Proj}_{\bar{\mathcal{H}}_*([0,1]) \to \mathcal{H}_*}$ defined in (6.7), define ν_n and ν to be the push-forward measures on $\mathcal{H}_*$ corresponding to $\bar{\mu}_n$ and $\bar{\mu}$, respectively. As $\mathsf{Proj}_{\bar{\mathcal{H}}_*([0,1]) \to \mathcal{H}_*}$ removes marks, we have $\nu_n = \mu_n$ and $\nu = \mu$. Now, note that $\bar{\mu}_n \Rightarrow \bar{\mu}$ implies $(\mathsf{Proj}_{\bar{\mathcal{H}}_*([0,1]) \to \mathcal{H}_*})_* \bar{\mu}_n \Rightarrow (\mathsf{Proj}_{\bar{\mathcal{H}}_*([0,1]) \to \mathcal{H}_*})_* \bar{\mu}$ which means that $\mu_n \Rightarrow (\mathsf{Proj}_{\bar{\mathcal{H}}_*([0,1]) \to \mathcal{H}_*})_* \bar{\mu}$. On the other hand, we know $\mu_n \Rightarrow \mu$. Thereby, $(\mathsf{Proj}_{\bar{\mathcal{H}}_*([0,1]) \to \mathcal{H}_*})_* \bar{\mu} = \mu$.

Note that, with the above construction, $\mathcal{L}_{H_n}$ is the pushforward of $\bar{\mu}_n$ under the mapping $[\bar{H}, i] \mapsto \partial\xi_{\bar{H}}(i)$ defined in (6.51). Therefore, as $\bar{\mu}_n \Rightarrow \bar{\mu}$, $\mathcal{L}_{H_n}$ converges weakly to the law of $\partial\xi_{\bar{H}}(i)$ under $\bar{\mu}$. Consequently, to show that $\mathcal{L}_{H_n} \Rightarrow \mathcal{L}$, it suffices to show that

$$\partial\xi_{\bar{H}}(i) = \partial\Theta_0([H, i]) \qquad \bar{\mu}\text{–a.s..} \tag{6.54}$$

From Proposition E.2 in Appendix E.5 we know that $\bar{\mu}$ is unimodular. Hence, we have

$$\begin{aligned}
&\int (\partial\xi_{\bar{H}}(i) - \partial\Theta_0([H, i]))^+ \, d\bar{\mu}(\bar{H}, i) \\
&\quad = \int (\partial_{\bar{H}}\xi(i) - \partial\Theta_0([H, i])) \, \mathbb{1}_{\partial\xi_{\bar{H}}(i) > \partial\Theta_0([H,i])} d\bar{\mu}([\bar{H}, i]) \\
&\quad = \int (\xi_{\bar{H}}(e, i) - \Theta_0([H, e, i])) \, \mathbb{1}_{\partial\xi_{\bar{H}}(i) > \partial\Theta_0([H,i])} d\vec{\bar{\mu}}([\bar{H}, e, i]) \\
&\quad = \int \frac{1}{|e|} \sum_{j \in e} (\xi_{\bar{H}}(e, j) - \Theta_0([H, e, j])) \, \mathbb{1}_{\partial\xi_{\bar{H}}(j) > \partial\Theta_0([H,j])} d\vec{\bar{\mu}}([\bar{H}, e, i),
\end{aligned}$$

where the last equality follows from unimodularity of $\bar{\mu}$. From Lemma 6.8, $\bar{\mu}$ is balanced in the sense of Definition 6.32 above. Also, Lemma 6.9 together with the balancedness of $\bar{\mu}$ and Θ_0, implies that the integrand is non–positive almost everywhere, which means that $\partial\xi_{\bar{H}}(i) \le \partial\Theta_0([H, i])$ $\bar{\mu}$-almost surely. It could be proved in a similar fashion that $\partial\Theta_0(H, i) \le \partial\xi_{\bar{H}}(i)$ $\bar{\mu}$-almost surely. This proves (6.54).

So far, we have shown that $\mathcal{L}_{H_n}$ has a subsequence $\mathcal{L}_{H_{n_k}}$ such that $\mathcal{L}_{H_{n_k}} \Rightarrow \mathcal{L}$. This argument could be repeated for any subsequence of H_n, i.e. any subsequence H_{n_k} has a further subsequence $H_{n_{k_l}}$ such that $\mathcal{L}_{H_{n_{k_l}}} \Rightarrow \mathcal{L}$. This implies that $\mathcal{L}_{H_n} \Rightarrow \mathcal{L}$ (see, for instance, Theorem 2.2 in [Bil71]). □

Proof of Lemma 6.8. First, we show that $\bar{\mu}$ satisfies (6.53). Define

$$A_{**} := \{[\bar{H}, e, i] \in \bar{\mathcal{H}}_{**}([0, 1]) : \forall j, j' \in e, \partial\xi_{\bar{H}}(j) > \partial\xi_{\bar{H}}(j') \Rightarrow \xi_{\bar{H}}(e, j) = 0\},$$

and

$$A_* := \{[\bar{H}, i] \in \bar{\mathcal{H}}_*([0, 1]) : \forall e \ni i, j, j' \in e, \partial\xi_{\bar{H}}(j) > \partial\xi_{\bar{H}}(j') \Rightarrow \xi_{\bar{H}}(e, j) = 0\}.$$

We claim balancedness of $\bar{\mu} \in \mathcal{P}(\mathcal{H}_*([0, 1]))$ is equivalent to $\bar{\mu}(A_*) = 1$. By definition, $\bar{\mu}$ being balanced means $\bar{\mu}(A_{**}) = \bar{\mu}(\bar{\mathcal{H}}_{**}([0, 1]))$. This is equivalent to $\int \mathbb{1}_{A_{**}} d\vec{\bar{\mu}} = \int 1 d\vec{\bar{\mu}}$, or $\int \partial\mathbb{1}_{A_{**}} d\bar{\mu} = \int \deg_H(i) d\bar{\mu}$. But we have $\partial\mathbb{1}_{A_{**}} \le \deg_H(i)$ pointwise. Therefore, $\partial\mathbb{1}_{A_{**}}(H, i, \xi) = \deg_H(i)$, $\bar{\mu}$–almost surely. This is equivalent to $[\bar{H}, e, i] \in A_{**}$ for all $e \ni i$, $\bar{\mu}$–almost surely, or $[\bar{H}, i] \in A_*$ $\bar{\mu}$–almost surely which is in turn equivalent to $\bar{\mu}(A_*) = 1$.

Now, we claim that A_* is closed in $\bar{\mathcal{H}}_*([0, 1])$ (with respect to the topology induced by the distance $\bar{d}_*$ defined in Section 6.2.10). Equivalently, we show that A_*^c is open. Take an arbitrary $[\bar{H}, i] \in A_*^c$. Since $[\bar{H}, i] \notin A_*$, there exists $e \ni i$ and $j, j' \in e$ such that

$\partial\xi_{\bar{H}}(j) > \partial\xi_{\bar{H}}(j')$ but $\xi_{\bar{H}}(e,j) > 0$. If $[\bar{H}', i'] \in \bar{\mathcal{H}}_*([0,1])$ is such that $d := \bar{d}_*([\bar{H}, i], [\bar{H}', i'])$ satisfies $d < 1/4$, then $[H, i] \equiv_3 [H', i']$ (recall that H and H' are the underlying unmarked hypergraphs associated to $\bar{H}$ and $\bar{H}'$, respectively). Moreover, if we define

$$\epsilon = \min\left\{\frac{\partial\xi_{\bar{H}}(j) - \partial\xi_{\bar{H}}(j')}{3}, \frac{\xi_{\bar{H}}(e,j)}{2}\right\},$$

and

$$\Delta = \max_{k : d_H(i,k) \le 3} \deg_H(k),$$

then if

$$d < \frac{1}{1 + \frac{\Delta}{\epsilon}},$$

one can check that $\partial\xi_{\bar{H}'}(\phi(j)) > \partial\xi_{\bar{H}'}(\phi(j'))$ while $\xi_{\bar{H}'}(\phi(e), \phi(j)) > 0$, where ϕ is the local isomorphism between H and H' following from $[H, i] \equiv_3 [H', i']$. This in particular means that $[\bar{H}', i'] \notin A_*$. Consequently, the ball with radius $\min\{1/4, 1/(1 + (\Delta/\epsilon))\}$ around $[\bar{H}, i]$ is in A_*^c. This means that A_* is closed.

Notice that since $\bar{\mu}_n$ is balanced, $\bar{\mu}_n(A_*) = 1$. On the other hand, as $\bar{\mu}_n \Rightarrow \bar{\mu}$ and A_* is closed, we have

$$\bar{\mu}(A_*) \ge \limsup \bar{\mu}_n(A_*) = 1,$$

which means that $\bar{\mu}(A_*) = 1$ and $\bar{\mu}$ is balanced.

Now we turn to showing that $\bar{\mu}$ also satisfies (6.52). Similar to the above approach, if we define

$$B_{**} := \{[\bar{H}, e, i] \in \mathcal{H}_{**}([0,1]) : \sum_{j \in e} \xi_{\bar{H}}(e,j) = 1\},$$

and

$$B_* := \{[\bar{H}, i] \in \mathcal{H}_*([0,1]) : [\bar{H}, e, i] \in B_{**}\, \forall e \ni i\},$$

we can show that for $\vec{\bar{\mu}}$–almost every $[\bar{H}, e, i] \in \mathcal{H}_{**}([0,1])$ we have $\sum_{j \in e} \xi_{\bar{H}}(e,j) = 1$. Hence, all the conditions in Definition 6.32 are satisfied and $\bar{\mu}$ is balanced. □

Proof of Lemma 6.9. Define $L := \min_i l_i$ and $A := \{1 \le i \le n : l_i = L\}$. Likewise, let $L' := \min_i l'_i$ and $A' := \{1 \le i \le n : l'_i = L'\}$. The given condition implies that $\theta_j = 0$ for $j \notin A$ and similarly $\theta'_j = 0$ for $j \notin A'$.

First assume that $L > L'$. In this case, we have

$$\begin{aligned}
\sum_i (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} &\overset{(a)}{=} \sum_{i \in A \cup A'} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} \\
&= \sum_{i \in A \setminus A'} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} + \sum_{i \in A'} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} \\
&\overset{(b)}{=} \sum_{i \in A \setminus A'} \theta_i \mathbb{1}_{L > l'_i} + \sum_{i \in A'} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} \\
&\overset{(c)}{=} \sum_{i \in A \setminus A'} \theta_i \mathbb{1}_{L > l'_i} + \sum_{i \in A'} (\theta_i - \theta'_i) \\
&\le \sum_{i \in A \setminus A'} \theta_i + \sum_{i \in A'} \theta_i - \theta'_i \\
&= 1 - 1 = 0,
\end{aligned}$$

where (a) holds since θ and θ' are zero outside $A \cup A'$, (b) uses $\theta'_i = 0$ for $i \notin A'$, and (c) uses the fact that for $i \in A'$, $l_i \ge L > L' = l'_i$.

Now, for the case $L \le L'$ we have

$$\begin{aligned}
\sum_{i=1}^{n} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} &= \sum_{i \in A} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} + \sum_{i \in A' \setminus A} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} \\
&\overset{(a)}{=} \sum_{i \in A' \setminus A} (\theta_i - \theta'_i)\mathbb{1}_{l_i > l'_i} \\
&\overset{(b)}{=} \sum_{i \in A' \setminus A} -\theta'_i \mathbb{1}_{l_i > l'_i} \\
&\le 0,
\end{aligned}$$

where (a) follows from the fact that for $i \in A$, $l_i = L \le L' \le l'_i$ and (b) employs $\theta_i = 0$ for $i \notin A$. This completes the proof. □

Proof of Lemma 6.10. We take a sequence $[\bar{H}_n, i_n]$ in $\bar{K}$ and try to find a converging subsequence. Notice that $[H_n, i_n]$ is a sequence in K. Therefore, it has a convergent subsequence. Without loss of generality, assume this subsequence is the whole sequence, converging to some $[H, i] \in \mathcal{H}_*$. By reducing to a further subsequence, we may assume that $d_{\mathcal{H}_*}([H_n, i_n], [H, i]) \le 1/(n+1)$. This means that, for all $m \ge n$, we have $(H_m, i_m) \equiv_n (H, i)$. Since (H, i) is locally finite, there are finitely many marks up to level n in (H, i) which all have values in the compact space Ξ. Hence, there is a subsequence where the marks in the first n levels in $\bar{H}_n$ are convergent. With this, we may associate marks to $(H, i)_n$ using the limiting values. Since this is true for all n, via a diagonalization argument we may construct a marked rooted hypergraph $(\bar{H}, i)$ together with a subsequence $(\bar{H}_{m_l}, i_{m_l})$ such that the underlying unmarked hypergraph is identical to H and also, for any integer k, the marks up to depth k in $(\bar{H}_{m_l}, i_{m_l})$ converge to those in $(\bar{H}, i)$ as $m \to \infty$. This means that $[\bar{H}_{m_l}, i_{m_l}] \to [\bar{H}, i]$ and completes the proof. $\square$

6.7 Response Functions

Let H be a fixed hypergraph (not necessary bounded) and $i \in V(H)$ be a vertex in H. Fix $\epsilon > 0$. The response function $\rho^{\epsilon}_{(H,i)} : \mathbb{R} \to \mathbb{R}$ is defined as follows: $\rho^{\epsilon}_{(H,i)}(t)$ is the total load at node i corresponding to the canonical ϵ–balanced allocation with respect to an external load with value t at node i (recall the definition of canonical ϵ–balanced allocations from Section 6.4.3). More precisely, given $t \in \mathbb{R}$, let $b_{t,i} : V(H) \to \mathbb{R}$ be the baseload function such that $b_{t,i}(i) = t$ and $b_{t,i}(j) = 0$ for $j \neq i$. Moreover, let $\theta^{b_{t,i}}_{\epsilon}$ be the canonical ϵ–balanced allocation on H with respect to the baseload $b_{t,i}$, as was defined in Section 6.4.3. We then define

$$\rho^{\epsilon}_{(H,i)}(t) := \partial_{b_{t,i}} \theta^{b_{t,i}}_{\epsilon}(i).$$

It turns out that this function has the following properties:

Proposition 6.13. *Given a vertex rooted hypergraph (H, i), for any $\epsilon > 0$ and $x < y$, we have*

$$0 \le \rho^{\epsilon}_{(H,i)}(y) - \rho^{\epsilon}_{(H,i)}(x) \le y - x. \tag{6.55}$$

Also,

$$0 \le \rho^{\epsilon}_{(H,i)}(x) - x \le \deg_H(i). \tag{6.56}$$

Proof. Let $\theta^{b_{x,i}}_{\epsilon}$ be the canonical ϵ–balanced allocation with respect to the baseload $b_{x,i}$, as defined above. Then, by definition, $\rho^{\epsilon}_{(H,i)}(x) = \partial_{b_{x,i}} \theta^{b_{x,i}}_{\epsilon}(i)$. Let $b_{y,i}$ and $\theta^{b_{y,i}}_{\epsilon}$ be defined similarly, with x being replaced by y. As $x < y$, Proposition 6.6 implies that $\partial_{b_{x,i}} \theta^{b_{x,i}}_{\epsilon}(i) \le \partial_{b_{y,i}} \theta^{b_{y,i}}_{\epsilon}(i)$, which means $\rho^{\epsilon}_{(H,i)}(x) \le \rho^{\epsilon}_{(H,i)}(y)$.

In order to show that $\rho^{\epsilon}_{(H,i)}(y) - \rho^{\epsilon}_{(H,i)}(x) \le y - x$, let $d_{x,i}$ be the baseload function such that $d_{x,i}(i) = 0$ and $d_{x,i}(j) = -x$ for $j \in V(H)$, $j \neq i$. Moreover, let $d_{y,i}$ be the baseload function such that $d_{y,i}(i) = 0$ and $d_{y,i}(j) = -y$ for $j \in V(H)$, $j \neq i$. Let $\theta^{b_{x,i},\Delta}_{\epsilon}$ and $\theta^{b_{y,i},\Delta}_{\epsilon}$ be the canonical ϵ–balanced allocations on H^{Δ} with respect to $b_{x,i}$ and $b_{y,i}$, respectively. It is easy to see that, on H^{Δ}, $\theta^{b_{x,i},\Delta}_{\epsilon}$ and $\theta^{b_{y,i},\Delta}_{\epsilon}$ are ϵ–balanced with respect to $d_{x,i}$ and $d_{y,i}$ as well, respectively. This is because $b_{x,i}(j) = d_{x,i}(j) + x$ and $b_{y,i}(j) = d_{y,i}(j) + y$, for all $j \in V(H)$. As $d_{x,i}(i) = 0$ and $b_{x,i}(i) = x$, we have

$$\partial_{d_{x,i}} \theta^{b_{x,i},\Delta}_{\epsilon}(i) = \partial_{b_{x,i}} \theta^{b_{x,i},\Delta}_{\epsilon}(i) - x.$$

Likewise, $\partial_{d_{y,i}} \theta^{b_{y,i},\Delta}_{\epsilon}(i) = \partial_{b_{y,i}} \theta^{b_{y,i},\Delta}_{\epsilon}(i) - y$. On the other hand, as $d_{x,i}(j) \ge d_{y,i}(j)$ for all $j \in V(H)$, using Proposition 6.5, $\partial_{d_{x,i}} \theta^{b_{x,i},\Delta}_{\epsilon}(i) \ge \partial_{d_{y,i}} \theta^{b_{y,i},\Delta}_{\epsilon}(i)$. Comparing with the above, this means that

$$\partial_{b_{x,i}} \theta^{b_{x,i},\Delta}_{\epsilon}(i) - x \ge \partial_{b_{y,i}} \theta^{b_{y,i},\Delta}_{\epsilon}(i) - y.$$

Sending $\Delta \to \infty$, we have $\partial_{b_{y,i}} \theta_\epsilon^{b_{y,i}}(i) - \partial_{b_{x,i}} \theta_\epsilon^{b_{x,i}} \le y - x$. This means that $\rho^\epsilon_{(H,i)}(y) - \rho^\epsilon_{(H,i)}(x) \le y - x$.

To show (6.56), simply note that $\rho^\epsilon_{(H,i)}(x) - x$ is $\partial \theta_\epsilon^{b_{x,i}}(i)$, which is the sum of $\deg_H(i)$ many numbers in the interval $[0, 1]$, hence is in the interval $[0, \deg_H(i)]$. $\square$

In view of Proposition 6.13, it is convenient to define the following class of functions:

Definition 6.33. *For $C > 0$, let $\Upsilon(C)$ denote the class of functions $g : \mathbb{R} \to \mathbb{R}$ satisfying the following conditions:*

(i) *g is nondecreasing.*

(ii) *g is nonexpansive, i.e. for $x \le y$ we have $g(y) - g(x) \le y - x$.*

(iii) $0 \le g(x) - x \le C$.

In this notation, Proposition 6.13 implies that for any vertex rooted hypergraph (H, i) and $\epsilon > 0$,

$$\rho^\epsilon_{(H,i)}(.) \in \Upsilon(\deg_H(i)).$$

It is easy to check the following properties of this class of functions:

Lemma 6.11. *For any function $g \in \Upsilon(C)$ we have*

(i) *g is continuous.*

(ii) $\lim_{x \to \pm\infty} g(x) = \pm\infty$.

(iii) *For any $t \in \mathbb{R}$, the set $\{x \in \mathbb{R} : g(x) = t\}$ is a nonempty bounded closed interval in $\mathbb{R}$.*

(iv) *The function $g^{-1} : \mathbb{R} \to \mathbb{R}$ defined as*

$$g^{-1}(t) := \max\{x \in \mathbb{R} : g(x) = t\}, \tag{6.57}$$

is well defined, nondecreasing and right–continuous. Moreover, if D denotes its set of discontinuities of g^{-1}, D is countable. Furthermore, the set $\{x : g(x) = t\}$ has exactly one element iff $t \notin D$.

(v) *Let D be the set of discontinuities of g^{-1} and $t \notin D$. If for some a, $a < g^{-1}(t)$, then we have $g(a) < t$. Moreover, if for some a we have $a > g^{-1}(t)$, then we have $g(a) > t$.*

6.7.1 Recursion for Response functions on Hypertrees

Recall from Section 6.4.5 that if T is a hypertree, then for $(e,i) \in \Psi(T)$, $T_{e\to i}$ is obtained by removing e from T and looking at the connected component of the resulting hypertree rooted at i.

Given $x \in \mathbb{R}$, let b_x be the baseload function on $T_{e\to i}$ with $b_x(i) = x$ and $b_x(j) = 0$, $j \neq i$. With this, let $\theta_\epsilon^{b_x}$ be the canonical ϵ–balanced allocation on $T_{e\to i}$ with respect to b_x. As was discussed above, $\rho^\epsilon_{T_{e\to i}}(x) = x + \partial\theta_\epsilon^{b_x}(i)$.

Now we attempt to find some recursive expressions for such response functions. Before that, we need the following general lemma, whose proof is postponed to the end of this section.

Lemma 6.12. *Given $C > 0$ and a collection of nondecreasing functions $g_i : [0,1] \to \mathbb{R}$, $1 \le i \le n$, the set of fixed point equations*

$$\theta_i = \frac{e^{-g_i(\theta_i)}}{\sum_{j=1}^n e^{-g_j(\theta_j)} + C} \qquad 1 \le i \le n, \tag{6.58}$$

has a unique solution $(\theta_1, \dots, \theta_n) \in [0,1]^n$.

Proposition 6.14. *Assume T is a locally finite hypertree (not necessarily bounded), $\epsilon > 0$, and $(e,i) \in \Psi(T)$. Then, $\rho^\epsilon_{T_{e\to i}}(.)$ is an invertible function, and for $t \in \mathbb{R}$, we have*

$$\left(\rho^\epsilon_{T_{e\to i}}\right)^{-1}(t) = t - \sum_{e'\ni i: e'\neq e}\left(1 - \sum_{j\in e', j\neq i}\zeta^\epsilon_{e',j}\right), \tag{6.59}$$

where $\{\zeta^\epsilon_{e',j}\}$ for $e' \ni i$, $e' \neq e$, $j \in e'$, $j \neq i$ are the unique solutions to the set of equations

$$\zeta^\epsilon_{e',j} = \frac{\exp\left(-\frac{\rho^\epsilon_{T_{e'\to j}}(\zeta^\epsilon_{e',j})}{\epsilon}\right)}{e^{-t/\epsilon} + \sum_{l\in e', l\neq i}\exp\left(-\frac{\rho^\epsilon_{T_{e'\to l}}(\zeta^\epsilon_{e',l})}{\epsilon}\right)}. \tag{6.60}$$

Proof. Note that Lemma 6.11 implies that the set $A(t) := \{x \in \mathbb{R} : \rho^\epsilon_{T_{e\to i}}(x) = t\}$ is not empty. Hence, in order to show that $\rho^\epsilon_{T_{e\to i}}(.)$ is invertible, we should show that $A(t)$ is a singleton for all $t \in \mathbb{R}$. By definition, $x \in A(t)$ means that we have

$$t = x + \sum_{e'\ni i, e'\neq e}\left(1 - \sum_{j\in e', j\neq i}\theta_\epsilon^{b_x}(e',j)\right). \tag{6.61}$$

For each (e',j) in the above summation, as $T_{e'\to j}$ is a subtree of $T_{e\to i}$, we can treat $\theta_\epsilon^{b_x}$ as an allocation on $T_{e'\to j}$ via the identity projection. With this, Proposition 6.8 implies that $\theta_\epsilon^{b_x}$ is the canonical ϵ–balanced allocation on $T_{e'\to j}$ with respect to the baseload function that evaluates to $\theta^\epsilon(e', j)$ at j and zero elsewhere. Thereby, by the definition of the response function, for each such pair (e', j), we have $\partial\theta^{b_x}_\epsilon(j) = \rho^\epsilon_{T_{e'\to j}}(\theta^\epsilon(e', j))$. Consequently,

$$\theta^{b_x}_\epsilon(e', j) = \frac{\exp\left(-\frac{\rho^\epsilon_{T_{e'\to j}}(\theta^{b_x}_\epsilon(e',j))}{\epsilon}\right)}{e^{-t/\epsilon} + \sum_{l\in e', l\neq i} \exp\left(-\frac{\rho^\epsilon_{T_{e'\to l}}(\theta^{b_x}_\epsilon(e',l))}{\epsilon}\right)}.$$

Lemma 6.12 guarantees that for each $e' \ni i$, $e' \neq e$, there is a unique set of solutions to these equations. From (6.61), we realize that for any $t \in \mathbb{R}$, there is a unique solution for x. This implies the invertibility of $\rho^\epsilon_{T_{e\to i}}$. Rearranging (6.61), we get (6.59), with $\zeta_{e',j} = \theta^{b_x}_\epsilon(e', j)$. □

Now, we will send ϵ to zero in the above Proposition and show that, under some conditions, the sequence of response functions converges pointwise to a limit which satisfies a certain fixed point equation. To do so, we need the following lemma, whose proof is given at the end of this section.

Lemma 6.13. *Assume $\{g_n\}_{n=1}^\infty$ is a sequence of functions in $\Upsilon(C)$ that converge pointwise to a function g, i.e. $g(x) = \lim_{n\to\infty} g_n(x)$ for all $x \in \mathbb{R}$. Then, g is also in $\Upsilon(C)$. Furthermore, if D_n denotes the set of discontinuities of g_n and D denotes the set of discontinuities of g, then for $t \notin (\bigcup_n D_n) \cup D$ we have*

$$\lim_{n\to\infty} g_n^{-1}(t) = g^{-1}(t). \tag{6.62}$$

Now, we can write recursive equations for the limit of $\rho^\epsilon_{T_{e\to i}}$, if it exists.

Proposition 6.15. *Assume T is a locally finite hypertree (not necessarily bounded) and ϵ_n is a sequence of positive numbers converging to zero, with θ_{ϵ_n} being the canonical ϵ_n–balanced allocation on T. If $l_i := \lim_{n\to\infty} \partial\theta_{\epsilon_n}(i)$ exists for all $i \in V(T)$, then, for all $(e, i) \in \Psi(T)$, $\rho^{\epsilon_n}_{T_{e\to i}}(.)$ converges pointwise to some $\rho_{T_{e\to i}}(.) \in \Upsilon(\deg_T(i) - 1)$. Moreover, for all $t \in \mathbb{R}$, we have*

$$\rho^{-1}_{T_{e\to i}}(t) = t - \sum_{e'\ni i: e'\neq e} \left[1 - \sum_{j\in e', j\neq i} \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+\right]_0^1, \tag{6.63}$$

where the inverse functions are defined as in (6.57). Furthermore, for a node $i \in V(T)$ we have

$$l_i > t \qquad \Longleftrightarrow \qquad \sum_{e\ni i} \left[1 - \sum_{j\in e, j\neq i} \left(\rho^{-1}_{T_{e\to j}}(t)\right)^+\right]_0^1 > t. \tag{6.64}$$

Proof. First we fix $x \in \mathbb{R}$ and $(e, i) \in \Psi(T)$ and show that, as $n \to \infty$, $\rho^{\epsilon_n}_{T_{e\to i}}(x)$ is convergent. We call the limit $\rho_{T_{e\to i}}(x)$. Let $\theta^{b_x}_{\epsilon_n}$ denote the canonical ϵ_n–balanced allocation on $T_{e\to i}$ with baseload b_x, which equals x at i and is zero elsewhere. By definition, $\rho^{\epsilon_n}_{T_{e\to i}}(x) = x + \partial\theta^{b_x}_{\epsilon_n}(i)$. Thus, it suffices to show that $\partial\theta^{b_x}_{\epsilon_n}(i)$ is convergent. Note that $\theta^{b_x}_{\epsilon_n}$ is a sequence in the compact space $[0,1]^{\Psi(T_{e\to i})}$ equipped with the product topology. On the other hand, $\partial\theta^{b_x}_{\epsilon_n}(i)$ is a bounded sequence depending only on a finite number of coordinates, namely $\deg_T(i)-1$. As a result, in order to show that $\partial\theta^{b_x}_{\epsilon_n}(i)$ is convergent, it suffices to show that if θ_1 and θ_2 are two subsequential limits of $\theta^{b_x}_{\epsilon_n}$ in $[0,1]^{\Psi(T_{e\to i})}$, then $\partial\theta_1(i)=\partial\theta_2(i)$. Passing to the limit in (6.2), we realize that both θ_1 and θ_2 are balanced allocations on $T_{e\to i}$ with respect to the baseload b_x. Using Proposition 6.2, it suffices to show that $\|\partial\theta_1-\partial\theta_2\|_{l^1(V(T_{e\to i}))}<\infty$. From Proposition 6.8, we know that the restriction of θ_{ϵ_n} to $T_{e\to i}$ is the canonical ϵ_n–balanced allocation with baseload $\theta_{\epsilon_n}(e,i)$ at i and zero elsewhere. Hence, if K is a finite subset of $V(T_{e\to i})\setminus\{i\}$, using Proposition 6.7, we have

$$\sum_{j\in K}|\partial\theta^{b_x}_{\epsilon_n}(j)-\partial\theta_{\epsilon_n}(j)|\le |x|+\theta_{\epsilon_n}(e,i)\le |x|+1.$$

Using the triangle inequality, for integers n and m,

$$\sum_{j\in K}|\partial\theta^{b_x}_{\epsilon_n}(j)-\partial\theta^{b_x}_{\epsilon_m}(j)|\le 2(|x|+1)+\sum_{j\in K}|\partial\theta_{\epsilon_n}(j)-\partial\theta_{\epsilon_m}(j)|.$$

Now, send n to infinity along the subsequence of $\theta^{b_x}_{\epsilon_n}$ that converges to θ_1. Likewise, send m to infinity along the subsequence of $\theta^{b_x}_{\epsilon_n}$ that converges to θ_2. Using the assumption that $\partial\theta_{\epsilon_n}(j)$ is convergent, and since K is finite, we get

$$\sum_{j\in K}|\partial\theta_1(j)-\partial\theta_2(j)|\le 2(|x|+1).$$

Since K is arbitrary, sending K to $V(T_{e\to i})\setminus\{i\}$ we get

$$\begin{aligned}\sum_{j\in V(T_{e\to i})}|\partial\theta_1(j)-\partial\theta_2(j)| &= |\partial\theta_1(i)-\partial\theta_2(i)|+\\ &\qquad\sum_{j\in V(T_{e\to i})\setminus\{i\}}|\partial\theta_1(j)-\partial\theta_2(j)|\\ &\le 2(|x|+1)+2\deg_T(i)<\infty,\end{aligned}$$

which means $\|\partial\theta_1-\partial\theta_2\|_{l^1(V(T_{e\to i}))}<\infty$. This means that for all $x\in\mathbb{R}$ and $(e,i)\in\Psi(T)$, $\rho^{\epsilon_n}_{T_{e\to i}}(x)\to\rho_{T_{e\to i}}(x)$. As $\rho^{\epsilon_n}_{T_{e\to i}}(.)\in\Upsilon(\deg_T(i)-1)$, this in particular implies that $\rho_{T_{e\to i}}(.)\in\Upsilon(\deg_T(i)-1)$.

Now, we prove (6.63). Using part (iv) of Lemma 6.11, both sides of (6.63) are right continuous functions of t. Hence, if D denotes the union of the discontinuity sets of $\rho^{-1}_{T_{e\to i}}$ and $\rho^{-1}_{T_{e'\to j}}$ for $e'\ni i$ and $j\in e'$, D is countable and hence D^c is dense in $\mathbb{R}$. Thus, due to the right continuity, it suffices to show (6.63) for $t\notin D$.

Now, take some $t\in D^c$. Using Lemma 6.13 we have

$$\rho^{-1}_{T_{e\to i}}(t)=\lim_{n\to\infty}\left(\rho^{\epsilon_n}_{T_{e\to i}}\right)^{-1}(t),$$

for t in a dense subset of D^c. By abuse of notation, we continue to denote this subset by D^c. Using Proposition 6.14, $(\rho^{\epsilon_n}_{T_{e\to i}})^{-1}$ is expressed in terms of $\zeta^{\epsilon_n}_{e',j}$ which are solutions to (6.60). Comparing this to (6.63), it suffices to prove that for each $e' \ni i, e' \neq e$, we have

$$\lim_{n\to\infty} 1 - \sum_{j\in e', j\neq i} \zeta^{\epsilon_n}_{e',j} = \left[1 - \sum_{j\in e', j\neq i} \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+\right]_0^1. \tag{6.65}$$

Fixing such an e', since for each $j \in e'$, $j \neq i$, the sequence $\{\zeta^{\epsilon_n}_{e',j}\}_{n=1}^\infty$ is in the compact set $[0, 1]$, it suffices to show that if for all $j \in e'$, $j \neq i$, there is a subsequence $\zeta^{\epsilon_{n_k}}_{e',j}$ converging to some $\zeta^*_{e',j}$, then

$$1 - \sum_{j\in e', j\neq i} \zeta^*_{e',j} = \left[1 - \sum_{j\in e', j\neq i} \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+\right]_0^1. \tag{6.66}$$

Without loss of generality and in order to simplify the notation, we may assume that the subsequence is the whole sequence, i.e. for all $j \in e'$, $j \neq i$, $\zeta^{\epsilon_n}_{e',j} \to \zeta^*_{e',j}$. We show (6.66) in different cases:

Case I: $\sum_{j\in e',j\neq i} \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+ \leq 1$: since $\sum_{j\in e',j\neq i} \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+$ is nonnegative, it suffices to show that $\zeta^*_{e',j} = \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+$ for each $j \in e'$, $j \neq i$. For such a j, we do this in two subcases:

Case Ia: First, assume $\rho^{-1}_{T_{e'\to j}}(t) \leq 0$, in which case we should show $\zeta^*_{e',j} = 0$. If this is not the case, as $\zeta^*_{e',j} \in [0, 1]$, there must be the case that $\zeta^*_{e',j} > 0 \geq \rho^{-1}_{T_{e'\to j}}(t)$. Then, part ($v$) of Lemma 6.11 implies that $\rho_{T_{e'\to j}}(\zeta^*_{e',j}) \geq t + \delta$ for some $\delta > 0$. With this,

$$\begin{aligned}
\rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^{\epsilon_n}_{e',j}) - t &= \rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^{\epsilon_n}_{e',j}) - \rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^*_{e',j}) \\
&\quad + \rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^*_{e',j}) - \rho_{T_{e'\to j}}(\zeta^*_{e',j}) \\
&\quad + \rho_{T_{e'\to j}}(\zeta^*_{e',j}) - t \\
&\geq -|\zeta^{\epsilon_n}_{e',j} - \zeta^*_{e',j}| \\
&\quad + \rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^*_{e',j}) - \rho_{T_{e'\to j}}(\zeta^*_{e',j}) \\
&\quad + \rho_{T_{e'\to j}}(\zeta^*_{e',j}) - t.
\end{aligned}$$

Note that the first two terms converge to zero as $n \to \infty$ and hence they could be made smaller than $\delta/3$ by choosing n large enough. Thus $\rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^{\epsilon_n}_{e',j}) - t \geq \delta/3$ for n large enough. On the other hand

$$\begin{aligned}
\zeta^{\epsilon_n}_{e',j} &\leq \frac{1}{1 + \exp\left(-\frac{t - \rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^{\epsilon_n}_{e',j})}{\epsilon_n}\right)} \\
&\leq \frac{1}{1 + e^{\delta/(3\epsilon_n)}}.
\end{aligned}$$

Sending n to infinity, since δ is fixed, we realize that $\zeta^{\epsilon_n}_{e',j}$ converges to zero, which is a contradiction with the assumption that $\zeta^*_{e',j} > 0$.

Case Ib: Now consider the case $\rho^{-1}_{T_{e'\to j}}(t) > 0$. If $\zeta^*_{e',j} > \rho^{-1}_{T_{e'\to j}}(t)$, following a similar argument as in case Ia, we can conclude that $\zeta^*_{e',j} = 0$, which is a contradiction. Hence, assume $\zeta^*_{e',j} < \rho^{-1}_{T_{e'\to j}}(t)$. Since $t \in D^c$ is a continuity point of $\rho^{-1}_{T_{e'\to j}}$, using Lemma 6.11 part (v), we have $\rho_{T_{e'\to j}}(\zeta^*_{e',j}) \leq t - \delta$ for some $\delta > 0$. An argument similar to that in case Ia implies $\rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^{\epsilon_n}_{e',j}) \leq t - \delta/3$ for n large enough. Now

$$
\begin{aligned}
1 - \sum_{l\in e', l\neq i} \zeta^*_{e',l} &= \lim_{n\to\infty} 1 - \sum_{l\in e', l\neq i} \zeta^{\epsilon_n}_{e',l} \\
&= \lim_{n\to\infty} \frac{1}{1 + \sum_{l\in e', l\neq i} \exp\left(-\frac{\rho^{\epsilon_n}_{T_{e'\to l}}(\zeta^{\epsilon_n}_{e',l}) - t}{\epsilon_n}\right)} \\
&\leq \lim_{n\to\infty} \frac{1}{1 + \exp\left(-\frac{\rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^{\epsilon_n}_{e',j}) - t}{\epsilon_n}\right)} \\
&\leq \frac{1}{1 + e^{\delta/(3\epsilon_n)}}.
\end{aligned}
$$

Since δ is fixed, sending n to infinity we realize that $\sum_{l\in e', l\neq i} \zeta^*_{e',l} = 1$. This, together with our earlier assumption of Case I, means that

$$
\sum_{l\in e', l\neq i} \zeta^*_{e',l} = 1 \geq \sum_{l\in e', l\neq i} \left(\rho^{-1}_{T_{e'\to l}}(t)\right)^+ = \sum_{l\in e', l\neq i, \rho^{-1}_{T_{e'\to l}}(t) > 0} \rho^{-1}_{T_{e'\to l}}(t).
$$

Since we have assumed that $\zeta^*_{e',j} < \rho^{-1}_{T_{e'\to j}}(t)$, this means there exists some $j' \in e'$, $j' \neq i$ such that $\rho^{-1}_{T_{e'\to j'}}(t) > 0$ and $\zeta^*_{e',j'} > \rho^{-1}_{T_{e'\to j}}(t) > 0$. This, as we discussed above, results in $\zeta^*_{e',j'} = 0$, which is a contradiction. Hence, $\zeta^*_{e',j}$ should be equal to $\rho^{-1}_{T_{e'\to j}}(t)$ and the proof of this case is complete.

Case II: $\sum_{j\in e', j\neq i} \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+ > 1$, in which case we need to show that $1 - \sum_{j\in e', j\neq i} \zeta^*_{e',j} = 0$ to conclude (6.66). Since

$$
\begin{aligned}
\sum_{j\in e', j\neq i} \zeta^*_{e',j} &= \lim_{n\to\infty} \sum_{j\in e', j\neq i} \zeta^{\epsilon_n}_{e',j} \\
&= \lim_{n\to\infty} \frac{\sum_{l\in e', l\neq i} \exp(-\rho^{\epsilon_n}_{T_{e'\to l}}(\zeta^{\epsilon_n}_{e',l})/\epsilon)}{e^{-t/\epsilon} + \sum_{l\in e', l\neq i} \exp(-\rho^{\epsilon_n}_{T_{e'\to l}}(\zeta^{\epsilon_n}_{e',l})/\epsilon)} \\
&\leq 1 < \sum_{j\in e', j\neq i} \left(\rho^{-1}_{T_{e'\to j}}(t)\right)^+,
\end{aligned}
$$

there should exist some $j^* \in e', j^* \neq i$ such that $\zeta^*_{e',j^*} < \rho^{-1}_{T_{e'\to j^*}}(t)$ and $\rho^{-1}_{T_{e'\to j^*}} > 0$. Since $t \notin D$, using Lemma 6.11 we have $\rho_{T_{e'\to j^*}}(\zeta^*_{e',j^*}) \leq t - \delta$ for some $\delta > 0$. Using calculations similar to those in case Ia above, for n large enough we have $\rho^{\epsilon_n}_{T_{e'\to j^*}}(\zeta^{\epsilon_n}_{e',j^*}) \leq t - \delta/3$. Now,

$$\begin{aligned} 1 - \sum_{j\in e', j\neq i} \zeta^{\epsilon_n}_{e',j} &= \frac{1}{1 + \sum_{j\in e', j\neq i} \exp\left(-\frac{\rho^{\epsilon_n}_{T_{e'\to j}}(\zeta^{\epsilon_n}_{e',j}) - t}{\epsilon_n}\right)} \\ &\leq \frac{1}{1 + \exp\left(-\frac{\rho^{\epsilon_n}_{T_{e'\to j^*}}(\zeta^{\epsilon_n}_{e',j^*}) - t}{\epsilon_n}\right)} \\ &\leq \frac{1}{1 + e^{\delta/(3\epsilon_n)}}. \end{aligned}$$

Since δ is fixed, sending n to infinity we conclude that $1 - \sum_{j\in e', j\neq i} \zeta^*_{e',j} = 0$, which completes the argument of this case.

Having verified (6.66) in all cases, we conclude (6.63). Now, we prove (6.64). Note that in the above discussion we started with the rooted tree $T_{e\to i}$. However, it can be verified that all the arguments are valid if we start with the tree T rooted at an arbitrary vertex $i \in V(T)$, i.e. (T, i). Convergence of $\rho_{(T,i)}$ is also similar. In this case, repeating the above argument, (6.63) becomes

$$\rho^{-1}_{(T,i)}(t) = t - \sum_{e\ni i} \left[1 - \sum_{j\in e, j\neq i} \left(\rho^{-1}_{T_{e\to j}}(t)\right)^+\right]_0^1. \tag{6.67}$$

On the other hand, note that

$$l_i = \lim_{n\to\infty} \partial\theta_n(i) = \lim_{n\to\infty} \rho^{\epsilon_n}_{(T,i)}(0) = \rho_{(T,i)}(0).$$

Hence, $l_i > t$ iff $\rho_{(T,i)}(0) > t$, or equivalently, using part (v) of Lemma 6.11, $\rho^{-1}_{(T,i)}(t) < 0$. Substituting into (6.67), we conclude (6.64), and the proof is complete. □

Proof of Lemma 6.12. Assume $(\theta_1, \ldots, \theta_n)$ and $(\theta'_1, \ldots, \theta'_n)$ are two distinct solutions to this set of equations. We claim that it cannot be the case that $\theta'_i > \theta_i$ for some i and $\theta'_j \leq \theta_j$ for some other $j \neq i$. Assume this holds. Since the right hand side of (6.58) is positive, all θ_i's and θ'_i's are positive. On the other hand, we have

$$\frac{e^{-g_i(\theta'_i)}}{e^{-g_j(\theta'_j)}} = \frac{\theta'_i}{\theta'_j} > \frac{\theta_i}{\theta_j} = \frac{e^{-g_i(\theta_i)}}{e^{-g_j(\theta_j)}},$$

which means

$$g_i(\theta'_i) + g_j(\theta_j) < g_j(\theta'_j) + g_i(\theta_i).$$

But $\theta'_i > \theta_i$ implies $g_i(\theta'_i) \geq g_i(\theta_i)$ since g_i is nondecreasing. On the other hand, $\theta'_j \leq \theta_j$ implies $g_j(\theta'_j) \geq g_j(\theta_j)$ which is a contradiction with the above inequality.

Hence, without loss of generality, we may assume that $\theta'_i \geq \theta_i$ for all $1 \leq i \leq n$. If it is not the case that $\theta_i = \theta'_i$ for all $1 \leq i \leq n$, then

$$\frac{\sum_{i=1}^n e^{-g_i(\theta_i)}}{\sum_{i=1}^n e^{-g_i(\theta_i)} + C} = \sum \theta_i < \sum \theta'_i = \frac{\sum_{i=1}^n e^{-g_i(\theta'_i)}}{\sum_{i=1}^n e^{-g_i(\theta'_i)} + C}.$$

On the other hand, $\theta'_i \geq \theta_i$ and g_i being nondecreasing implies that $e^{-g_i(\theta_i)} \geq e^{-g_i(\theta'_i)}$ for all $1 \leq i \leq n$, which is in contradiction with the above inequality. □

Proof of Lemma 6.13. Sending n to infinity in the three conditions of Definition 6.33 and using the fact that g_n converges pointwise to g implies that g is in $\Upsilon(C)$. To show (6.62), given $t \notin (\bigcup_n D_n) \cup D$, define $x_n := g_n^{-1}(t)$ and $x := g^{-1}(t)$. Since $g_n \in \Upsilon(C)$, we have $x_n \leq g_n(x_n) \leq x_n + C$ or $x_n \in [t - C, t]$, which is a compact set. Hence, it suffices to show that any subsequential limit of x_n is equal to x. Thus, without loss of generality, we may assume that $x_n \to x'$, and we show that $x' = x$.

If $x' < x$, since t is a continuity point for g, Lemma 6.11 part (iv) implies that $g(x') < t$. Thereby, $g(x') \leq t - \delta$ for some $\delta > 0$. Now,

$$\begin{aligned} g_n(x_n) &= g_n(x_n) - g_n(x') + g_n(x') - g(x') + g(x') \\ &\leq |x_n - x'| + |g_n(x') - g(x')| + g(x'), \end{aligned}$$

where the last inequality employs the fact that $g_n \in \Upsilon(C)$. Since $x_n \to x'$ and g_n converges pointwise to g, for large n the first two terms could be made smaller than $\delta/3$. Thus, for large n, $g_n(x_n) \leq t - \delta/3$, which is a contradiction with $g_n(x_n) = t$. The assumption $x' > x$ similarly results in contradiction. As a result, $x' = x$, and the proof is complete. □

6.8 Characterization of the Mean Excess Function for Galton Watson Processes

In this section, we prove Theorem 6.2. This is done in two steps. First, in Section 6.8.1, we prove that for any set of fixed points $\{Q_l\}_{l \geq 2}$, the LHS of (6.11) is no less than the RHS. This is proved in Proposition 6.16. Later, in Section 6.8.2, we show that there exists a set of fixed points achieving the maximum in (6.11). This is done in Proposition 6.17.

6.8.1 Lower Bound

In this section, we use the indexing notation $\mathbb{N}_{\text{vertex}}$ and $\mathbb{N}_{\text{edge}}$, which was introduced in Section 6.2.9. The level of a vertex $(s_1, e_1, i_1, \dots, s_k, e_k, i_k) \in \mathbb{N}_{\text{vertex}}$ is defined to be k, and the level of $\emptyset$ is defined to be zero. Likewise, the level of an edge $(s_1, e_1, i_1, \dots, s_k, e_k) \in \mathbb{N}_{\text{edge}}$ is defined to be k. Also, recall that for $v \in \mathbb{N}_{\text{vertex}}$, $s \geq 2$ and $e \geq 1$, (v, s, e) is an element in $\mathbb{N}_{\text{edge}}$ obtained by concatenating (s, e) to the end of the string representing v in $\mathbb{N}_{\text{vertex}}$. Likewise, for $s \geq 2$, $e \geq 1$ and $1 \leq i \leq s - 1$, $(v, s, e, i) \in \mathbb{N}_{\text{vertex}}$ is defined similarly.

We need the following tool before proving our lower bound. In the following lemma, $\mathsf{Proj}_{\bar{\mathcal{H}}_*(\mathbb{R})\to\mathcal{H}_*} : \bar{\mathcal{H}}_*(\mathbb{R}) \to \mathcal{H}_*$ is the projection defined in (6.7).

Lemma 6.14. *Let $t \in \mathbb{R}$ together with distributions P and $\{Q_k\}_{k\geq 2} \in \mathcal{Q}$ be given as in Theorem 6.2. Given a probability distribution W on Λ, there is a random marked rooted tree $(\bar{\mathbb{T}}_W, \emptyset)$, with marks taking values in $\mathbb{R}$, and with vertex set and edge set $\mathbb{N}_{vertex}$ and $\mathbb{N}_{edge}$, respectively, such that the underlying unmarked rooted tree is a Galton Watson tree such that the type of the root is distributed according to W, and the type of a non–root vertex $v = (s_1, e_1, i_1, \dots, s_r, e_r, i_r)$ in the subtree below v is distributed according to $\hat{P}_{s_r}$. Moreover, the marks of $\bar{\mathbb{T}}_W$ satisfy*

$$\xi_{\bar{\mathbb{T}}_W}(e,i) = t - \sum_{e' \ni i, e' \neq e} \left[1 - \sum_{j \in e', j \neq i} \xi_{\bar{\mathbb{T}}_W}(e',j)^+\right]_0^1, \tag{6.68}$$

for all $(e,i) \in \Psi(\bar{\mathbb{T}}_W)$. Furthermore, for any $L \geq 1$, conditioned on the structure of the tree up to depth L, the set of marks from edges in level L towards vertices in that level are independent, and for any edge–vertex pair (e,i) both in level L, $\xi_{\bar{\mathbb{T}}_W}(e,i)$ is distributed according to Q_k where k is the size of e.

In particular, when $W = P$, the measure $\nu \in \mathcal{P}(\bar{\mathcal{H}}_*(\mathbb{R}))$, which is defined to be the law of $[\bar{\mathbb{T}}_P, \emptyset]$, is unimodular and $(\mathsf{Proj}_{\bar{\mathcal{H}}_*(\mathbb{R})\to\mathcal{H}_*})_*\nu = \mathsf{UGWHT}(P)$.

Proof. We first generate the collection of random variables $\Gamma_\emptyset, (\Gamma_v, X_v)_{v\in\mathbb{N}_{\text{vertex}}\setminus\{\emptyset\}}$, such that Γ_v for $v \in \mathbb{N}_{\text{vertex}}$ takes value in Λ and X_v for $v \in \mathbb{N}_{\text{vertex}} \setminus \{\emptyset\}$ takes values in $\mathbb{R}$, with the following properties: (i) $(\Gamma_v)_{v\in\mathbb{N}_{\text{vertex}}}$ are independent from each other such that $\Gamma_\emptyset$ has law W and for $v = (s_1, e_1, i_1, \dots, s_r, e_r, i_r) \in \mathbb{N}_{\text{vertex}}$, Γ_v has law $\hat{P}_{s_r}$; (ii) For any $v \in \mathbb{N}_{\text{vertex}} \setminus \{\emptyset\}$, we have

$$X_v = t - \sum_{k=2}^{h(\Gamma_v)} \sum_{l=1}^{\Gamma_v(k)} \left[1 - X^+_{(v,k,l,1)} - \cdots - X^+_{(v,k,l,k-1)}\right]_0^1; \tag{6.69}$$

(iii) For any $L \geq 1$, X_v for nodes v at level L are independent and for $v = (s_1, e_1, i_1, \dots, s_L, e_L, i_L)$ at level L, X_v is distributed according to Q_{s_L}.

We construct the law of the above random variables satisfying the above conditions using Kolmogorov's extension theorem (see, for instance, [Tao11]). For an integer $L \geq 1$, define A_L to be the set of nodes in $\mathbb{N}_{\text{vertex}}$ with level at most L. For each $L \geq 1$, we introduce the law of a subset of the above family of random variables, namely $\Gamma_\emptyset, (X_v, \Gamma_v)_{v\in A_L\setminus\{\emptyset\}}$, and denote this law by ν_L. To start with, we generate Γ_v, $v \in A_L$ independently such that $\Gamma_\emptyset$ has law W and Γ_v for $v = (s_1, e_1, i_1, \dots, s_r, e_r, i_r)$, X_v has law $\hat{P}_{s_r}$. In the next step, we generate X_v for nodes v with depth equal to L independently such that X_v for $v = (s_1, e_1, i_1, \dots, s_L, e_L, i_L)$ has law Q_{s_L}. Next, we define X_v for nodes at levels 1 through $L-1$ using the relation (6.69) inductively starting from level $L-1$ all the way up to level 1. Using the fact that $Q_k = F^{(k)}_{P,t}(\{Q_l\}_{l\geq 2})$, see (6.9), and also that Γ_v for $v = (s_1, e_1, i_1, \dots, s_r, e_r, i_r)$ has law $\hat{P}_{s_r}$, it is evident that the set of measures $\{\nu_L\}_{L\geq 1}$ are consistent. Therefore, Kolmogorov's extension theorem implies that the set of random variables with the conditions stated above exist.

Now, we turn these random variables into a marked random rooted tree $\bar{\mathbb{T}}_W$ having vertex set and edge set $\mathbb{N}_{\text{vertex}}$ and $\mathbb{N}_{\text{edge}}$, respectively. To do so, we first construct the underlying unmarked tree $\bar{\mathbb{T}}_W$ given the types Γ_v, $v \in \mathbb{N}_{\text{vertex}}$. In the next step, for any edge e and vertex v being at the same level in the tree, we set $\xi_{\bar{\mathbb{T}}_W}(e, v) := X_v$. It can be easily seen that the "upward" marks, i.e. marks from edges towards nodes above them, are immediately unambiguously defined. To see this, for an edge e at level 1, we define $\xi_{\bar{\mathbb{T}}_W}(e, \emptyset)$ using (6.68). We then inductively go down one level at a time, to define all the other upward marks.

Now, we show that the measure ν, which is defined to be the law of $[\bar{\mathbb{T}}_P, \emptyset]$, is unimodular. Our proof technique is similar to that of Lemma E.2 in Appendix E.6. Take a Borel function $f : \bar{\mathcal{H}}_{**}(\mathbb{R}) \to [0, \infty)$ and note that due to our above construction,

$$\int f d\vec{\nu} = \int \partial f d\nu = \mathbb{E}\left[\sum_{k=2}^{h(\Gamma_\emptyset)} \sum_{l=1}^{\Gamma_\emptyset(k)} f(\bar{\mathbb{T}}_P, (k, l), \emptyset)\right].$$

Using the symmetry in the construction, we have

$$\int f d\vec{\nu} = \sum_{\gamma \in \Lambda} P(\gamma) \sum_{k=2}^{h(\gamma)} \gamma(k) \mathbb{E}\left[f(\bar{\mathbb{T}}_P, (k, 1), \emptyset) | \Gamma_\emptyset = \gamma\right].$$

As all the terms are nonnegative, we may change the order of summation. Also using the definition of $\hat{P}$, if Γ is a random variable with law P, we get

$$\int f d\vec{\nu} = \sum_{k=2}^{\infty} \mathbb{E}\left[\Gamma(k)\right] \sum_{\gamma \in \Lambda} \hat{P}_k(\gamma) \mathbb{E}\left[f(\bar{\mathbb{T}}_P, (k, 1), \emptyset) | \Gamma_\emptyset = \gamma + \mathbb{e}_k\right].$$

Now, for each $k \geq 2$, define $\widetilde{P}_k$ to be the law of the random variable $\Gamma_k + \mathbb{e}_k$ where Γ_k has law $\hat{P}_k$. With this, for each $k \geq 2$, the inner summation over γ in the above expression could be interpreted as an expectation with respect to a tree with root type distribution $\widetilde{P}_k$, i.e. $\bar{\mathbb{T}}_{\widetilde{P}_k}$. In fact, this shows that

$$\int f d\vec{\nu} = \sum_{k=2}^{\infty} \mathbb{E}\left[\Gamma(k)\right] \mathbb{E}\left[f(\bar{\mathbb{T}}_{\widetilde{P}_k}, (k, 1), \emptyset)\right]. \tag{6.70}$$

Now, note that for each $k \geq 2$, due to the definition of $\widetilde{P}_k$, the tree $(\bar{\mathbb{T}}_{\widetilde{P}_k})_{(k,1) \to \emptyset}$ rooted at $\emptyset$ (which we recall is obtained by removing $(k, 1)$ and then taking the subtree rooted at $\emptyset$) has an underlying unmarked structure which is precisely $\text{GWT}_k(P)$. This, in particular, implies that $\xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k, 1), \emptyset)$ has law Q_k. Also, by construction, for $1 \leq i \leq k - 1$, $(\bar{\mathbb{T}}_{\widetilde{P}_k})_{(k,1) \to (k,1,i)}$ have independent underlying unmarked structures, all with law $\text{GWT}_k(P)$. Moreover, the downward marks in $(\bar{\mathbb{T}}_{\widetilde{P}_k})_{(k,1)\to\emptyset}$ and $(\bar{\mathbb{T}}_{\widetilde{P}_k})_{(k,1)\to(k,1,i)}$ for $1 \le i \le k-1$ are independent from each other and have the same distribution. Thereby, $\xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1),v)$ for $v \in (k,1)$ are independent and all have distribution Q_k. Now, we claim that for all $v \in (k,1)$, $(\bar{\mathbb{T}}_{\widetilde{P}_k})_{(k,1)\to v}$ are equal in distribution. To see this, note that for all $v \in (k,1)$, the unmarked structure of $(\bar{\mathbb{T}}_{\widetilde{P}_k})_{(k,1)\to v}$ is $\mathrm{GWT}_k(P)$ and the downward marks are constructed following the same recipes. As was discussed above, to construct upward marks, we start from the root and go down inductively. The fact that $\xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1),v)$ for $v \in (k,1)$ are i.i.d. with law Q_k guarantees that the downward marks in $(\bar{\mathbb{T}}_{\widetilde{P}_k})_{(k,1)\to v}$ are equally distributed for all $v \in (k,1)$. This in particular implies that for $1 \le i \le k-1$,

$$\mathbb{E}\left[f(\bar{\mathbb{T}}_{\widetilde{P}_k},(k,1),\emptyset)\right] = \mathbb{E}\left[f(\bar{\mathbb{T}}_{\widetilde{P}_k},(k,1),(k,1,i))\right].$$

Using this and writing (6.70) for $\int \nabla f d\mu$, we conclude that $\int f d\vec{\nu} = \int \nabla f d\vec{\nu}$ which completes the proof of the unimodularity of ν. □

Before proving our lower bound, we state a modified version of our variational representation in Proposition 6.11 and a general lemma. The proof of the following lemmas are given at the end of this section.

Lemma 6.15. *Assume μ is a distribution on $\mathcal{H}_*$ with $\deg(\mu) < \infty$ and ν is a unimodular distribution on $\bar{\mathcal{H}}_*(\mathbb{R})$ such that $\left(\mathsf{Proj}_{\bar{\mathcal{H}}_*(\mathbb{R})\to\mathcal{H}_*}\right)_*\nu = \mu$. Then, for any Borel allocation $\Theta : \mathcal{H}_{**} \to [0,1]$ and any function $f : \bar{\mathcal{H}}_*(\mathbb{R}) \to [0,1]$, we have*

$$\int (\partial\Theta - t)^+ d\mu \ge \int \tilde{f}_{min} d\vec{\nu} - t\int f d\nu,$$

where

$$\tilde{f}_{min}([\bar{H},e,i]) := \frac{1}{|e|}\min_{j\in e} f([\bar{H},j]).$$

Lemma 6.16. *Assume $x_1,\dots,x_n$ are real numbers. Then $x_i < \left[1-\sum_{j\neq i} x_j^+\right]_0^1$ for all $1 \le i \le n$ if and only if $\sum x_i^+ < 1$.*

Proposition 6.16. *Assume P is a distribution on Λ such that $\mathbb{E}\left[\|\Gamma\|_1\right] < \infty$ where Γ has law P. Then, with $\mu = \mathsf{UGWHT}(P)$, for any $t \in \mathbb{R}$, and any set of probability distributions on real numbers $\{Q_k\}_{k\ge 2}$ such that for all $k \ge 2$ we have $Q_k = F^{(k)}_{P,t}(\{Q_l\}_{l\ge 2})$, it holds that*

$$\Phi_\mu(t) \ge \left(\sum_{k=2}^{\infty}\frac{\mathbb{E}\left[\Gamma(k)\right]}{k}\mathbb{P}\left(\sum_{i=1}^{k} X_{k,i}^+ < 1\right)\right) - t\mathbb{P}\left(\sum_{k=2}^{h(\Gamma)}\sum_{i=1}^{\Gamma(k)} Y_{k,i} > t\right).$$

Here, in the first expression, Γ is a random variable on Λ with law P and $\{X_{k,i}\}_{k,i}$ are independent such that $X_{k,i}$ has law Q_k. Also, in the second expression, Γ has law P and $\{Y_{k,i}\}_{k,i}$ are independent from each other and from Γ such that $Y_{k,i}$ has the law of the random variable $[1-(Z_1^+ + \cdots + Z_{k-1}^+)]_0^1$ where Z_j are i.i.d. with law Q_k.

Proof. Note that the condition $\mathbb{E}\left[\|\Gamma\|_1\right] < \infty$ guarantees that $\deg(\mu) < \infty$. Define the functions $F : \bar{\mathcal{H}}_{**}(\mathbb{R}) \to [0,1]$ and $f : \bar{\mathcal{H}}_* \to [0,1]$ as

$$F([\bar{H}, e, i]) = \left[1 - \sum_{j \in e, j \neq i} \xi_{\bar{H}}(e,j)^+\right]_0^1,$$

and $f([\bar{H}, i]) := \mathbb{1}_{\partial F([\bar{H},i]) > t}$. Using the unimodular measure ν constructed in Lemma 6.14 and the variational characterization in Lemma 6.15, we have

$$\Phi_\mu(t) \geq \int \tilde{f}_{\min} d\vec{\nu} - t \int f d\nu, \tag{6.71}$$

where

$$\tilde{f}_{\min}(\bar{H}, e, j) = \frac{1}{|e|} \min_{j \in e} f(\bar{H}, j) = \frac{1}{|e|} \mathbb{1}\left[\partial F(\bar{H}, j) > t \ \forall j \in e\right].$$

Following the proof of Lemma 6.14, and in particular Equation (6.70) therein, we have

$$\int \tilde{f}_{\min} d\vec{\nu} = \sum_{k \geq 2} \mathbb{E}\left[\Gamma(k)\right] \mathbb{E}\left[\tilde{f}_{\min}(\bar{\mathbb{T}}_{\widetilde{P}_k}, (k,1), \emptyset)\right]. \tag{6.72}$$

Due to the definition of f, $\tilde{f}_{\min}(\bar{\mathbb{T}}_{\widetilde{P}_k}, (k,1), \emptyset) = \frac{1}{k}\mathbb{1}\left[\partial F(\bar{\mathbb{T}}_{\widetilde{P}_k}, v) > t \ \ \forall v \in (k,1)\right]$. For $v \in (k,1)$,

$$\begin{aligned}
\partial F(\bar{\mathbb{T}}_{\widetilde{P}_k}, v) &= \sum_{e' \ni v} F(\bar{\mathbb{T}}_{\widetilde{P}_k}, e', v) \\
&= \left[1 - \sum_{w \in (k,1), w \neq v} \xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1), w)^+\right]_0^1 + \\
&\qquad \sum_{e' \ni v, e' \neq (k,1)} \left[1 - \sum_{w \in e', w \neq v} \xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}(e', w)^+\right]_0^1.
\end{aligned}$$

Using (6.68), this yields

$$\partial F(\bar{\mathbb{T}}_{\widetilde{P}_k}, v) = \left[1 - \sum_{w \in (k,1), w \neq v} \xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1), w)^+\right]_0^1 + t - \xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1), v).$$

Therefore, $\partial F(\bar{\mathbb{T}}_{\widetilde{P}_k}, v) > t$ for all $v \in (k,1)$ if and only if for all $v \in (k,1)$,

$$\xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1), v) < \left[1 - \sum_{w \in (k,1), w \neq v} \xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1), w)^+\right]_0^1.$$

Using Lemma 6.16, this is equivalent to $\sum_{v\in(k,1)} \xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1),v)^+ < 1$. Therefore,

$$\mathbb{E}\left[\tilde{f}_{\min}(\bar{\mathbb{T}}_{\widetilde{P}_k},(k,1),\emptyset)\right] = \frac{1}{k}\mathbb{P}\left(\sum_{v\in(k,1)} \xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1),v)^+ < 1\right).$$

But as was shown in Lemma 6.14, $\xi_{\bar{\mathbb{T}}_{\widetilde{P}_k}}((k,1),v)$ for $v \in (k,1)$ are i.i.d. with law Q_k. Consequently, substituting in (6.72) we get

$$\int \tilde{f}_{\min} d\vec{\nu} = \sum_{k=2}^{\infty} \frac{\mathbb{E}\left[\Gamma(k)\right]}{k}\mathbb{P}\left(\sum_{i=1}^{k} X_{k,i}^+ < 1\right), \tag{6.73}$$

where $X_{k,i}$, $k \geq 2, 1 \leq i \leq k$ are independent such that $X_{k,i}$ has law Q_k.

On the other hand,

$$\begin{aligned}\int f d\nu = \mathbb{P}\left(\partial F(\bar{\mathbb{T}}_P,\emptyset) > t\right) &= \mathbb{P}\left(\sum_{k=2}^{h(\Gamma_\emptyset)}\sum_{i=1}^{\Gamma_\emptyset(k)} F(\bar{\mathbb{T}}_P,(k,i),\emptyset) > t\right)\\ &= \mathbb{P}\left(\sum_{k=2}^{h(\Gamma_\emptyset)}\sum_{i=1}^{\Gamma_\emptyset(k)} \left[1 - \sum_{j=1}^{k-1} \xi_{\bar{\mathbb{T}}_P}((k,i),(k,i,j))^+\right]_0^1 > t\right).\end{aligned}$$

But, as we saw in Lemma 6.14, $\xi_{\bar{\mathbb{T}}_P}((k,i),(k,i,j))$ for $k \geq 2$, $i \leq \Gamma_\emptyset(k)$, $1 \leq j \leq k-1$, are independent, and $\xi_{\bar{\mathbb{T}}_P}((k,i),(k,i,j))$ has law Q_k. Thereby,

$$\int f d\nu = \mathbb{P}\left(\sum_{k=2}^{h(\Gamma)}\sum_{i=1}^{\Gamma(k)} Y_{k,i} > t\right), \tag{6.74}$$

with $Y_{k,i}$ being independent from each other such that $Y_{k,i}$ has the law of the random variable $[1-(Z_1^+ + \cdots + Z_{k-1}^+)]_0^1$ with Z_j's being i.i.d. with law Q_k. The proof is complete by substituting (6.73) and (6.74) into (6.71). □

Proof of Lemma 6.15. By interpreting Θ as a function on $\mathcal{H}_{**}(\mathbb{R})$ via $\Theta(\bar{H},e,i) := \Theta(H,e,i)$ (where we recall that H is the underlying unmarked hypergraph associated to $\bar{H}$), and also using the inequality $x^+ \geq xy$ which holds for $y \in [0,1]$, we have

$$\int (\partial\Theta - t)^+ d\mu \int (\partial\Theta - t)^+ d\nu \geq \int f\partial\Theta d\nu - t\int f d\nu. \tag{6.75}$$

Since $\deg(\mu) < \infty$, all the integrals are finite and well defined. Due to the definition of $\vec{\nu}$ and the unimodularity of ν we have $\int f\partial\Theta d\nu = \int f\Theta d\vec{\nu} = \int \nabla(f\Theta) d\nu$. On the other hand,

$$\nabla(f\Theta)(\bar{H},e,i) = \frac{1}{|e|}\sum_{j\in e} f(\bar{H},j)\Theta(H,e,j) \geq \frac{1}{|e|}\min_{j\in e} f(\bar{H},j) = \tilde{f}_{\min}(\bar{H},i),$$

where the inequality holds since $\sum_{j\in e}\Theta(H,e,j) = 1$ and $\Theta([H,e,j]) \geq 0$ for all $j \in e$. Substituting this into (6.75) completes the proof. □

Proof of Lemma 6.16. First, assume that $x_i < \left[1 - \sum_{j\neq i} x_j^+\right]_0^1$ for all i. If $x_i \le 0$ for all i then nothing remains to be proved. Hence, assume that $x_i > 0$ for some i. Since $x_i < \left[1 - \sum_{j\neq i} x_j^+\right]_0^1$, we have $0 < \sum_{j\neq i} x_j^+ < 1$, which means that $1-\sum_{j\neq i} x_j^+ \in [0,1]$. Therefore, $\left[1 - \sum_{j\neq i} x_j^+\right]_0^1 = 1 - \sum_{j\neq i} x_j^+$. On the other hand,

$$x_i^+ = x_i < \left[1 - \sum_{j\neq i} x_j^+\right]_0^1 = 1 - \sum_{j\neq i} x_j^+,$$

which implies $\sum x_i^+ < 1$.

In order to prove the other direction, take $1 \le i \le n$ and note that

$$x_i \le x_i^+ = \sum_{k=1}^n x_k^+ - \sum_{j\neq i} x_j^+ < 1 - \sum_{j\neq i} x_j^+.$$

Moreover, $\sum_{j\neq i} x_j^+ \le \sum x_i^+ < 1$. Thereby, $1 - \sum_{j\neq i} x_j^+ \in [0,1]$ and $1 - \sum_{j\neq i} x_j^+ = \left[1 - \sum_{j\neq i} x_j^+\right]_0^1$. Substituting this into the above inequality completes the proof. □

6.8.2 Upper Bound

In this section, we show that there exists a family of probability distributions $\{Q_k\}_{k\geq 2}$ satisfying the fixed points equations (6.10) and achieving the maximum on the RHS of (6.11).

Proposition 6.17. *Assume P is a distribution on Λ such that $\mathbb{E}\left[\|\Gamma\|_1\right] < \infty$, where Γ has law P. Let $\mu = \mathsf{UGWHT}(P)$. Given $t \in \mathbb{R}$, there exists a family of probability distributions $\{Q_l\}_{l\geq 2}$ on the set of real numbers such that, for each $k \geq 2$, $Q_k = F^{(k)}_{P,t}(\{Q_l\}_{l\geq 2})$, and such that we have*

$$\Phi_\mu(t) = \left(\sum_{k=2}^\infty \frac{\mathbb{E}\left[\Gamma(k)\right]}{k} \mathbb{P}\left(\sum_{i=1}^k X^+_{k,i} < 1\right)\right) - t\mathbb{P}\left(\sum_{k=2}^{h(\Gamma)}\sum_{i=1}^{\Gamma(k)} Y_{k,i} > t\right).$$

Here, in the first expression, Γ is a random variable on Λ with law P and $\{X_{k,i}\}_{k,i}$ are i.i.d. such that $X_{k,i}$ has law Q_k. Also, in the second expression, Γ has law P and $\{Y_{k,i}\}_{k,i}$ are independent from each other and from Γ such that $Y_{k,i}$ has the law of the random variable $[1-(Z_1^+ + \cdots + Z_{k-1}^+)]_0^1$ where Z_j are i.i.d. with law Q_k.

Proof. The condition $\mathbb{E}\left[\|\Gamma\|_1\right] < \infty$ guarantees that $\deg(\mu) < \infty$. Therefore, Proposition 6.10 implies that there exists a sequence $\epsilon_n \downarrow 0$ such that the sequence of ϵ_n–balanced allocations Θ_{ϵ_n} converges $\vec{\mu}$–a.e. to a Borel allocation Θ_0 which is balanced with respect to μ.

As $\vec{\mu}$ is supported on $\mathcal{T}_{**}$, Proposition 6.3 then implies that for $\vec{\mu}$–almost all $[T, e, i] \in \mathcal{T}_{**}$, we have that for all $(e', i') \in \Psi(T)$, $\Theta_{\epsilon_n}(T, e', i') \to \Theta_0(T, e', i')$. Since all hypertrees in $\mathcal{T}_{**}$ are locally finite, this means that for $\vec{\mu}$–almost all $[T, e, i] \in \mathcal{T}_{**}$, we have that for all $i' \in V(T)$, $\partial\Theta_{\epsilon_n}(T, i) \to \partial\Theta_0(T, i)$. Recall from Remark 6.12 that $\Theta_{\epsilon_n}(T, e, i) = \theta^T_{\epsilon_n}(e, i)$ where $\theta^T_{\epsilon_n}$ is the canonical ϵ_n–balanced allocation for T. Thereby, for $\vec{\mu}$–almost $[T, e, i] \in \mathcal{T}_{**}$, the conditions of Proposition 6.15 are satisfied. Thereby, for $\vec{\mu}$–almost all $[T, e, i]$, for all $(e', i') \in \Psi(T)$, $\rho^{\epsilon_n}_{T_{e' \to i'}}(.)$ converges pointwise to some $\rho_{T_{e' \to i'}}(.)$. Moreover, for $\vec{\mu}$–almost all $[T, e, i]$, we have that for all $(e', i') \in \Psi(T)$,

$$\rho^{-1}_{T_{e' \to i'}}(t) = t - \sum_{e'' \ni i' : e'' \neq e'} \left[1 - \sum_{j \in e'', j \neq i'} \left(\rho^{-1}_{T_{e'' \to j}}(t) \right)^+ \right]_0^1, \tag{6.76}$$

and

$$\partial\Theta_0(T, i') > t \qquad \Longleftrightarrow \qquad \sum_{e'' \ni i'} \left[1 - \sum_{j \in e'', j \neq i'} \left(\rho^{-1}_{T_{e'' \to j}}(t) \right)^+ \right]_0^1 > t. \tag{6.77}$$

For $[T, e, i]$ such that $\Theta_{\epsilon_n}(e', i')$ is not convergent for some $(e', i') \in \Psi(T)$, we may define $\rho^{-1}_{T_{e'' \to i''}}(.)$ arbitrarily for $(e'', i'') \in \Psi(T)$. This will not impact our argument, as this happens only on a measure zero set. Using Lemma 6.1 part (ii), we conclude that for μ–almost all $[T, i] \in \mathcal{T}_*$, we have that for all $(e', i') \in \Psi(T)$, (6.76) and (6.77) hold.

With this, we define the functions F and G on $\mathcal{T}_{**}$ as follows:

$$G(T, e, i) := \rho^{-1}_{T_{e \to i}}(t),$$

and

$$F(T, e, i) := \left[1 - \sum_{j \in e, j \neq i} \left(\rho^{-1}_{T_{e \to j}}(t) \right)^+ \right]_0^1 = \left[1 - \sum_{j \in e, j \neq i} G(T, e, j)^+ \right]_0^1.$$

Moreover, define the function $f : \mathcal{T}_* \to \{0, 1\}$ as $f(T, i) = \mathbb{1}_{\partial F(T,i) > t}$. From (6.77), μ–a.s. we have $f = \mathbb{1}_{\partial\Theta_0 > t}$. Hence, using the variational characterization in Proposition 6.11, we have

$$\int (\partial\Theta_0 - t)^+ d\mu = \int \tilde{f}_{\min} d\vec{\mu} - t \int f d\mu, \tag{6.78}$$

where $\tilde{f}_{\min} : \mathcal{H}_{**} \to [0, 1]$ is defined as

$$\tilde{f}_{\min}(T, e, i) = \frac{1}{|e|} \min_{j \in e} f(T, j).$$

With this and the definition of f, we have

$$\tilde{f}_{\min}(T, e, i) = \begin{cases} \frac{1}{|e|} & \partial F(T, j) > t \ \forall j \in e, \\ 0 & \text{otherwise.} \end{cases}$$

From (6.76), for $\vec{\mu}$–almost every $[T, e, i]$ we have

$$G(T, e, j) = t - \sum_{e' \ni j, e' \neq e} F(T, e', j) = t - \partial F(T, j) + F(T, e, j).$$

Therefore, $\vec{\mu}$–almost everywhere, $\partial F(T, j) > t$ iff $F(T, e, j) > G(T, e, j)$. Consequently, $\vec{\mu}$–a.e., we have

$$\tilde{f}_{\min}(T, e, i) = \begin{cases} \frac{1}{|e|} & F(T, e, j) > G(T, e, j) \ \forall j \in e, \\ 0 & \text{otherwise.} \end{cases} \tag{6.79}$$

Note that, by definition, we have $F(T, e, j) = \left[1 - \sum_{l \in e, l \neq j} G(T, e, l)^+\right]_0^1$. Thereby, using Lemma 6.16 in Section 6.8.1, $\vec{\mu}$–a.e. we have

$$\tilde{f}_{\min}(T, e, i) = \begin{cases} \frac{1}{|e|} & \sum_{j \in e} G(T, e, j)^+ < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the condition $\mathbb{E}\left[\|\Gamma\|_1\right] < \infty$ in particular implies that for all $k \geq 2$ we have $\mathbb{E}\left[\Gamma(k)\right] < \infty$. With this, using Lemma E.2 in Appendix E.6, we have

$$\int \tilde{f}_{\min} d\vec{\mu} = \sum_{k=2}^{\infty} \mathbb{E}\left[\Gamma(k)\right] \sum_{\gamma \in \Lambda} \hat{P}_k(\gamma) \mathbb{E}\left[f(\mathbb{T}, (k, 1), \emptyset) | \Gamma_\emptyset = \gamma + \mathbb{e}_k\right],$$

where $\mathbb{T}$ is the random rooted hypertree of Definition 6.28. Following the argument in the proof of Proposition 6.4 in Appendix E.6, this can be written as

$$\int \tilde{f}_{\min} d\vec{\mu} = \sum_{k=2}^{\infty} \mathbb{E}\left[\Gamma(k)\right] \mathbb{E}\left[\tilde{f}_{\min}(\tilde{\mathbb{T}}_k, (k, 1), \emptyset)\right],$$

where for $k \geq 2$, $\tilde{\mathbb{T}}_k$ is a tree with root $\emptyset$ that has an edge $(k, 1)$ of size k connected to the root, with the type of the other edges connected to the root being $\hat{P}_k$, and with the subtrees at the other vertices of all the edges (including the edge $(k, 1)$) generated according to the rules of $\mathsf{UGWHT}(P)$. Now, using the definition of $\tilde{f}_{\min}$, we have

$$\int \tilde{f}_{\min} d\vec{\mu} = \sum_{k=2}^{\infty} \frac{\mathbb{E}\left[\Gamma(k)\right]}{k} \mathbb{P}\left(\sum_{v \in (k,1)} G(\tilde{\mathbb{T}}_k, (k, 1), v)^+ < 1\right).$$

Now, for every k and $1 \leq i \leq k - 1$, let $\tilde{\mathbb{T}}_{k,i}$ be the hypertree below vertex $(k, 1, i)$ rooted at $(k, 1, i)$. Moreover, let $\tilde{\mathbb{T}}_{k,k}$ be the hypertree rooted at $\emptyset$ obtained from $\tilde{\mathbb{T}}_k$ by removing the edge $(k, 1)$ and all its subtree. Now, due to the construction of $\tilde{\mathbb{T}}_k$, $\tilde{\mathbb{T}}_{k,i}$, for $1 \leq i \leq k$

are i.i.d. $\mathrm{GWT}_k(P)$ hypertrees. Hence, $G(\tilde{\mathbb{T}}_k, (k,1), v)$ for $v \in (k,1)$ are independent and identically distributed. Let Q_k be the common distribution. This means that

$$\int \tilde{f}_{\min} d\vec{\mu} = \sum_{k=2}^{\infty} \frac{\mathbb{E}\left[\Gamma(k)\right]}{k} \mathbb{P}\left(\sum_{i=1}^{k} X_{k,i}^+ < 1\right), \tag{6.80}$$

where for each $k \geq 2$, $X_{k,i}$, $1 \leq i \leq k$ are i.i.d. with law Q_k.

On the other hand, with $\mathbb{T}$ being the random rooted hypertree of Definition 6.28, we have

$$\begin{aligned}\int f d\mu &= \mathbb{P}\left(\partial F(\mathbb{T}, \emptyset) > t\right) = \mathbb{P}\left(\sum_{k=2}^{h(\Gamma_\emptyset)} \sum_{i=1}^{\Gamma_\emptyset(k)} F(\mathbb{T}, (k,i), \emptyset) > t\right) \\ &= \mathbb{P}\left(\sum_{k=2}^{h(\Gamma_\emptyset)} \sum_{i=1}^{\Gamma_\emptyset(k)} \left[1 - \sum_{j=1}^{k-1} G(\mathbb{T}, (k,i), (k,i,j))^+\right]_0^1 > t\right),\end{aligned}$$

where $\Gamma_\emptyset$ is the type of the root in $\mathbb{T}$. But by definition, $G(\mathbb{T}, (k,i), (k,i,j)) = \rho^{-1}_{\mathbb{T}_{(k,i)\to(k,i,j)}}(t)$. But $\mathbb{T}_{(k,i)\to(k,i,j)}$ for $2 \leq k \leq \Gamma_\emptyset$, $1 \leq i \leq \Gamma_\emptyset(k)$, $1 \leq j \leq k-1$ are independent and $\mathbb{T}_{(k,i)\to(k,i,j)}$ has law $\mathrm{GWT}_k(P)$. Comparing this to the above definition of the distributions Q_k, $k \geq 2$, we realize that $G(\mathbb{T}, (k,i), (k,i,j))$ for $2 \leq k \leq \Gamma_\emptyset$, $1 \leq i \leq \Gamma_\emptyset(k)$, $1 \leq j \leq k-1$ are independent and $G(\mathbb{T}, (k,i), (k,i,j))$ has law Q_k. Consequently, as $\Gamma_\emptyset$ in the above expression has law P, we have

$$\int f d\mu = \mathbb{P}\left(\sum_{k=2}^{h(\Gamma)} \sum_{i=1}^{\Gamma(k)} Y_{k,i} > t\right), \tag{6.81}$$

where Γ has law P and $\{Y_{k,i}\}_{k,i}$ are independent from each other and from Γ such that $Y_{k,i}$ has the law of the random variable $[1 - (Z_1^+ + \cdots + Z_{k-1}^+)]_0^1$ where Z_j are i.i.d. with law Q_k. This together with (6.80) and the variational expression (6.78), completes the proof. □

6.9 Convergence of Maximum Load

In this section, we first introduce our configuration model and conditions under which it converges to the unimodular Galton–Watson hypertree model defined in Section 6.2.9. This is done in Section 6.9.1 below, specifically Theorem 6.4. We then state the conditions under which we prove Theorem 6.3 in Section 6.9.2, Proposition 6.18 and give the proof.

Before introducing our configuration model, we need to formally define multihypergraphs. Here, we only work with *finite* multihypergraphs.

Definition 6.34. *A finite multihypergraph $H = \langle V, E = (e_j, j \in J)\rangle$ consists of a finite vertex set V together with a finite edge index set J such that each hyperedge e_j is a multiset of vertices in V.*

Here, the assumption of e_j being a multiset allows for vertices to appear more than once in each edge. In this case, we call such an edge "improper". Moreover, it might be the case that $e_j = e_{j'}$ for $j \neq j' \in I$, in which case we call e_j and $e_{j'}$ "multiple edges".

6.9.1 Configuration model on Hypergraphs

We proposed a generalized Galton Watson process for hypertrees in Section 6.2.9 and showed that it is unimodular. In this section, we propose a configuration model which converges to it in the local weak sense under certain conditions.

Assume that, for each integer n, a type sequence $\gamma^{(n)} = (\gamma_1^{(n)}, \dots, \gamma_n^{(n)})$ is given such that $\gamma_i^{(n)} \in \Lambda$ and

$$\gamma_i^{(n)}(k) = 0 \qquad \forall\, 1 \le i \le n, k > n, \text{and} \tag{6.82a}$$

$$k \Bigg| \sum_{i=1}^n \gamma_i^{(n)}(k) \qquad \forall\, 2 \le k \le n, \tag{6.82b}$$

where the latter means that k divides $\sum_{i=1}^n \gamma_i^{(n)}(k)$. In what follows, we generate a random multihypergraph on the vertex set $\{1, \dots, n\}$ such that the type of node i is $\gamma_i^{(n)}$. For each $2 \le k \le n$ and $1 \le i \le n$, we attach $\gamma_i^{(n)}(k)$ many objects $e^k_{i,1}, \dots, e^k_{i,\gamma_i^{(n)}(k)}$ called "$1/k$–edges" to the node i . For each k, let $\Delta^{(n)}(k)$ be defined as the set of all $1/k$–edges, i.e.

$$\Delta^{(n)}(k) := \bigcup_{i=1}^n \{e^k_{i,1}, \dots, e^k_{i,\gamma_i^{(n)}(k)}\},$$

and let $\Delta^{(n)} := \cup_k \Delta^{(n)}(k)$ be the set of all "partial edges", where by a partial edge we mean a $1/k$–edge for some k. Also, let $\Delta_i^{(n)}$ be the set of all partial edges connected to a node $i \in \{1, \dots, n\}$, i.e.

$$\Delta_i^{(n)} = \bigcup_{k:\gamma_i^{(n)}(k)>0} \{e^k_{i,1}, \dots, e^k_{i,\gamma_i^{(n)}(k)}\}.$$

For a partial edge $e \in \Delta^{(n)}$, define $\nu(e)$ to be the node it corresponds to, i.e. $\nu(e^k_{i,j}) = i$. Also, define $|e|$ to be the size of e, i.e. $|e^k_{i,j}| = k$.

We say that a permutation σ_k is a k–matching on the set $\Delta^{(n)}(k)$ if it is a permutation on $\Delta^{(n)}(k)$ with no fixed points and also with all the cycles having size exactly equal to k. Due to the condition $k | \sum_{i=1}^n \gamma_i^{(n)}(k)$, such k–matchings exist for all k such that $\Delta^{(n)}(k) \neq \emptyset$. In fact, if for a finite nonempty set A, whose cardinality is divisible by k, we denote the set of k–matchings on A by $\mathcal{M}_k(A)$, it can be easily checked that $|\mathcal{M}_k(A)| = \frac{|A|!}{k^{|A|/k}(|A|/k)!}$.

Given this, for each k such that $\Delta^{(n)}(k) \neq \emptyset$, we pick σ_k uniformly at random from $\mathcal{M}_k(\Delta^{(n)}(k))$, independently over k. With this, we generate a random multihypergraph H_n on the set of vertices $\{1, \dots, n\}$. This is done by identifying each cycle of σ_k such that

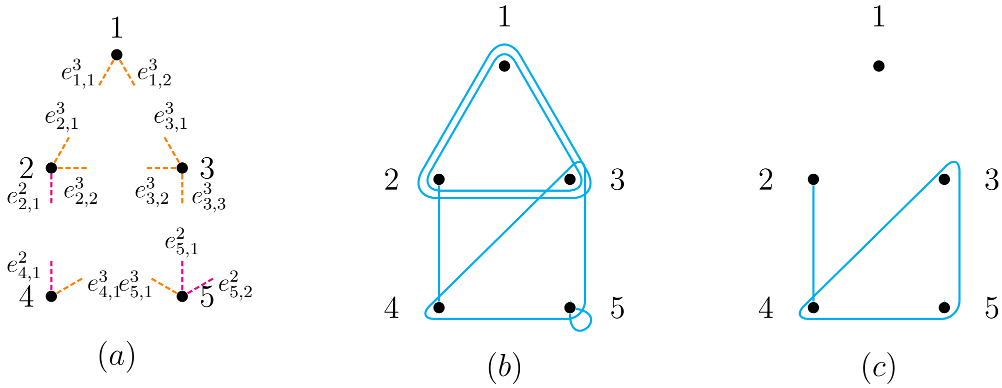

Figure 6.5: An example of an instance of the configuration model on $n = 5$ vertices with vertex types $\gamma_1^{(n)} = (0,2)$, $\gamma_2^{(n)} = (1,2)$, $\gamma_3^{(n)} = (0,3)$, $\gamma_4^{(n)} = (1,1)$, $\gamma_5^{(n)} = (1,2)$. Here, a type γ is represented by a vector where the first coordinate is $\gamma(2)$, the second is $\gamma(3)$ and so on. (a) illustrates the set of partial edges connected to vertices. (b) depicts the multihypergraph H_n formed by matching the partial edges using the permutations σ_2 and σ_3 which are represented in the cycle notation as $\sigma_2 = (e^2_{2,1}, e^2_{4,1})(e^2_{5,1}, e^2_{5,2})$ and $\sigma_3 = (e^3_{1,1}, e^3_{2,1}, e^3_{3,1})(e^3_{1,2}, e^3_{2,2}, e^3_{3,2})(e^3_{3,3}, e^3_{4,1}, e^3_{5,1})$. (c) illustrates the simple hypergraph H_n^e formed by removing the multiple edges of size 3 on vertices $1,2,3$ and also the improper edge on vertex 5.

$\Delta^{(n)}(k) \neq \emptyset$ of the form $(e, \sigma_k(e), \dots, \sigma_k^{(k-1)}(e))$ with the edge $\{\nu(e), \nu(\sigma_k(e)), \dots, \nu(\sigma_k^{(k-1)}(e))\}$ in H_n. Here, $e \in \Delta^{(n)}(k)$ and $\sigma_k^{(l)}$ denotes the permutation σ_k begin applied l times. Note that it is possible that two realizations of permutations as above result in the same multihypergraph H_n. As an example, if $n = 3$, $\gamma_1^{(n)}(2) = 2$, $\gamma_2^{(n)}(2) = \gamma_3^{(n)}(2) = 1$, and $\gamma_i^{(n)}(k) = 0$ for $k \neq 2$, $1 \le i \le 3$, then the two permutations σ_2 and σ'_2 on $\Delta^{(n)}(2)$ presented in the cycle notation as $\sigma_2 = (e^2_{1,1}, e^2_{2,1})(e^2_{1,2}, e^2_{3,1})$ and $\sigma'_2 = (e^2_{1,2}, e^2_{2,1})(e^2_{1,1}, e^2_{3,1})$ would result in the same multihypergraph (which turns out to be a simple graph in this example).

Note that, in general, H_n might not be simple. In particular, it might contain edges which contain a vertex more than once (we call such an edge "improper"), or it might be the case that two edges exist in H_n with the exact same multiset of vertices (we call such an edge a "multiple edge"), or these two can happen simultaneously. Having generated H_n, we may generate a simple hypergraph H_n^e by deleting all such improper edges and multiple edges in H_n, i.e. we first remove all improper edges and, subsequently, we delete all edges with the same set of endpoints. See Figure 6.5 for an example of an instance of the configuration model.

If one fixes a type sequence $\gamma^{(n)} = (\gamma_1^{(n)}, \dots, \gamma_n^{(n)})$ for each n, then the above configuration model generates a sequence of random hypergraphs H_n^e. Since H_n^e is random, $u_{H_n^e}$ is also a random probability distribution over $\mathcal{H}_*$. Let $\mathbb{E}\left[u_{H_n^e}\right]$ be the expectation with respect to the randomness of H_n^e, i.e. a weighted average over all the possible configurations of H_n^e (the number of such configurations is indeed finite for each n). Hence, $\mathbb{E}\left[u_{H_n^e}\right]$ is a sequence of probability distributions over $\mathcal{H}_*$ and one can ask whether it has a weak limit. On the other hand, one can build all H_n^e on a common probability space under which H_n^e are independent for different n. Then, for some $\mu \in \mathcal{P}(\mathcal{H}_*)$, we say that $u_{H_n^e} \Rightarrow \mu$ almost surely when outside a measure zero set in this common probability space, this convergence holds. The following theorem shows that under some conditions on the type sequence, H_n^e has a local weak limit. See Appendix E.7 for a proof.

Theorem 6.4. *Assume that a probability distribution P on Λ is given and define $I := \{k \geq 2 : P(\{\gamma \in \Lambda : \gamma(k) > 0\}) > 0\}$. Furthermore, let $\gamma^{(n)} = (\gamma_1^{(n)}, \dots, \gamma_n^{(n)})$ be a type sequence satisfying (6.82a) and (6.82b) such that*

$$\gamma_i^{(n)}(k) = 0 \qquad \forall k \notin I \quad \forall n, i, \tag{6.83a}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\gamma_i^{(n)} = \gamma} = P(\gamma) \qquad \forall \gamma \in \Lambda, \tag{6.83b}$$

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|\gamma_i^{(n)}\|_1^2 < \infty. \tag{6.83c}$$

Additionally, assume that there are constants $c_1, c_2, c_3, \epsilon > 0$ such that for n large enough,

$$\max_{1\leq i\leq n} \|\gamma_i^{(n)}\|_1 \leq c_1 (\log n)^{c_2}, \tag{6.84a}$$

$$\max_{1\leq i\leq n} h(\gamma_i^{(n)}) \leq c_1 (\log n)^{c_2}, \tag{6.84b}$$

$$\forall 2 \leq k \leq n \quad \Delta^{(n)}(k) = \emptyset \quad \text{ or } \quad |\Delta^{(n)}(k)| \geq c_3 n^{\epsilon}. \tag{6.84c}$$

Then, if H_n^e is the random hypergraph generated from the configuration model corresponding to $\gamma^{(n)}$ described above, we have $u_{H_n^e} \Rightarrow \mathsf{UGWHT}(P)$ almost surely.

Remark 6.13. *Note that the index set I would allow us to consider cases where there are only certain edge sizes in our model. For instance, when $I = \{2\}$, this theorem reduces to a statement for the graph pairing model, and when $I = \{k : k \geq 2\}$ it allows for all edge sizes to be present.*

Remark 6.14. *For a fixed k and for each n, define X_n to be an integer valued random variable taking value $\gamma_i^{(n)}(k)$ with probability $1/n$ for $1 \leq i \leq n$. Then, the condition (6.83c) implies that the sequence $\{X_n\}$ is uniformly integrable. Also, (6.83b) implies that $X_n \xrightarrow{d} \Gamma(k)$ where Γ has law P. Thus, $\mathbb{E}[X_n] \to \mathbb{E}[\Gamma(k)]$, i.e.*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \gamma_i^{(n)}(k) = \mathbb{E}[\Gamma(k)] \quad \forall k > 1. \tag{6.85}$$

This identity is useful in our future analysis.

Remark 6.15. *It can be proved that, under some regularity conditions for P, when the type sequence is generated i.i.d., the conditions of Theorem 6.4 are satisfied with probability one. However, we omit the proof of this claim here, since it is not central to our discussion.*

We will use the following simplified version of the above theorem in this section.

Corollary 6.2. *Assume that a probability distribution P over Λ is given such that for a finite set $I \subset \{2, 3, \dots, \}$, if Γ is a random variable with law P, we have $\mathbb{P}(\Gamma(k) > 0) > 0$ for $k \in I$ and $\mathbb{P}(\Gamma(k) > 0) = 0$ for $k \notin I$. Furthermore, assume $\gamma^{(n)} = (\gamma_1^{(n)}, \dots, \gamma_n^{(n)})$ is a type sequence satisfying (6.82a), (6.82b), (6.83a) and (6.83b). Additionally, assume that for some $\theta > 0$,*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n e^{\theta \|\gamma_i^{(n)}\|_1} < \infty. \tag{6.86}$$

Then, $u_{H_n^e} \Rightarrow \mathsf{UGWHT}(P)$ almost surely.

Proof. We check that in this regime, all the conditions of Theorem 6.4 are satisfied. Note that

$$\frac{1}{n}\sum_{i=1}^n \gamma_i^{(n)}(k)^2 \le \frac{2!}{\theta^2}\frac{1}{n}\sum_{i=1}^n e^{\theta\gamma_i^{(n)}(k)} \le \frac{2!}{\theta^2}\frac{1}{n}\sum_{i=1}^n e^{\theta\|\gamma_i^{(n)}\|_1},$$

which, together with (6.86), implies (6.83c). In order to show (6.84a), note that (6.86) implies there is a constant $\lambda < \infty$ such that $\frac{1}{n}\sum_{i=1}^n e^{\theta\|\gamma_i^{(n)}\|_1} < \lambda$ for all n. Now, $\exp(\theta \max_{1\le i\le n} \|\gamma_i^{(n)}\|_1) \le \sum_{i=1}^n e^{\theta\|\gamma_i^{(n)}\|_1} \le n\lambda$. Hence, $\max_{1\le i\le n}\|\gamma_i^{(n)}\|_1 \le (\log n + \log\lambda)/\theta$, which shows (6.84a). On the other hand, $h(\gamma_i^{(n)}) \le \max\{k : k \in I\}$ for all i and n. Hence, (6.84b) also holds.

For a fixed $k \in I$, using (6.85), which follows from (6.83c) which was proved above, for n large enough we have $|\Delta^{(n)}(k)| = \sum_{i=1}^n \gamma_i^{(n)}(k) \ge n\mathbb{E}[\Gamma(k)]/2$. Since there are finitely many $k \in I$, for n large enough we have $|\Delta^{(n)}(k)| \ge n \min_{k'\in I}\mathbb{E}[\Gamma(k')]/2$. But for $k' \in I$, we have $\mathbb{E}[\Gamma(k')] \ge \mathbb{P}(\Gamma(k') > 0) > 0$ by assumption. Consequently, (6.84c) holds with $\epsilon = 1$ and $c_3 := \min_{k'\in I}\mathbb{E}[\Gamma(k')]/2 > 0$. This means that all the conditions of Theorem 6.4 are satisfied and $u_{H_n^e} \Rightarrow \mathsf{UGWHT}(P)$ almost surely. □

6.9.2 Statement of the result

Here, we state the conditions under which we prove Theorem 6.3.

Proposition 6.18. *Assume that a probability distribution P over Λ is given such that for a finite set $I \subset \{2, 3, \dots, \}$, if Γ is a random variable with law P, we have $\mathbb{P}(\Gamma(k) > 0) > 0$ for $k \in I$ and $\mathbb{P}(\Gamma(k) > 0) = 0$ for $k \notin I$. Moreover, assume that for all $k \in I$, $\mathbb{E}[\Gamma(k)] < \infty$. Also, with k_{min} being the minimum element in I, assume that $\mathbb{P}(\Gamma(k_{min}) = 0) + \mathbb{P}(\Gamma(k_{min}) = 1) < 1$. Moreover, assume that a sequence of types $\gamma^{(n)} = (\gamma_1^{(n)}, \dots, \gamma_n^{(n)})$ is given satisfying (6.82a), (6.82b), (6.83a) (6.83b) and (6.86). Then, if H_n^e is the simple hypergraph generated* *by the configuration model in Section 6.9.1, $\varrho(H_n^e)$ converges in probability to $\varrho(\mu)$, where $\mu = \mathsf{UGWHT}(P)$.*

Before proving this proposition, we need the following two lemmas, whose proofs are given at the end of this section.

Lemma 6.17. *With the assumptions of Proposition 6.18, there is a constant $c_4 > 0$ such that for n large enough, for any subset $S \subset \{1, \dots, n\}$, the number of edges in H_n^e with all endpoints in S is stochastically dominated by the sum of s independent Bernoulli random variables, each having mean $c_4 s^{k_{min}-1}/n^{k_{min}-1}$, where $s := \sum_{i \in S} \|\gamma_i^{(n)}\|_1$.*

Lemma 6.18. *With the assumptions of Proposition 6.18, if $t > \frac{1}{k_{min}-1}$, there exists $\delta > 0$ such that if $Z_{\delta,t}^{(n)}$ denotes the number of subsets S of $\{1, \dots, n\}$ with size at most $n\delta$ where H_n^e has at least $t|S|$ many edges with all endpoints in S, we have $\mathbb{P}\left(Z_{\delta,t}^{(n)} > 0\right) \to 0$ as $n \to \infty$.*

Proof of Proposition 6.18. Let μ_n denote $u_{H_n^e}$, which is a random probability distribution on $\mathcal{H}_*$. Then, Corollary 6.2 guarantees that $\mu_n \Rightarrow \mu$ almost surely. As a result, if $\mathcal{L}_n$ is the law of the balanced load for H_n^e and $\mathcal{L}$ is the law of $\partial\Theta$, where Θ is the balanced allocation corresponding to μ, then, using Theorem 6.1, we have $\mathcal{L}_n \Rightarrow \mathcal{L}$ almost surely. Now, let $t := \varrho(\mu)$, and fix $\epsilon > 0$. Due to the definition of $\varrho(\mu)$, $\mathcal{L}((t-\epsilon, \infty)) > 0$. As a result, using the portmanteau theorem (Theorem 2.1 in Chapter 2) since $\mu_n \Rightarrow \mu$ almost surely, we have

$$\liminf_n \mathcal{L}_n((t-\epsilon, \infty)) > 0 \qquad a.s..$$

This means that

$$\mathbb{P}\left(\varrho(H_n^e) \le t - \epsilon\right) = \mathbb{P}\left(\mathcal{L}_n((t-\epsilon, \infty)) = 0\right) \to 0. \tag{6.87}$$

Now we show that $\mathbb{P}\left(\varrho(H_n^e) \ge t + \epsilon\right)$ also converges to zero. To do so, fix some $\delta > 0$ and note that

$$\begin{aligned} \mathbb{P}\left(\varrho(H_n^e) \ge t + \epsilon\right) &= \mathbb{P}\left(\mathcal{L}_n([t+\epsilon, \infty)) > 0\right) \\ &= \mathbb{P}\left(\mathcal{L}_n([t+\epsilon, \infty)) > \delta\right) + \mathbb{P}\left(0 < \mathcal{L}_n([t+\epsilon, \infty)) \le \delta\right). \end{aligned} \tag{6.88}$$

The portmanteau theorem implies that $\mathbb{P}\left(\mathcal{L}_n([t+\epsilon, \infty)) > \delta\right)$ converges to zero. Now, we argue that the second term also converges to zero. If θ_n denotes the balanced allocation on H_n^e, then, by the definition of $\mathcal{L}_n$, on the event $0 < \mathcal{L}_n([t+\epsilon, \infty)) \le \delta$, the set $S := \{1 \le i \le n : \partial\theta_n(i) \ge t + \epsilon\}$ is non–empty and $|S| \le \delta n$. Now if $e \in E(H_n^e)$ with $i, j \in e$ such that $i \in S$ and $j \in S^c$, then, since $\partial\theta_n(j) < t + \epsilon \le \partial\theta_n(i)$, we have $\theta_n(e, i) = 0$. Hence, for all $i \in S$, $\partial\theta_n(i) = \sum_{e \subseteq S, e \in E(H_n^e)} \theta_n(e, i)$. Thus,

$$\sum_{i \in S} \partial\theta_n(i) = |\{e \in E(H_n^e) : e \subseteq S\}| = |E_{H_n^e}(S)|.$$

On the other hand, we have $\sum_{i \in S} \partial\theta_n(i) \ge |S|(t+\epsilon)$. Hence, $|E_{H_n^e}(S)| \ge (t+\epsilon)|S|$, where $E_{H_n^e}(S)$ denotes the set of edges in H_n^e with all endpoints in S. As a result, if $Z_{\delta,t+\epsilon}^{(n)}$ denotes the number of subsets $S \subset \{1,\dots,n\}$ with size at most $n\delta$ such that H_n^e has at least $(t+\epsilon)|S|$ many edges with all endpoints in S, we have

$$\mathbb{P}\left(0 < \mathcal{L}_n([t+\epsilon,\infty)) < \delta\right) \le \mathbb{P}\Bigg(\exists\, S \subseteq \{1,\dots,n\},\ 0 < |S| \le \delta n, \\ |E_{H_n^e}(S)| \ge (t+\epsilon)|S|\Bigg) \\ \le \mathbb{P}\left(Z^{(n)}_{\delta,t+\epsilon} > 0\right).$$

If we have $t \ge 1/(k_{\min}-1)$, Lemma 6.18 above implies that $\mathbb{P}\left(Z^{(n)}_{\delta,t+\epsilon} > 0\right)$ goes to zero as $n \to \infty$ and we are done. We now argue why this is the case. For this, note that Proposition 6.10 together with Proposition 6.3 imply that there exists a sequence of ϵ_m–balanced allocations Θ_{ϵ_m} such that for μ–almost all $[H,i] \in \mathcal{H}_*$, for all vertices $j \in V(H)$, $\partial\Theta_{\epsilon_m}(H,j) \to \partial\Theta_0(H,j)$ where Θ_0 is a balanced allocation with respect to μ. Moreover, using Remark 6.12, $\partial\Theta_{\epsilon_m}(H,j) = \partial\theta^H_{\epsilon_m}(j)$ where $\theta^H_{\epsilon_n}$ is the canonical ϵ–balanced allocation on H.

On the other hand, the assumption $\mathbb{P}\left(\Gamma(k_{\min})=0\right)+\mathbb{P}\left(\Gamma(k_{\min})=1\right) < 1$ guarantees that for all integer $M \ge 2$, there is a nonzero probability under μ that the Galton–Watson process has a finite sub–hypertree containing the root, and having M edges with all these edge having size $k_{\min}$. It can be easily seen that a finite hypertree with M edges all having size c has $1+M(c-1)$ vertices and there is a balanced allocation on such a hypertree such that all the vertices get the same amount of load, which is equal to $\frac{M}{1+M(c-1)}$. Motivated by the discussion in the previous paragraph, for μ–almost all $[H,i] \in \mathcal{H}_*$, $\partial\Theta_0(H,i) = \lim_{m\to\infty} \partial\theta^H_{\epsilon_m}(i)$ where $\theta^H_{\epsilon_m}$ is the canonical ϵ_m–balanced allocation on H. This, together with Proposition 6.6, implies that for each integer M, there is a nonzero probability under μ that $\partial\Theta_0(H,i)$ is at least $M/(1+M(k_{\min}-1))$. Sending $M \to \infty$, this means that $t = \varrho(\mu) \ge 1/(k_{\min}-1)$. As was discussed above, using Lemma 6.18, we have $\mathbb{P}\left(Z^{(n)}_{\delta,t+\epsilon} > 0\right) \to 0$. Thus, $\mathbb{P}\left(\varrho(H_n^e) \ge t+\epsilon\right)$ goes to zero as n goes to infinity. This together with (6.87) proves that $\varrho(H_n^e) \xrightarrow{p} \varrho(\mu)$. □

Proof of Lemma 6.17. For $k \in I$, let $s_k := \sum_{i \in S} \gamma_i^{(n)}(k)$ and $m_k := |\Delta^{(n)}(k)|$. As the set of edges in H_n^e is a subset of that of H_n, we may prove the result for H_n instead. This allows us to directly analyze the configuration model. Let A be the set of the partial edges connected to vertices in S. Note that $|A| = s$. We order the elements in A arbitrarily. At time $t=1$, we pick the smallest element in A. Let k_1 be the size of this edge. Then, we choose $k_1 - 1$ other partial edges in $\Delta^{(n)}(k_1)$ uniformly at random to form an edge in H_n. We continue this process until all the elements in A are used up. More precisely, at time t, we pick the smallest available partial edge in A, namely e_t, and if k_t is the size of e_t, we match e_t with $k_t - 1$ other elements in the available partial edges in $\Delta^{(n)}(k_t)$ uniformly at random to form an edge in H_n. At time t, let $s_{k_t}(t)$ and $m_{k_t}(t)$ be the number of available partial edges of size k_t in A and $\Delta^{(n)}(k_t)$, respectively. With this, if p_t denotes the probability that e_t is matched with partial edges all inside A,

$$p_t = \frac{s_{k_t}(t)-1}{m_{k_t}(t)-1} \times \frac{s_{k_t}(t)-2}{m_{k_t}(t)-2} \times \cdots \times \frac{s_{k_t}(t)-(k_t-1)}{m_{k_t}(t)-(k_t-1)} \leq \left(\frac{s_{k_t}(t)}{m_{k_t}(t)}\right)^{k_t-1}.$$

Note that if l_t denotes the number of partial edges of size k_t among $e_1,\dots,e_{t-1}$, $m_{k_t}(t)$ is precisely $m_{k_t} - l_t k_t$. On the other hand, $s_{k_t}(t) \leq s_{k_t} - l_t$, where equality holds only if all the partial edges of size k among $e_1,\dots,e_{t-1}$ are matched with partial edges outside A. With this,

$$p_t \leq \left(\frac{k_t s_{k_t}}{m_{k_t}}\right)^{k_t-1}.$$

This is because, if $k_t s_{k_t} \leq m_{k_t}$, the even stronger inequality $p_t \leq \left(\frac{s_{k_t}}{m_{k_t}}\right)^{k_t-1}$ holds, while otherwise the RHS is at least 1, and the inequality is trivial. Furthermore, as we saw in the proof of Corollary 6.2, there exists $\alpha > 0$ such that for n large enough and all $k \in I$, $m_k \geq n\alpha$. This together with the fact that $s_k \leq s$ for all $k \in I$ implies

$$p_t \leq k_{\max}^{k_{\max}-1}\left(\frac{s}{n\alpha}\right)^{k_{\min}-1},$$

where $k_{\max}$ denotes the maximum element in I. As this upper bound is a constant, and the above process can continue for at most s steps until we match all partial edges in A, the number of edges with all endpoints in S is stochastically dominated by Binomial$(s, k_{\max}^{k_{\max}-1}\left(\frac{s}{n\alpha}\right)^{k_{\min}-1})$. The proof is complete by setting $c_4 := k_{\max}^{k_{\max}-1}/\alpha^{k_{\min}-1}$. □

Proof of Lemma 6.18. For positive integers L and r, let $X_{L,r}^{(n)}$ denote the number of subsets $S \subset \{1,\dots,n\}$ with size L such that H_n^e has at least r many edges with all endpoints in S. With this, $\mathbb{E}\left[Z_{\delta,t}^{(n)}\right] = \sum_{L=1}^{\lfloor n\delta \rfloor} \mathbb{E}\left[X_{L,\lceil Lt \rceil}^{(n)}\right]$. Now, fix integers L and r such that $L \leq n\delta$, with $\delta > 0$ sufficiently small. Let $\mathcal{S}_L$ denote the set of $S \subset \{1,\dots,n\}$ with size equal to L. Using Lemma 6.17 and the fact that for a binomial random variable Z, $\mathbb{P}(Z \geq r) \leq (\mathbb{E}[Z])^r / r!$, we have

$$\mathbb{E}\left[X_{L,r}^{(n)}\right] \leq \sum_{S \in \mathcal{S}_L} \frac{1}{r!}\left(c_4 \frac{\left(\sum_{i \in S} \|\gamma_i^{(n)}\|_1\right)^{k_{\min}}}{n^{k_{\min}-1}}\right)^r.$$

Using the inequality $x^m \leq m! e^x$ which holds for $x \geq 0$ and integer m, this yields

$$\begin{aligned}\mathbb{E}\left[X_{L,r}^{(n)}\right] &\leq \sum_{S \in \mathcal{S}_L} \frac{(rk_{\min})! c_4^r}{r! \theta^{rk_{\min}} n^{r(k_{\min}-1)}} \prod_{i \in S} e^{\theta \|\gamma_i^{(n)}\|_1} \\ &\leq \frac{(rk_{\min})! c_4^r}{r! \theta^{rk_{\min}} n^{r(k_{\min}-1)}} \frac{1}{L!}\left(\sum_{i=1}^n e^{\theta \|\gamma_i^{(n)}\|_1}\right)^L,\end{aligned}$$

where $\theta > 0$ is as in the statement of Corollary 6.2. Using the assumption (6.86), there exists $\lambda > 0$ such that $\sum_{i=1}^n e^{\theta \|\gamma_i^{(n)}\|_1} < n\lambda$ for all n. Using this, together with the inequalities $L! \geq (L/e)^L$ and $(rk_{\min})!/r! \leq (rk_{\min})^{r(k_{\min}-1)}$, and rearranging the terms, this yields

$$\mathbb{E}\left[X_{L,r}^{(n)}\right] \leq \left(\frac{c_5 r}{n}\right)^{r(k_{\min}-1)} \left(\frac{en\lambda}{L}\right)^L,$$

where $c_5 := k_{\min} c_4^{\frac{1}{k_{\min}-1}} / \theta^{\frac{k_{\min}}{k_{\min}-1}}$. Note that we may assume $c_5 > 1$; otherwise, we may replace it with $c_5 \vee 1$ which makes this quantity even bigger; for, the exponent of c_5 is positive.

Using this bound for $r = \lceil Lt \rceil$, we have

$$\begin{aligned}\mathbb{E}\left[X_{L,\lceil Lt \rceil}^{(n)}\right] &\leq \left(\frac{c_5 \lceil Lt \rceil}{n}\right)^{\lceil Lt \rceil (k_{\min}-1)} \left(\frac{en\lambda}{L}\right)^L \\ &= \left(\frac{c_5 \lceil Lt \rceil}{L}\right)^{\lceil Lt \rceil (k_{\min}-1)} (e\lambda)^L \left(\frac{L}{n}\right)^{\lceil Lt \rceil (k_{\min}-1)-L}.\end{aligned}$$

Using $c_5 > 1$, $L/n < 1$, $Lt \leq \lceil Lt \rceil \leq L(t+1)$ and the assumption $t > 1/(k_{\min} - 1)$, we get the upper bound

$$\mathbb{E}\left[X_{L,\lceil Lt \rceil}^{(n)}\right] \leq f\left(\frac{L}{n}\right)^L,$$

where

$$f(x) := c_6 x^{t(k_{\min}-1)-1},$$

with $c_6 := e\lambda(c_5(t+1))^{(t+1)(k_{\min}-1)}$. Note that $f(L/n)^L = \exp(-ng(L/n))$, where

$$g(x) := -x \log c_6 - (t(k_{\min} - 1) - 1)x \log x.$$

As $t > 1/(k_{\min} - 1)$, there exists $a > 0$ such that $g(x)$ is strictly increasing and strictly positive in $(0, a)$. Now, we choose $\delta > 0$ such that $\delta < a$ and also $f(\delta) < 1$. This is possible since the assumption $t > 1/(k_{\min} - 1)$ guarantees $f(\delta) \to 0$ as $\delta \downarrow 0$. With this, for any $0 < \zeta < \delta$, we have

$$\begin{aligned}\mathbb{E}\left[Z_{\delta,t}^{(n)}\right] &\leq \sum_{L=1}^{\lfloor n\delta \rfloor} f(L/n)^L \\ &= \sum_{L=1}^{\lfloor n\zeta \rfloor} f(L/n)^L + \sum_{L=\lfloor n\zeta \rfloor+1}^{\lfloor n\delta \rfloor} \exp(-ng(L/n)).\end{aligned}$$

Using the facts that f is increasing in $(0, \infty)$, g is increasing in $(0, a)$ and $0 < L/n \leq \delta < a$, we have

$$\mathbb{E}\left[Z_{\delta,t}^{(n)}\right] \leq \left(\sum_{L=1}^{\lfloor n\zeta \rfloor} f(\zeta)^L\right) + n\delta \exp(-ng(\zeta)).$$

But $f(\zeta) < f(\delta) < 1$. Therefore,

$$\mathbb{E}\left[Z_{\delta,t}^{(n)}\right] \leq \frac{f(\zeta)}{1-f(\zeta)} + n\delta\exp(-ng(\zeta)).$$

Now, by sending n to infinity, the second term vanishes, because $g(\zeta) > 0$, and we get $\limsup \mathbb{E}\left[Z_{\delta,t}^{(n)}\right] \leq f(\zeta)/(1-f(\zeta))$. Furthermore, by sending $\zeta \to 0$, we get $\mathbb{E}\left[Z_{\delta,t}^{(n)}\right] \to 0$ which means $\mathbb{P}\left(Z_{\delta,t}^{(n)} > 0\right) \to 0$, as $Z_{\delta,t}^{(n)}$ is integer valued. □

6.10 Conclusion

We studied the asymptotic behavior of balanced allocations for a sequence of hypergraphs converging to a local weak limit. This is done by defining and analyzing balanced Borel allocations directly on the limit. We expressed the mean excess for the Galton–Watson limit in terms of fixed point distributional equations. Moreover, we proved the convergence of the maximum load under some conditions.

Chapter 7

Concluding Remarks

In this thesis, we discussed two main category of problems on large sparse graphs, namely the graphical data compression, and load balancing. We employed the framework of local weak convergence, or so called the objective method, to make sense of a notion of stochastic processes for sparse marked graphs. We also discussed a notion of entropy for such processes on sparse marked graphs, which we called the *marked BC entropy*. In the process of studying the problem of compression, we realized that this notion of entropy is indeed the information theoretic limit of compression for our framework. In particular, we introduced a universal lossless compression scheme which is capable of compressing a sequence of sparse marked graphs converging in the local weak sense, without a priori knowing the limit. Specifically, this compression scheme is capable of achieving the marked BC entropy associated to this unknown limit. Furthermore, we discussed a distributed compression scheme for sparse marked graphs. In particular, we introduced a version of the Slepian–Wolf theorem for sparse marked graphs which characterizes the rate region when the statistical description of the distributed graphical data can be modeled as being one of two types – as a member of a sequence of marked sparse Erdős–Rényi ensembles or as a member of a sequence of marked configuration model ensembles. Furthermore, we gave a generalization of this result for Erdős–Rényi and configuration model ensembles with more than two sources.

In addition to studying the problem of compression, we studied the problem of load balancing in networks. We did this by modeling the problem as a hypergraph where each hyperedge represents a task carrying one unit of load, and each node represents a server. In this model, the load of each hyperedge can be distributed among it endpoints. An allocation is a way of distributing this load. Motivated by the work of Hajek [Haj90], we studied balanced allocations, which are roughly speaking those allocations in which no demand desires to change its allocation. We analyzed the properties of the empirical distribution of the loads faced by the resources in balanced allocations for a sequence of hypergraphs converging in the local weak sense as their size goes to infinity. In the special case of unimodular hypergraph Galton–Watson processes, we characterized the asymptotic empirical load distribution at a typical resource via a fixed point distributional equation. Moreover, we characterized the asymptotics of the maximum load at a resource under some additional conditions. We achieved these in particular by generalizing the local weak convergence theory to hypergraphs. Our work is an extension to hypergraphs of Anantharam and Salez [AS16], which considered load balancing in graphs, and is aimed at more comprehensively resolving conjectures of Hajek [Haj90].

Motivated by the fact that real–world graphical data is usually sparse, the framework discussed in this thesis has broad applicability in studying problems involving sparse graphical data. Additionally, since we allow for graphs to be marked, our framework allows for modeling both the interaction between objects in a complex network, as well as additional data associated to such objects as a marked graph. Examples of such scenarios include social networks, internet graphs, and genomics and proteomics data.

Problems studied in this thesis should be considered as examples showing the wide-range applicability of the local weak convergence theory and the notion of marked BC entropy. In fact, this framework provides a viewpoint of stationary stochastic processes for sparse marked graphs. The theory of time series is the engine driving an enormous range of applications in areas such as control theory, communications, information theory and signal processing. It is to be expected that a theory of stationary stochastic processes for combinatorial structures, in particular graphs, would eventually have a similarly wide-ranging impact.

Appendix A

Proofs for Chapter 2

A.1 Proof of Lemma 2.1

We first prove that if the condition mentioned in Lemma 2.1 is satisfied, then $\mu_n \Rightarrow \mu$. Let $f : \bar{\mathcal{G}}_* \to \mathbb{R}$ be a uniformly continuous and bounded function. Since f is uniformly continuous, for fixed $\epsilon > 0$ there exists $\delta > 0$ such that $|f([G,o]) - f([G',o'])| < \epsilon$ for all $[G,o]$ and $[G',o']$ such that $\bar{d}_*([G,o],[G',o']) < \delta$. For this δ, choose k such that $1/(1+k) < \delta$. Note that since Ξ and Θ are finite there are countably many locally finite rooted trees with marks in Ξ and Θ and depth at most k. Therefore, one can find countably many rooted trees $\{(T_j, i_j)\}_{j=1}^\infty$ with depth at most k such that $A^k_{(T_j,i_j)} \cap \bar{\mathcal{T}}_*$ partitions $\bar{\mathcal{T}}_*$. On the other hand, as μ is a probability measure with its support being a subset of $\bar{\mathcal{T}}_*$, one can find finitely many of these (T_j, i_j), which we may index without loss of generality by $1 \le j \le m$, such that $\sum_{j=1}^m \mu(A^k_{(T_j,i_j)}) \ge 1 - \epsilon$. To simplify the notation, we use A_j for $A^k_{(T_j,i_j)}$, $1 \le j \le m$. Note that if $[G,o] \in A_j$, $\bar{d}_*([G,o],[T_j,i_j]) \le \frac{1}{1+k} < \delta$. Hence, if $\mathcal{A}$ denotes $\cup_{j=1}^m A_j$, we have

$$
\begin{aligned}
&\left| \int f d\mu - \sum_{j=1}^m f([T_j,i_j])\mu(A_j) \right| \\
&\quad \le \sum_{j=1}^m \left| \int_{A_j} f d\mu - f([T_j,i_j])\mu(A_j) \right| + \|f\|_\infty \mu(\mathcal{A}^c) \\
&\quad \le \epsilon(1 + \|f\|_\infty),
\end{aligned}
$$

where the last inequality uses the facts that $\mu(\mathcal{A}^c) \leq \epsilon$, $1/(1+k) < \delta$, and $|f([G,o]) - f([T_j, i_j])| < \epsilon$ for $[G,o] \in A_j$. Similarly, we have

$$\begin{aligned}
\left| \int f d\mu_n - \sum_{j=1}^{m} f([T_j, i_j])\mu(A_j) \right| \\
&\leq \left| \int f d\mu_n - \sum_{j=1}^{m} f([T_j, i_j])\mu_n(A_j) \right| \\
&\quad + \sum_{j=1}^{m} |f([T_j, i_j])||\mu_n(A_j) - \mu(A_j)| \\
&\leq \|f\|_\infty \left(1 - \sum_{j=1}^{m} \mu_n(A_j)\right) + \epsilon \\
&\quad + \|f\|_\infty \sum_{j=1}^{m} |\mu_n(A_j) - \mu(A_j)|.
\end{aligned}$$

Combining the two preceding inequalities, we have

$$\begin{aligned}
\left| \int f d\mu_n - \int f d\mu \right| &\leq \|f\|_\infty \left(1 - \sum_{j=1}^{m} \mu_n(A_j)\right) \\
&\quad + \|f\|_\infty \sum_{j=1}^{m} |\mu_n(A_j) - \mu(A_j)| + \epsilon(2 + \|f\|_\infty).
\end{aligned}$$

Now, as n goes to infinity, $\mu_n(A_j) \to \mu(A_j)$ by assumption and also $1 - \sum_{j=1}^{m} \mu_n(A_j) \to \mu(\mathcal{A}^c) \leq \epsilon$. Thus,

$$\limsup_{n\to\infty} \left| \int f d\mu_n - \int f d\mu \right| \leq 2\epsilon(1 + \|f\|_\infty).$$

Since $\|f\|_\infty < \infty$ and $\epsilon > 0$ is arbitrary, $\int f d\mu_n \to \int f d\mu$, whereby $\mu_n \Rightarrow \mu$.

For the converse, fix an integer $h \geq 0$ and a rooted marked tree (T, i) with depth at most h. Since $\mathbb{1}_{A^h_{(T,i)}}([G,o]) = \mathbb{1}_{A^h_{(T,i)}}([G',o'])$ when $\bar{d}_*([G,o],[G',o']) < 1/(1+h)$, we see that $\mathbb{1}_{A^h_{(T,i)}}$ is a bounded continuous function. This immediately implies that

$$\mu_n(A^h_{(T,i)}) = \int \mathbb{1}_{A^h_{(T,i)}} d\mu_n \to \int \mathbb{1}_{A^h_{(T,i)}} d\mu = \mu(A^h_{(T,i)}),$$

which completes the proof.

A.2 Some Properties of Marked Rooted Trees of Finite Depth

In this section we gather some useful properties of marked rooted trees of finite depth, which are used at various points during the discussion.

Given a marked rooted tree (T, o), integers $k, l \geq 1$, $t \in \Xi \times \bar{\mathcal{T}}_*^{k-1}$, and $t' \in \Xi \times \bar{\mathcal{T}}_*^{l-1}$, define

$$E_{k,l}(t, t')(T, o) := |\{v \sim_T o : T(v, o)_{k-1} \equiv t, T(o, v)_{l-1} \equiv t'\}|.$$

When $k = l$ this reduces to the notation we defined in Section 2.7, i.e. $E_{k,k}(t, t')(T, o)$ is the same as $E_k(t, t')(T, o)$.

Lemma A.1. *Assume (T, o) is a rooted marked tree with finite depth, and v and v' are offspring of the root. Then, if $T(o, v) \equiv T(o, v')$ and $\xi_T(v, o) = \xi_T(v', o)$, we have $T(v, o) \equiv T(v', o)$.*

Proof. We construct the rooted automorphism $f : V(T) \to V(T)$ as follows. We set $f(o) = o$, $f(v) = v'$ and $f(v') = v$. Moreover, we use the isomorphism $T(o, v) \equiv T(o, v')$ to map the nodes in the subtree of v to the nodes in the subtree of v' and vice versa. Finally, we set f to be the identity map on the rest of the tree. Indeed, f is an adjacency preserving bijection. On the other hand, the assumptions $\xi_T(v, o) \equiv \xi_T(v', o)$ and $T(o, v) \equiv T(o, v')$ imply that f preserves the marks. Therefore, f is an automorphism which maps $T(v, o)$ to $T(v', o)$. This completes the proof. □

Lemma A.2. *Assume $t^{(1)}, t^{(2)} \in \Xi \times \bar{\mathcal{T}}_*^h$ and $t' \in \Xi \times \bar{\mathcal{T}}_*^k$ are given such that $t^{(1)} \oplus t' = t^{(2)} \oplus t'$. Further, assume that $t^{(1)}[m] = t^{(2)}[m]$. Then, we have $t^{(1)} = t^{(2)}$.*

Proof. Let (T, o) be an arbitrary member of the equivalence class $t^{(1)} \oplus t'$. Therefore, we have $T(v, o) \equiv t^{(1)}$ and $T(o, v) \equiv t'$ for some $v \sim_T o$. On the other hand, by assumption, (T, o) is also a member of the equivalence class $t^{(2)} \oplus t'$. This means that $T(v', o) \equiv t^{(2)}$ and $T(o, v') \equiv t'$ for some $v' \sim_T o$. Moreover, by assumption, $\xi_T(v, o) = t^{(1)}[m] = t^{(2)}[m] = \xi_T(v', o)$. Since $T(o, v) \equiv T(o, v') \equiv t'$, Lemma A.1 above implies that $T(v, o) \equiv T(v', o)$, or equivalently $t^{(1)} = t^{(2)}$. □

Lemma A.3. *Assume (T, o) is a rooted marked tree with depth at most $k \geq 1$. Moreover, assume that, for some $l \geq 1$, $t \in \Xi \times \bar{\mathcal{T}}_*^l$, and $t' \in \Xi \times \bar{\mathcal{T}}_*^{k-1}$, we have $E_{l+1,k}(t, t')(T, o) > 0$. Then,*

$$E_{l+1,k}(t, t')(T, o) = |\{v \sim_T o : T(o, v)_{k-1} \equiv t', \xi_T(v, o) = t[m]\}|.$$

Proof. From the definition of $E_{l+1,k}(t, t')(T, o)$, we have

$$\begin{aligned} E_{l+1,k}(t, t')(T, o) &= |\{v \sim_T o : T(o, v)_{k-1} \equiv t', T(v, o)_l \equiv t\}| \\ &\leq |\{v \sim_T o : T(o, v)_{k-1} \equiv t', \xi_T(v, o) = t[m]\}|. \end{aligned}$$

Now, we show the inequality in the opposite direction. The assumption $E_{l+1,k}(t,t')(T,o) > 0$ implies that there exists $v \sim_T o$ such that $T(o,v)_{k-1} \equiv t'$ and $T(v,o)_l \equiv t$. This in particular means that $\xi_T(v,o) = t[m]$. On the other hand, if $v' \sim_T o$ is such that $T(o,v')_{k-1} \equiv t'$ and $\xi_T(v',o) = t[m]$, Lemma A.1 above implies that $T(v',o) \equiv T(v,o)$, which means $T(v',o)_l \equiv t$. This establishes the other direction of the inequality and completes the proof. □

Lemma A.4. *Given $h \geq 1$ and two marked rooted trees (T,o) and (T',o') with depth at most h, assume that the mark at the root in T and T' are the same and also, for all $t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$, we have $E_h(t,t')(T,o) = E_h(t,t')(T',o')$. Then, $(T,o) \equiv (T',o')$.*

Proof. Note that the assumption regarding the mark at the root in T and that in T' being equal is necessary in this statement. To see this, consider the example where (T,o) and (T',o') are isolated roots, then we automatically have $E_h(t,t')(T,o) = E_h(t,t')(T',o') = 0$ for all $t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$, but $(T,o) \equiv (T',o')$ only when the marks at the root in the two rooted trees are the same.

Since the root marks in $[T,o]$ and $[T',o']$ are the same, it suffices to show that for all $x \in \Xi$ and $t \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$ we have

$$|\{v \sim_T o : \xi_T(v,o) = x, T(o,v)_{h-1} \equiv t\}| = |\{v \sim_{T'} o' : \xi_{T'}(v,o') = x, T'(o',v)_{h-1} \equiv t\}|. \tag{A.1}$$

Note that if $|\{v \sim_T o : \xi_T(v,o) = x, T(o,v)_{h-1} \equiv t\}| > 0$ then there exists $v \sim_T o$ such that $\xi_T(v,o) = x$ and $T(o,v)_{h-1} \equiv t$. This means that with $t' := T[v,o]_{h-1}$, we have $E_h(t',t)(T,o) > 0$. Moreover, Lemma A.3 implies that $E_h(t',t)(T,o) = |\{v \sim_T o : \xi_T(v,o) = x, T(o,v)_{h-1} \equiv t\}|$. On the other hand, from the hypothesis of this lemma, we also know that $E_h(t',t)(T',o') = E_h(t',t)(T,o) > 0$. Another usage of Lemma A.3 shows (A.1). So far, we have shown that $|\{v \sim_T o : \xi_T(v,o) = x, T(o,v)_{h-1} \equiv t\}| > 0$ implies (A.1). Similarly, $|\{v \sim_{T'} o : \xi_{T'}(v,o') = x, T'(o',v)_{h-1} \equiv t\}| > 0$ implies (A.1). This completes the proof. □

A.3 Some Properties of Unimodular Galton–Watson Trees with given Neighborhood Distribution

Fix $h \geq 1$ and $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ admissible. In this section, we prove some properties of $\mathsf{UGWT}_h(P)$.

Given $[T,o] \in \bar{\mathcal{T}}_*$, and $v \in V(T)$ with $v \neq o$, let $p(v)$ denote the parent node of v. For such v, we denote $(T[p(v),v]_{h-1}, T[v,p(v)]_{h-1})$ by $c(v)$ and $(T[v,p(v)]_{h-1}, T[p(v),v]_{h-1})$ by $\bar{c}(v)$. Let

$$\gamma_{[T,o]}(v) = \begin{cases} P([T,o]_h) & v = o, \\ \widehat{P}_{c(v)}(T[p(v),v]_h) & v \neq o. \end{cases} \tag{A.2}$$

When the marked rooted tree $[T,o]$ is clear from the context we will simply write $\gamma(v)$ for $\gamma_{[T,o]}(v)$.

Lemma A.5. *Given $[T,o] \in \bar{\mathcal{T}}_*$ and $v \in V(T)$ with $dist_T(v,o) = k$ where $k \geq 1$, let $o = v_0, v_1, \dots, v_k = v$ denote the path connecting v to the root.*

Then, if $\gamma(v_i) > 0$ for all $0 \leq i \leq k-1$, we have $P([T,v_i]_h) > 0$ for $0 \leq i \leq k-1$, and $e_P(c(v_i)) > 0$ for $1 \leq i \leq k$.

Proof. We prove this by induction on k. First, consider $k = 1$. In this case, we have $\gamma(v_0) = \gamma(o) = P([T,o]_h)$, and so the hypothesis that $\gamma(v_0) > 0$ implies that $P([T,v_0]_h) = P([T,o]_h) > 0$, which establishes the first claim. Using this, we get

$$\begin{aligned} e_P(\bar{c}(v)) &= e_P(T[v,o]_{h-1}, T[o,v]_{h-1}) \geq P([T,o]_h) E_h(T[v,o]_{h-1}, T[o,v]_{h-1})([T,o]_h) \\ &\geq P([T,o]_h) > 0, \end{aligned}$$

where the last equality uses the fact that, by definition, $E_h(T[v,o]_{h-1}, T[o,v]_{h-1})([T,o]_h) \geq 1$. But, since P is admissible, we have $e_P(c(v)) = e_P(\bar{c}(v)) > 0$ which completes the proof for $k = 1$.

Now, for $k > 1$, we have $e_P(c(v_{k-1})) > 0$ from the induction hypothesis. This implies that, with $t := T[v_{k-2}, v_{k-1}]_{h-1}$, $t' := T[v_{k-1}, v_{k-2}]_{h-1}$ and $\tilde{t} := T[v_{k-2}, v_{k-1}]_h$, we have

$$\begin{aligned} 0 < \gamma(v_{k-1}) &= \widehat{P}_{t,t'}(\tilde{t}) \\ &= \frac{P(\tilde{t} \oplus t') E_h(t,t')(\tilde{t} \oplus t')}{e_P(t,t')}. \end{aligned}$$

In particular, we have $P(\tilde{t} \oplus t') > 0$. But $\tilde{t} \oplus t' = [T, v_{k-1}]_h$. This together with the induction hypothesis implies that $P([T,v_i]_h) > 0$ for $0 \leq i \leq k-1$. Moreover, we have

$$e_P(\bar{c}(v)) \geq P([T,v_{k-1}]_h) E_h(\bar{c}(v))([T,v_{k-1}]_h) \geq P([T,v_{k-1}]_h) > 0,$$

where the last equality follows from the fact that, by definition, we have

$$E_h(\bar{c}(v))([T,v_{k-1}]_h) \geq 1.$$

The proof is complete by noting that, since P is admissible, we have $e_P(\bar{c}(v)) = e_P(c(v))$. □

Corollary A.1. *Let $\mu = \mathsf{UGWT}_h(P)$ with $h \geq 1$ and let $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$, i.e. P admissible. Then, for μ–almost all $[T,o] \in \bar{\mathcal{T}}_*$, we have $\gamma_{[T,o]}(v) > 0$ for all $v \in V(T)$ and $e_P(c(w)) > 0$ for all $w \in V(T) \setminus \{o\}$.*

Proof. First recall that $\gamma_{[T,o]}(o) = P([T,o]_h)$ is the probability of sampling $[T,o]_h$ in the process of generating $[T,o]$ with law $\mu = \mathsf{UGWT}_h(P)$. Hence, μ–almost surely, we have $\gamma_{[T,o]}(o) > 0$. Moreover, for a vertex $v \in V(T) \setminus \{o\}$, $\gamma_{[T,o]}(v) = \widehat{P}_{c(v)}(T[p(v),v]_h)$ is the probability of sampling $T[p(v),v]_h$ given $T[p(v),v]_{h-1}$ and $T[v,p(v)]_{h-1}$ in the process of generating $[T,o]$ with law μ. Since there are countably many vertices in $[T,o] \in \bar{\mathcal{T}}_*$, μ–almost surely we have $\gamma_{[T,o]}(v) > 0$ for all $v \in V(T)$.

Motivated by the above discussion, if $[T,o] \in \bar{\mathcal{T}}_*$ is outside a measure zero set with respect to μ, we have $\gamma_{[T,o]}(v) > 0$ for all $v \in V(T)$. Thus, Lemma A.5 above implies that $e_P(c(w)) > 0$ for all $w \in V(T) \setminus \{o\}$ and completes the proof. □

A.4 A Convergence Property of Unimodular Galton–Watson Trees with respect to the Neighborhood Distribution

In this section we give the proof of Lemma 2.4.

Proof of Lemma 2.4. Let $\mu^{(n)} := \mathsf{UGWT}_h(P^{(n)})$ and $\mu := \mathsf{UGWT}_h(P)$. We claim that for any integer $l \in \mathbb{N}$ and $[\hat{T}, \hat{o}] \in \bar{\mathcal{T}}_*^l$, we have

$$\lim_{n\to\infty} \mu^{(n)}(A_{[\hat{T},\hat{o}]}) = \mu(A_{[\hat{T},\hat{o}]}), \tag{A.3}$$

where

$$A_{[\hat{T},\hat{o}]} := \{[T, o] \in \bar{\mathcal{T}}_* : [T, o]_l = [\hat{T}, \hat{o}]\}.$$

Before proving our claim, we show why this implies $\mu^{(n)} \Rightarrow \mu$. To do this, we take a bounded and uniformly continuous function $f : \bar{\mathcal{T}}_* \to \mathbb{R}$ and show that $\int f d\mu^{(n)} \to \int f d\mu$. Fix $\epsilon > 0$. Due to the local topology on $\bar{\mathcal{T}}_*$, there is $l \in \mathbb{N}$ such that for all $[\hat{T}, \hat{o}] \in \bar{\mathcal{T}}_*^l$ and $[T, o] \in A_{[\hat{T},\hat{o}]}$, we have $d_*([T, o], [\hat{T}, \hat{o}]) < \epsilon$. Recall that d_* denotes the local metric on $\bar{\mathcal{T}}_*$. Since f is uniformly continuous, this implies that $|f([T, o]) - f([\hat{T}, \hat{o}])| < \eta(\epsilon)$ where $\eta(\epsilon) \to 0$ as $\epsilon \to 0$. Now, fix a finite collection $\mathcal{S}$ of marked rooted trees $[\hat{T}, \hat{o}] \in \bar{\mathcal{T}}_*^l$ such that

$$\sum_{[\hat{T},\hat{o}]\in\mathcal{S}} \mu(A_{[\hat{T},\hat{o}]}) > 1 - \epsilon.$$

Then, (A.3) implies that, for n large enough, we have

$$\sum_{[\hat{T},\hat{o}]\in\mathcal{S}} \mu^{(n)}(A_{[\hat{T},\hat{o}]}) > 1 - 2\epsilon.$$

This implies that

$$\left| \int f d\mu^{(n)} - \int f d\mu \right| \le 2\eta(\epsilon) + 3\epsilon \|f\|_\infty + \sum_{[\hat{T},\hat{o}]\in\mathcal{S}} |f([\hat{T}, \hat{o}])| |\mu^{(n)}(A_{[\hat{T},\hat{o}]}) - \mu(A_{[\hat{T},\hat{o}]})|.$$

By first sending n to infinity and then ϵ to zero, we get our desired result.

We now get back to proving our claim in (A.3). First, observe that, by the definition of $\mathsf{UGWT}_h(P)$, we have

$$\mu(A_{[\hat{T},\hat{o}]}) = C \prod_{v\in B} \gamma(v), \tag{A.4}$$

where C is a constant which only depends on $[\hat{T}, \hat{o}]$, and $B := \{v \in V(\hat{T}) : \mathrm{dist}_{\hat{T}}(v, \hat{o}) \le (l-h)_+\}$ where $(l-h)_+ := \max\{l-h, 0\}$. Here, we have employed the notation $\gamma(v) = \gamma_{[\hat{T},\hat{o}]}(v)$

from (A.2) in Appendix A.3. On the other hand, if we define $\gamma^{(n)}$ by replacing P with $P^{(n)}$ and $\widehat{P}_{c(v)}$ with $\widehat{P}^{(n)}_{c(v)}$ in the definition of γ, we have

$$\mu^{(n)}(A_{[\hat{T},\hat{o}]}) = C \prod_{v \in B} \gamma^{(n)}(v). \tag{A.5}$$

Note that, as C only depends on $[\hat{T},\hat{o}]$, the constants on (A.4) and (A.5) are the same. With this, we show (A.3) by considering two cases.

Case 1, $\mu(A_{[\hat{T},\hat{o}]}) > 0$: Using (A.4), this means that for all $v \in B$, we have $\gamma(v) > 0$. In particular, $P([\hat{T},\hat{o}]_h) > 0$ and Lemma A.5 above implies that for all $v \in B$, $v \neq \hat{o}$, we have $e_P(c(v)) > 0$. Hence, for $v \in B$, $v \neq \hat{o}$, we have

$$0 < \gamma(v) = \widehat{P}_{c(v)}(\hat{T}[p(v), v]_h) = \frac{P([\hat{T}, v]_h) E_h(c(v))([\hat{T}, v]_h)}{e_P(c(v))}.$$

Consequently, we have $P([\hat{T}, v]_h) > 0$. As a result, for n large enough, we have $P^{(n)}([\hat{T}, v]_h) > 0$. On the other hand, since $e_{P^{(n)}}(c(v)) \to e_P(c(v))$, for n large enough, we have $e_{P^{(n)}}(c(v)) > 0$ for all $v \in B$, $v \neq \hat{o}$. Therefore, for $v \in B$, $v \neq \hat{o}$ and n large enough, we have

$$\gamma^{(n)}(v) = \frac{P^{(n)}([\hat{T}, v]_h) E_h(c(v))([\hat{T}, v]_h)}{e_{P^{(n)}}(c(v))} \to \frac{P([\hat{T}, v]_h) E_h(c(v))([\hat{T}, v]_h)}{e_P(c(v))} = \gamma(v). \tag{A.6}$$

Moreover,

$$\gamma^{(n)}(\hat{o}) = P^{(n)}([\hat{T},\hat{o}]_h) \to P([\hat{T},\hat{o}]_h) = \gamma(\hat{o}).$$

Thus, together with (A.6), and comparing with (A.4) and (A.5), we realize that $\mu^{(n)}(A_{[\hat{T},\hat{o}]}) \to \mu(A_{[\hat{T},\hat{o}]})$.

Case 2, $\mu(A_{[\hat{T},\hat{o}]}) = 0$: Using (A.4), there is at least one node $v \in B$ such that $\gamma(v) = 0$. If $\gamma(\hat{o}) = P([\hat{T},\hat{o}]_h) = 0$, we have $P^{(n)}([\hat{T},\hat{o}]_h) \to P([\hat{T},\hat{o}]_h) = 0$. Hence, $\mu^{(n)}(A_{[\hat{T},\hat{o}]}) \to 0$ and we are done. Otherwise, let $v \in B$, $v \neq \hat{o}$, be a node with minimal depth such that $\gamma(v) = 0$, i.e. if $1 \leq k = \mathrm{dist}_{\hat{T}}(v, \hat{o})$ and $\hat{o} = v_0, v_1, \dots, v_k = v$ is the path connecting v to the root, we have $\gamma(v_i) > 0$ for $0 \leq i \leq k-1$ and $\gamma(v_k) = 0$. Using Lemma A.5, we conclude that $e_P(c(v_k)) > 0$ and thus

$$\begin{aligned} 0 = \gamma(v_k) &= \frac{P([\hat{T}, v_k]_h) E_h(c(v_k))([\hat{T}, v_k]_h)}{e_P(c(v_k))} \\ &\geq \frac{P([\hat{T}, v_k]_h)}{e_P(c(v_k))}, \end{aligned}$$

where the last line uses the fact that $E_h(c(v_k))([\hat{T}, v_k]_h) \geq 1$. This implies that $P([\hat{T}, v_k]_h) = 0$. Furthermore, since $P^{(n)}([\hat{T}, v_k]_h) \to P([\hat{T}, v_k]_h) = 0$ and $e_{P^{(n)}}(c(v_k)) \to e_P(c(v)) > 0$, we realize that for n large enough,

$$\gamma^{(n)}(v_k) = \frac{P^{(n)}([\hat{T}, v_k]_h) E_h(c(v_k))([\hat{T}, v_k]_h)}{e_{P^{(n)}}(c(v_k))} \to 0.$$

Consequently, using (A.5), we have $\mu^{(n)}(A_{[\hat{T},\hat{o}]}) \to 0 = \mu(A_{[\hat{T},\hat{o}]})$ which completes the proof.

A.5 Unimodularity of $\mathsf{UGWT}_h(P)$

We give a proof of Lemma 2.5. Let $h \geq 1$ and $P \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ be an admissible probability distribution. Let $\mu = \mathsf{UGWT}_h(P)$. In order to show that μ is unimodular, we need to show that for any Borel function $f : \bar{\mathcal{T}}_{**} \to \mathbb{R}_+$, we have

$$\mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, o, v) \right] = \mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, v, o) \right].$$

Without loss of generality, we may assume that $\deg(\mu) > 0$, since otherwise nothing remains to be proved. We have

$$\begin{aligned}
\mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, o, v) \right] &= \sum_{g \in \bar{\mathcal{T}}_*^h} P(g) \mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, o, v) \middle| (T, o)_h \equiv g \right] \\
&= \sum_{g \in \bar{\mathcal{T}}_*^h : \deg(g) > 0} \deg(g) P(g) \mathbb{E}_\mu \left[\frac{1}{\deg(g)} \sum_{v \sim_T o} f(T, o, v) \middle| (T, o)_h \equiv g \right],
\end{aligned} \tag{A.7}$$

where $\deg(g)$ denotes the degree at the root in g. Define the probability distribution $\widetilde{P} \in \mathcal{P}(\bar{\mathcal{T}}_*^h)$ such that

$$\widetilde{P}([T, o]) := \frac{P([T, o]) \deg_T(o)}{d},$$

where $d := \mathbb{E}_P [\deg_T(o)]$ is the expected degree at the root in P. Moreover, define the probability measure $\tilde{\mu} \in \mathcal{P}(\bar{\mathcal{T}}_*)$ in a way identical to $\mathsf{UGWT}_h(P)$, with the exception that $(T, o)_h$ in $\tilde{\mu}$ is sampled from $\widetilde{P}$ instead of P, and we use the distributions $\widehat{P}_{t,t'}$ to extend $(T, o)_h$ exactly as in $\mathsf{UGWT}_h(P)$. Since, by definition, conditioned on $(T, o)_h$, the distribution of (T, o) is the same in μ and $\tilde{\mu}$, we may write (A.7) as follows

$$\mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, o, v) \right] = d \sum_{g \in \bar{\mathcal{T}}_*^h} \widetilde{P}(g) \mathbb{E}_{\tilde{\mu}} \left[\frac{1}{\deg(g)} \sum_{v \sim_T o} f(T, o, v) \middle| (T, o)_h \equiv g \right].$$

With $\hat{v}$ being a node chosen uniformly at random among the nodes $v \sim_T o$ adjacent to the root in $[T, o] \sim \tilde{\mu}$, we may rewrite the above expression as follows,

$$\mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, o, v) \right] = d \mathbb{E}_{\tilde{\mu}} \left[f(T, o, \hat{v}) \right]. \tag{A.8}$$

Note that, $\tilde{\mu}$–almost surely, $\deg_T(o) > 0$ and $\hat{v}$ is well defined. Now, we find the distribution of $[T, o, \hat{v}] \in \bar{\mathcal{T}}_{**}$ when $[T, o] \sim \tilde{\mu}$ and $\hat{v}$ is chosen uniformly at random among the neighbors of the root, as was defined above.

In order to do so, we define the probability measure $\nu \in \mathcal{P}(\bar{\mathcal{T}}_{**})$ to be the law of $[H, o, o']$ where H is a connected random marked tree with two distinguished adjacent vertices o and o', defined as follows. We first sample t, t' from the distribution $\pi_P(t,t') = e_P(t,t')/d$, and construct H such that $H(o', o) = H(o', o)_{h-1} \equiv t$ and $H(o, o') = H(o, o')_{h-1} \equiv t'$. Then, similar to the construction of $\mathsf{UGWT}_h(P)$, we extend $H(o', o)$ and $H(o, o')$ inductively to construct H. More precisely, first we sample $\tilde{t}$ from $\widehat{P}_{t,t'}(.)$ and use it to add at most one layer to $H(o', o)_{h-1}$ so that $H(o', o)_h \equiv \tilde{t}$. Similarly, we sample $\tilde{t}'$ from $\widehat{P}_{t',t}(.)$ and use it to add at most one layer to $H(o, o')_{h-1}$ so that $H(o, o')_h \equiv \tilde{t}'$. Next, independently for $v \sim_H o, v \neq o'$, we sample $\tilde{t}$ from $\widehat{P}_{H[o,v]_{h-1}, H[v,o]_{h-1}}(.)$ and use it to add at most one layer to $H(o, v)_{h-1}$ such that $H(o, v)_h \equiv \tilde{t}$. We apply the same procedure to $w \sim_H o', w \neq o$. We continue this procedure inductively and define ν to be the law of $[H, o, o']$.

We now claim that if $[T, o]$ has distribution $\tilde{\mu}$ and $\hat{v}$ is chosen uniformly at random among the neighbors of the root in T as above, then $[T, o, \hat{v}]$ has distribution ν. Before proving this, we show how it completes the proof of the unimodularity of μ. Note that, with this claim proved, (A.8) becomes

$$\mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, o, v) \right] = d\mathbb{E}_\nu \left[f(H, o, o') \right].$$

Similarly, we have

$$\mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, v, o) \right] = d\mathbb{E}_\nu \left[f(H, o', o) \right].$$

However, the admissibility of P implies that $\pi_P(t, t') = \pi_P(t', t)$ for all $t, t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$. Therefore, ν is symmetric in the sense that $[H, o, o']$ and $[H, o', o]$ have the same distribution. Therefore, we have $\mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, o, v)\right] = \mathbb{E}_\mu \left[\sum_{v \sim_T o} f(T, v, o)\right]$, which is precisely what we needed to show.

Therefore, it remains to prove that with $[T, o] \sim \tilde{\mu}$ and $\hat{v}$ defined as above, $[T, o, \hat{v}] \sim \nu$. First, we claim that since $[T, o]_h \sim \widetilde{P}$, we have $(T[\hat{v}, o]_{h-1}, T[o, \hat{v}]_{h-1})$ has distribution π_P. In order to show this, note that due to the definition of $\widetilde{P}$ above, for $[T, o] \sim \widetilde{P}$, we have $\deg_T(o) \geq 1$ almost surely. Let Q be the distribution of $(T[\hat{v}, o]_{h-1}, T[o, \hat{v}]_{h-1})$ with $[T, o]$ and $\hat{v}$ as stated. Then, for $t, t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$, we have

$$\begin{aligned}
Q(t, t') &= \sum_{[T,o] \in \bar{\mathcal{T}}_*^h : \deg_T(o) \geq 1} \widetilde{P}([T, o]) \frac{E_h(t, t')(T, o)}{\deg_T(o)} \\
&= \sum_{[T,o] \in \bar{\mathcal{T}}_*^h : \deg_T(o) \geq 1} \frac{P([T, o]) \deg_T(o)}{d} \frac{E_h(t, t')(T, o)}{\deg_T(o)} \\
&= \frac{e_P(t, t')}{d} = \pi_P(t, t'),
\end{aligned}$$

which completes the proof of our claim. This, in particular, implies that, $\tilde{\mu}$–almost surely, we have $\pi_P(T[\hat{v},o]_{h-1}, T[o,\hat{v}]_{h-1}) > 0$. Moreover, we claim that for $t, t' \in \bar{\mathcal{T}}_*^{h-1}$ such that $\pi_P(t,t') > 0$ and $\tilde{t} \in \bar{\mathcal{T}}_*^h$ such that $\tilde{t}_{h-1} = t$, we have

$$\mathbb{P}_{\tilde{\mu}}\left(T[\hat{v},o]_h = \tilde{t} \,\middle|\, T[o,\hat{v}]_{h-1} = t', T[\hat{v},o]_{h-1} = t\right) = \widehat{P}_{t,t'}(\tilde{t}). \tag{A.9}$$

In order to show this, first note that, as was mentioned above, we have

$$\mathbb{P}_{\tilde{\mu}}\left(T[o,\hat{v}]_{h-1} = t', T[\hat{v},o]_{h-1} = t\right) = \pi_P(t,t'). \tag{A.10}$$

On the other hand, we have

$$\begin{aligned}
\mathbb{P}_{\tilde{\mu}}\left(T[\hat{v},o]_h = \tilde{t}, T[o,\hat{v}]_{h-1} = t'\right) &\overset{(a)}{=} \mathbb{P}_{\tilde{\mu}}\left([T,o]_h = \tilde{t} \oplus t'\right) \frac{1}{\deg(\tilde{t} \oplus t')} E_{h+1,h}(\tilde{t},t')(\tilde{t} \oplus t') \\
&= \frac{\widetilde{P}(\tilde{t} \oplus t')}{\deg(\tilde{t} \oplus t')} E_{h+1,h}(\tilde{t},t')(\tilde{t} \oplus t') \\
&= \frac{1}{d} P(\tilde{t} \oplus t') E_{h+1,h}(\tilde{t},t')(\tilde{t} \oplus t'),
\end{aligned}$$

where (a) is obtained by employing the assumption that $\hat{v}$ is chosen uniformly at random among the neighbors of the root, and the fact that conditioned on $[T,o]_h = \tilde{t} \oplus t'$, there are precisely $E_{h+1,h}(\tilde{t},t')(\tilde{t} \oplus t')$ many $v \sim_T o$ such that $T[v,o]_h = \tilde{t}$ and $T[o,v]_{h-1} = t'$. Here, $\deg(\tilde{t} \oplus t')$ denotes the degree at the root in $\tilde{t} \oplus t'$. Using this together with (A.10), we get

$$\begin{aligned}
\mathbb{P}_{\tilde{\mu}}\left(T[\hat{v},o]_h = \tilde{t} \,\middle|\, T[o,\hat{v}]_{h-1} = t', T[\hat{v},o]_{h-1} = t\right) &= \frac{P(\tilde{t} \oplus t') E_{h+1,h}(\tilde{t},t')(\tilde{t} \oplus t')}{d\pi_P(t,t')} \\
&= \frac{P(\tilde{t} \oplus t') E_{h+1,h}(\tilde{t},t')(\tilde{t} \oplus t')}{e_P(t,t')}.
\end{aligned} \tag{A.11}$$

Now, if $\tilde{o}$ denotes the root in $\tilde{t} \oplus t'$, we have

$$\begin{aligned}
E_{h+1,h}(\tilde{t},t')(\tilde{t} \oplus t') &\overset{(a)}{=} |\{v \sim_{\tilde{t} \oplus t'} \tilde{o} : (\tilde{t} \oplus t')(\tilde{o},v)_{h-1} \equiv t', \xi_{\tilde{t} \oplus t'}(v,\tilde{o}) = \tilde{t}[m]\}| \\
&\overset{(b)}{=} |\{v \sim_{\tilde{t} \oplus t'} \tilde{o} : (\tilde{t} \oplus t')(\tilde{o},v)_{h-1} \equiv t', \xi_{\tilde{t} \oplus t'}(v,\tilde{o}) = t[m]\}| \\
&\overset{(c)}{=} E_h(t,t')(\tilde{t} \oplus t'),
\end{aligned}$$

where in (a) we have used Lemma A.3, (b) is implied by the fact that $t[m] = \tilde{t}[m]$, and in (c) we have again used Lemma A.3. Substituting this into (A.11) and comparing with the definition of $\widehat{P}_{t,t'}$, we arrive at (A.9).

So far, we have shown that the distribution of $(T[o,\hat{v}]_{h-1}, T[\hat{v},o]_h)$ is the same as that of $(H[o,o']_{h-1}, H[o',o]_h)$ when $[H,o,o'] \sim \nu$. Observe that $T[o,\hat{v}]_{h-1}$ and $T[\hat{v},o]_h$ together form $[T,o]_h$. Moreover, by definition, conditioned on $(T,o)_h$, (T,o) is constructed using the $\widehat{P}_{t,t'}(.)$ distributions, in a way similar to the process of defining (H,o,o') above. Consequently, the distribution of $[T,o,\hat{v}]$ is identical to ν. As was discussed above, this completes the proof of the unimodularity of $\mathsf{UGWT}_h(P)$. □

A.6 Proof of Proposition 2.1

In this section, we prove Proposition 2.1.

Proof of Proposition 2.1. Let $\mu := \mathsf{UGWT}_h(P)$. Using induction, it suffices to show (2.8) only for $k = h+1$, i.e. with $Q := \mu_{h+1}$, we claim that

$$\mathsf{UGWT}_{h+1}(Q) = \mu. \tag{A.12}$$

Recall from Section 2.7 that μ is the law of $[T,o]$ where $(T,o)_h$ is sampled from P and the distributions $(\widehat{P}_{t,t'} : t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1})$ are used to extend depth $h-1$ rooted trees to depth h rooted trees in a recursive fashion. However, this process is equivalent to the following: first, we sample $(T,o)_{h+1}$ using Q and then recursively use $(\widehat{P}_{t,t'} : t,t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1})$ to extend depth $h-1$ trees to depth h trees, starting from nodes at depth 2. More precisely, for v being an offspring of the root and w being an offspring of v, we extend $T(v,w)_{h-1}$ to $T(v,w)_h$ using $\widehat{P}_{T[v,w]_{h-1},T[w,v]_{h-1}}$. This is done independently for all nodes w with depth 2. Equivalently, for each offspring v of the root, $T(o,v)_h$ is extended to $T(o,v)_{h+1}$, independent from all other offspring v' of the root. However, motivated by the above discussion, in order to extend $T(o,v)_h$, we need to know $T[v,w]_{h-1}$ and $T[w,v]_{h-1}$ for the offspring w of v. But this is known if we are given $T(o,v)_h$ and $T(v,o)_h$ (in fact, it is easy to see that even knowing $T(o,v)_h$ and $T(v,o)_{h-2}$ is sufficient). In other words, the distribution of $T[o,v]_{h+1}$ is uniquely determined by knowing $T[o,v]_h$ and $T[v,o]_h$. Motivated by this, for $s,s' \in \Xi \times \bar{\mathcal{T}}_*^h$ such that $e_Q(s,s') = \mathbb{E}_\mu\left[E_{h+1}(s,s')(T,o)\right] > 0$ and $\tilde{s} \in \bar{\mathcal{T}}_*^{h+1}$ such that $\tilde{s}_h = s$, define $\widetilde{P}_{s,s'}(\tilde{s})$ to be the probability of $T(o,v)_{h+1} \equiv \tilde{s}$ given $T(o,v)_h \equiv s$ and $T(v,o)_h \equiv s'$. The unimodularity of μ implies that if $e_Q(s,s') = 0$ for some $s,s' \in \Xi \times \bar{\mathcal{T}}_*^h$, the probability under μ of observing a node w with parent v such that $T(v,w)_h \equiv s$ and $T(w,v)_h \equiv s'$ is zero; therefore, we may define $\widetilde{P}_{s,s'}$ arbitrarily for such s,s'. Continuing this argument recursively for nodes at higher depths, we realize that μ is the law of (T,o) where $(T,o)_{h+1}$ is sampled from Q and then $(\widetilde{P}_{s,s'}(.) : s,s' \in \Xi \times \bar{\mathcal{T}}_*^h)$ is used to extend subtrees of depth h to subtrees of depth $h+1$. Comparing this with the construction of $\mathsf{UGWT}_{h+1}(Q)$, we realize that in order to show (A.12), it suffices to show that for every $s,s' \in \Xi \times \bar{\mathcal{T}}_*^h$ with $e_Q(s,s') > 0$, and for all $\tilde{s} \in \Xi \times \bar{\mathcal{T}}_*^{h+1}$, we have

$$\widehat{Q}_{s,s'}(\tilde{s}) = \widetilde{P}_{s,s'}(\tilde{s}), \tag{A.13}$$

where $\widehat{Q}_{s,s'}$ is defined using (2.7) based on the distribution Q. More precisely, for $s,s' \in \Xi \times \bar{\mathcal{T}}_*^h$ such that $e_Q(s,s') > 0$, and $\tilde{s} \in \Xi \times \bar{\mathcal{T}}_*^{h+1}$, we have

$$\widehat{Q}_{s,s'}(\tilde{s}) = \mathbb{1}\left[\tilde{s}_h = s\right] \frac{Q(\tilde{s} \oplus s') E_{h+1}(s,s')(\tilde{s} \oplus s')}{e_Q(s,s')}. \tag{A.14}$$

We now fix $s,s' \in \Xi \times \bar{\mathcal{T}}_*^h$ and show (A.13). Without loss of generality, we may assume that $\tilde{s}_h = s$, since otherwise both sides of (A.13) are zero. We claim that

$$\widetilde{P}_{s,s'}(\tilde{s}) = \frac{\mathbb{E}_\mu\left[E_{h+1,h+2}(s',\tilde{s})(T,o)\right]}{\mathbb{E}_\mu\left[E_{h+1}(s',s)(T,o)\right]}. \tag{A.15}$$

To see this, note that

$$\mathbb{E}_\mu\left[E_{h+1,h+2}(s',\tilde{s})(T,o)\right] = \sum_{r\in\bar{\mathcal{T}}_*^{h+1}} Q(r)\mathbb{E}_\mu\left[\sum_{v\sim_T o} \mathbb{1}\left[T(v,o)_h \equiv s', T(o,v)_{h+1}\equiv \tilde{s}\right] \middle| (T,o)_{h+1}\equiv r\right]$$
$$= \sum_{r\in\bar{\mathcal{T}}_*^{h+1}} Q(r)\mathbb{E}_\mu\left[\sum_{v\sim_T o} \mathbb{1}\left[T(v,o)_h \equiv s', T(o,v)_h\equiv s\right]\mathbb{1}\left[T(o,v)_{h+1}\equiv \tilde{s}\right] \middle| (T,o)_{h+1}\equiv r\right].$$

Note that the event $\{T(o,v)_{h+1}\equiv\tilde{s}\}$ is conditionally independent of the event $\{(T,o)_{h+1}\equiv r\}$, given the event $\{T(v,o)_h\equiv s', T(o,v)_h\equiv s\}$. Therefore,

$$\mathbb{E}_\mu\left[E_{h+1,h+2}(s',\tilde{s})(T,o)\right] = \sum_{r\in\bar{\mathcal{T}}_*^{h+1}} Q(r)E_{h+1}(s',s)(r)\widetilde{P}_{s,s'}(\tilde{s}) = \mathbb{E}_\mu\left[E_{h+1}(s',s)(T,o)\right]\widetilde{P}_{s,s'}(\tilde{s}). \tag{A.16}$$

Using the unimodularity of μ, we have

$$\mathbb{E}_\mu\left[E_{h+1}(s',s)(T,o)\right] = \mathbb{E}_\mu\left[E_{h+1}(s,s')(T,o)\right] = e_Q(s,s'). \tag{A.17}$$

Thereby, $e_Q(s,s')>0$ implies that $\mathbb{E}_\mu\left[E_{h+1}(s',s)(T,o)\right]>0$. Therefore, dividing both sides of (A.16) by $\mathbb{E}_\mu\left[E_{h+1}(s',s)(T,o)\right]$, we arrive at (A.15). Now, we simplify the right hand side of (A.15) to establish (A.13). Using the unimodularity of μ for the numerator, we have

$$\mathbb{E}_\mu\left[E_{h+1,h+2}(s',\tilde{s})(T,o)\right] = \mathbb{E}_\mu\left[E_{h+2,h+1}(\tilde{s},s')(T,o)\right].$$

Observe that $E_{h+2,h+1}(\tilde{s},s')(T,o)>0$ iff $(T,o)_{h+1}\equiv\tilde{s}\oplus s'$. On the other hand, if $(T,o)_{h+1}\equiv\tilde{s}\oplus s'$, we have $E_{h+2,h+1}(\tilde{s},s')(T,o) = E_{h+2,h+1}(\tilde{s},s')(\tilde{s}\oplus s')$. Consequently,

$$\begin{aligned}\mathbb{E}_\mu\left[E_{h+1,h+2}(s',\tilde{s})(T,o)\right] &= \mathbb{P}_\mu\left((T,o)_{h+1}\equiv\tilde{s}\oplus s'\right)E_{h+2,h+1}(\tilde{s},s')(\tilde{s}\oplus s')\\ &= Q(\tilde{s}\oplus s')E_{h+2,h+1}(\tilde{s},s')(\tilde{s}\oplus s').\end{aligned} \tag{A.18}$$

Note that $\tilde{s}\oplus s'$ by construction has the property that $E_{h+2,h+1}(\tilde{s},s')(\tilde{s}\oplus s')\geq 1$. Thereby, Lemma A.3 implies that $E_{h+2,h+1}(\tilde{s},s')(\tilde{s}\oplus s') = |\{v\sim_{\tilde{s}\oplus s'} o : (\tilde{s}\oplus s')(o,v)_h\equiv s', \xi_{\tilde{s}\oplus s'}(v,o) = \tilde{s}[m]\}|$. Here, o denotes the root in $\tilde{s}\oplus s'$. Likewise, since $\tilde{s}_h = s$, $E_{h+1}(s,s')(\tilde{s}\oplus s')\geq 1$, and another usage of Lemma A.3 implies that $E_{h+1}(s,s')(\tilde{s}\oplus s') = |\{v\sim_{\tilde{s}\oplus s'} o : (\tilde{s}\oplus s')(o,v)_h\equiv s', \xi_{\tilde{s}\oplus s'}(v,o) = s[m]\}|$. Also, $\tilde{s}_h = s$ in particular means $s[m]=\tilde{s}[m]$. Therefore, $E_{h+2,h+1}(\tilde{s},s')(\tilde{s}\oplus s') = E_{h+1}(s,s')(\tilde{s}\oplus s')$. Substituting into (A.18), we get

$$\mathbb{E}_\mu\left[E_{h+1,h+2}(s',\tilde{s})(T,o)\right] = Q(\tilde{s}\oplus s')E_{h+1}(s,s')(\tilde{s}\oplus s'). \tag{A.19}$$

Putting (A.17) and (A.19) back into (A.15) and comparing with (A.14), we arrive at (A.13), which completes the proof. □

A.7 Proof of Lemma 2.6

In this section we prove Lemma 2.6. First, we state the following lemma from [BC15] which will be useful in the proof.

Lemma A.6 (Lemma 5.4 in [BC15]). *Let $P = \{p_x, x \in \mathcal{X}\}$ be a probability measure on a discrete space $\mathcal{X}$ such that $H(P) < \infty$. Let $(\ell_x)_{x\in\mathcal{X}}$ be a sequence with $\ell_x \in \mathbb{Z}_+$, $x \in \mathcal{X}$, such that $\sum_x p_x \ell_x \log \ell_x < \infty$. Then $-\sum_x p_x \ell_x \log p_x < \infty$.*

Proof of Lemma 2.6. By Lemma 2.3, since μ is unimodular, $\widetilde{P}$ is admissible. Also,

$$\mathbb{E}_{\widetilde{P}}\left[\deg_T(o)\log\deg_T(o)\right] = \mathbb{E}_P\left[\deg_T(o)\log\deg_T(o)\right] < \infty.$$

Therefore, we only need to verify that $H(\widetilde{P}) < \infty$. Define $\nu := \mathsf{UGWT}_h(P)$ and let $P' := \nu_{h+1} \in \mathcal{P}(\bar{\mathcal{T}}_*^{h+1})$ be the distribution of the $h+1$–neighborhood of the root in ν. Here we have again used Lemma 2.3 to note that the unimodularity of μ implies that P is admissible, and hence $\mathsf{UGWT}_h(P)$ is well-defined. Now, we claim that

$$\sum_{s\in\bar{\mathcal{T}}_*^{h+1}} \widetilde{P}(s)\log\frac{1}{P'(s)} < \infty. \tag{A.20}$$

Using Gibbs' inequality, this implies that $H(\widetilde{P}) < \infty$ and completes the proof. Hence, it suffices to show (A.20).

Recall that, by the definition of $\mathsf{UGWT}_h(P)$, for $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$ we have

$$P'([T,o]) = CP([T,o]_h)\prod_{v\sim_T o}\widehat{P}_{T[o,v]_{h-1},T[v,o]_{h-1}}(T[o,v]_h), \tag{A.21}$$

where $C \geq 1$ is a constant that only depends on $[T,o]$ and counts the number of extensions of $[T,o]_h$ that result in $[T,o]$. Now, take $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$ such that $P'([T,o]) > 0$ and note that, using (A.21), we have $P([T,o]_h) > 0$. This, together with the fact that P is admissible, implies that for all $v \sim_T o$ we have

$$\begin{aligned} e_P(T[o,v]_{h-1}, T[v,o]_{h-1}) &= e_P(T[v,o]_{h-1}, T[o,v]_{h-1}) \\ &\geq P([T,o]_h)E_h(T[v,o]_{h-1}, T[o,v]_{h-1})(T,o) \\ &\geq P([T,o]_h) > 0. \end{aligned} \tag{A.22}$$

Therefore, for $v \sim_T o$, with $t := T[o,v]_{h-1}$ and $t' := T[v,o]_{h-1}$, we have $e_P(t,t') > 0$ and, using (2.7),

$$\begin{aligned} \widehat{P}_{t,t'}(T[o,v]_h) &= \frac{P([T,v]_h)E_h(t,t')([T,v]_h)}{e_P(t,t')} \\ &\geq \frac{P([T,v]_h)}{e_P(t,t')}, \end{aligned} \tag{A.23}$$

where the last line follows from the fact that $E_h(t,t')([T,v]_h) \geq 1$. Note that, as we have assumed $P'([T,o]) > 0$, from (A.21) we have $\widehat{P}_{t,t'}(T[o,v]_h) > 0$. Thereby, the first line in (A.23) implies that $P([T,v]_h) > 0$. So far, we have shown that for $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$ such that $P'([T,o]) > 0$, for all $v \sim_T o$ we have $P([T,v]_h) > 0$ and

$$\widehat{P}_{T[o,v]_{h-1},T[v,o]_{h-1}}(T[o,v]_h) \geq \frac{P([T,v]_h)}{e_P(T[o,v]_{h-1},T[v,o]_{h-1})}.$$

Substituting this in (A.21), we realize that for $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$ with $P'([T,o]) > 0$, we have

$$\log \frac{1}{P'([T,o])} \leq \log \frac{1}{P([T,o]_h)} + \sum_{v \sim_T o} \log \frac{1}{P([T,v]_h)} + \sum_{v \sim_T o} \log e_P(T[o,v]_{h-1},T[v,o]_{h-1}). \tag{A.24}$$

Next, we claim that $\widetilde{P} \ll P'$. Observe that from (A.21), for $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$, $P'([T,o]) = 0$ implies that either $P([T,o]_h) = 0$ or $P([T,o]_h) > 0$ and $\widehat{P}_{T[o,v]_{h-1},T[v,o]_{h-1}}(T[o,v]_h) = 0$ for some $v \sim_T o$. But if $P([T,o]_h) > 0$, (A.22) implies that for all $v \sim_T o$, we have $e_P(T[o,v]_{h-1},T[v,o]_{h-1}) > 0$. Therefore, using (A.23), if $\widehat{P}_{T[o,v]_{h-1},T[v,o]_{h-1}}(T[o,v]_h) = 0$ for some $v \sim_T o$, it must be the case that $P([T,v]_h) = 0$. Consequently, $P'([T,o]) = 0$ implies that either $P([T,o]_h) = 0$ or $P([T,o]_h) > 0$ and for some $v \sim_T o$, we have $P([T,v]_h) = 0$. Note that since $\widetilde{P}_h = P$, if $P([T,o]_h) = 0$, we have $\widetilde{P}([T,o]) = 0$. Now, we claim that if $P([T,v]_h) = 0$ for some $v \sim_T o$, then $\widetilde{P}([T,o]) = 0$. In order to establish this claim, using $\widetilde{P} = \mu_{h+1}$, we have

$$\begin{aligned}
\mathbb{E}_{\widetilde{P}}\left[\sum_{v \sim_T o} \mathbb{1}\left[P([T,v]_h) = 0\right]\right] &= \int \sum_{v \sim_T o} \mathbb{1}\left[P([T,v]_h) = 0\right] d\mu([T,o]) \\
&\stackrel{(a)}{=} \int \sum_{v \sim_T o} \mathbb{1}\left[P([T,o]_h) = 0\right] d\mu([T,o]) \\
&= \int \deg_T(o) \mathbb{1}\left[P([T,o]_h = 0)\right] d\mu([T,o]) \\
&\stackrel{(b)}{=} 0,
\end{aligned}$$

where (a) uses the unimodularity of μ and (b) uses the fact that $\mu_h = P$. This means that for $\widetilde{P}$–almost all $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$, $P([T,v]_h) > 0$ for all $v \sim_T o$. Equivalently, if $P([T,v]_h) = 0$ for $v \sim_T o$, we have $\widetilde{P}([T,o]) = 0$. To sum up, we showed that for $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$, $P'([T,o]) = 0$ implies $\widetilde{P}([T,o]) = 0$ and hence $\widetilde{P} \ll P'$. As a result, using this and (A.24), we may write the LHS of (A.20) as

$$\begin{aligned}\sum_{s\in\bar{\mathcal{T}}_*^{h+1}} \widetilde{P}(s)\log\frac{1}{P'(s)} &= \sum_{s\in\bar{\mathcal{T}}_*^{h+1}:P'(s)>0} \widetilde{P}(s)\log\frac{1}{P'(s)} \\ &\le \sum_{[T,o]\in\bar{\mathcal{T}}_*^{h+1}:P'([T,o])>0} \widetilde{P}([T,o])\Bigg(\log\frac{1}{P([T,o]_h)} + \sum_{v\sim_T o}\log\frac{1}{P([T,v]_h)} \\ &\qquad + \sum_{v\sim_T o}\log e_P(T[o,v]_{h-1},T[v,o]_{h-1})\Bigg).\end{aligned} \tag{A.25}$$

We may bound each component separately as follows. First, note that the facts $\widetilde{P}\ll P'$, $\widetilde{P}=\mu_{h+1}$ and $P=\mu_h$ imply that

$$\begin{aligned}\sum_{[T,o]\in\bar{\mathcal{T}}_*^{h+1}:P'([T,o])>0} \widetilde{P}([T,o])\log\frac{1}{P([T,o]_h)} &= \int \log\frac{1}{P([T,o]_h)}d\mu([T,o]) \\ &= -\sum_{s\in\bar{\mathcal{T}}_*^h} P(s)\log P(s) = H(P)<\infty.\end{aligned} \tag{A.26}$$

We also have

$$\begin{aligned}\sum_{[T,o]\in\bar{\mathcal{T}}_*^{h+1}:P'([T,o])>0} \widetilde{P}([T,o])\sum_{v\sim_T o}\log\frac{1}{P([T,v]_h)} &= \int\sum_{v\sim_T o}\log\frac{1}{P([T,v]_h)}d\mu([T,o]) \\ &\overset{(a)}{=} \int\sum_{v\sim_T o}\log\frac{1}{P([T,o]_h)}d\mu([T,o]) \\ &= \int \deg_T(o)\log\frac{1}{P([T,o]_h)}d\mu([T,o]) \\ &= -\sum_{[T,o]\in\bar{\mathcal{T}}_*^h}\deg_T(o)P([T,o])\log P([T,o]) \\ &\overset{(b)}{<}\infty\end{aligned} \tag{A.27}$$

where (a) follows from unimodularity of μ and (b) follows from Lemma A.6 and the fact that since P is strongly admissible, i.e. $P\in\mathcal{P}_h$, we have $\mathbb{E}_P\left[\deg_T(o)\log\deg_T(o)\right]<\infty$. Finally, for the third component, note that since

$$d=\mathbb{E}_P\left[\deg_T(o)\right]=\mathbb{E}_P\left[\sum_{t,t'\in\Xi\times\bar{\mathcal{T}}_*^{h-1}}E_h(t,t')(T,o)\right]=\sum_{t,t'\in\Xi\times\bar{\mathcal{T}}_*^{h-1}}e_P(t,t'),$$

we have $e_P(T[o,v]_{h-1}, T[v,o]_{h-1}) \leq d$ for all $[T,o] \in \bar{\mathcal{T}}_*^{h+1}$ and $v \sim_T o$. Consequently,

$$\begin{aligned}\sum_{[T,o]\in\bar{\mathcal{T}}_*^{h+1}:P'([T,o])>0} \widetilde{P}([T,o]) \sum_{v\sim_T o} \log e_P(T[o,v]_{h-1}, T[v,o]_{h-1}) \\ = \int \sum_{v\sim_T o} \log e_P(T[o,v]_{h-1}, T[v,o]_{h-1}) d\mu([T,o]) \\ \leq \int \deg_T(o)(\log d) d\mu([T,o]) = d\log d < \infty.\end{aligned} \tag{A.28}$$

Putting (A.26), (A.27) and (A.28) back in (A.25) we arrive at (A.20), which completes the proof. □

A.8 Proof of Proposition 2.2

In this section, we prove Proposition 2.2.

Proof of Proposition 2.2. Let $\mu := \mathsf{UGWT}_h(P)$. If P has a finite support then, as implied by Lemma 3.2 in Section 3.3.6 and Proposition 3.7 in Section 3.3.7, μ is sofic. If P does not have a finite support then, along the lines of the proof of Proposition 3.3 in Section 3.4.2, for $k > 1$, let $\mu^{(k)}$ be the law of $[T^{(k)}, o]$ obtained from $[T,o] \sim \mu$ as follows. For each vertex $v \in V(T)$, we remove all the edges connected to v if $\deg_T(v) \geq k$. Then, we let $T^{(k)}$ denote the connected component of the root in the resulting forest. As was shown in the proof of Proposition 3.3, with $P_k := (\mu^{(k)})_h$, as $k \to \infty$, we have $P_k \Rightarrow P$ and $e_{P_k}(t,t') \to e_P(t,t')$ for all $t, t' \in \Xi \times \bar{\mathcal{T}}_*^{h-1}$ (see (3.44)). Note that, from Lemma 2.5 in Appendix A.5, μ is unimodular. Thereby, it is easy to see that $\mu^{(k)}$ is also unimodular. Furthermore, from Lemma 2.3, P_k is admissible. The above discussion together with Lemma 2.4 in Appendix A.4 implies that $\mathsf{UGWT}_h(P_k) \Rightarrow \mu$. On the other hand, as we have discussed above, since P_k has a finite support, $\mathsf{UGWT}_h(P_k)$ is sofic. Therefore, a diagonal argument implies that μ is also sofic and completes the proof. It is worth recalling that we have earlier directly shown the unimodularity of μ in Appendix A.5. However, in general, being sofic might be a stronger property than being unimodular for all one knows at the moment. □

Appendix B

Proofs for Chapter 3

B.1 Calculations for Deriving (3.2)

First note that since $u^{(n)}(\theta)/n \to q_\theta$ for all $\theta \in \Theta$, we have

$$\log \frac{n!}{\prod_{\theta\in\Theta} u^{(n)}(\theta)!} = nH(Q) + o(n). \tag{B.1}$$

Furthermore, since $m^{(n)}(x,x')/n \to d_{x,x'}$ for $x \neq x'$, we have

$$\begin{aligned}
\log 2^{\sum_{x<x'} m^{(n)}(x,x')} &= n \sum_{x<x'} \frac{m^{(n)}(x,x')}{n} \log 2 \\
&= n \left(\sum_{x<x'} d_{x,x'} \log 2 + o(1) \right) \\
&= n \sum_{x<x'} d_{x,x'} \log 2 + o(n).
\end{aligned} \tag{B.2}$$

Moreover, from Stirling's approximation for the factorial, for a positive integer k we have $\log k! = k \log k - k + O(\log k)$. Moreover, since for all $x \neq x' \in \Xi$ we have $m^{(n)}(x,x')/n \to d_{x,x'} < \infty$ and for $x \in \Xi$ we have $m^{(n)}(x,x)/n \to d_{x,x}/2 < \infty$, we conclude that we have $m^{(n)}(x,x') = O(n)$ for all $x,x' \in \Xi$, and so

$$\begin{aligned}
\log \frac{\frac{n(n-1)}{2}!}{\prod_{x\leq x'\in\Xi} m^{(n)}(x,x')! \times \left(\frac{n(n-1)}{2} - \|\vec{m}^{(n)}\|_1\right)!} &= \frac{n(n-1)}{2} \log \frac{n(n-1)}{2} - \frac{n(n-1)}{2} \\
&- \sum_{x\leq x'} \left(m^{(n)}(x,x') \log m^{(n)}(x,x') - m^{(n)}(x,x')\right) \\
&- \left[\left(\frac{n(n-1)}{2} - \|\vec{m}^{(n)}\|_1\right) \log \left(\frac{n(n-1)}{2} - \|\vec{m}^{(n)}\|_1\right) - \left(\frac{n(n-1)}{2} - \|\vec{m}^{(n)}\|_1\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&\quad + O(\log n) \\
&= \frac{n(n-1)}{2}\log\frac{n(n-1)}{2} - \sum_{x\le x'} m^{(n)}(x,x')\log m^{(n)}(x,x') \\
&\quad - \frac{n(n-1)}{2}\log\left(\frac{n(n-1)}{2} - \|\vec{m}^{(n)}\|_1\right) + \|\vec{m}^{(n)}\|_1 \log\left(\frac{n(n-1)}{2} - \|\vec{m}^{(n)}\|_1\right) + o(n) \\
&= -\frac{n(n-1)}{2}\log\left(1 - \frac{2\|\vec{m}^{(n)}\|_1}{n(n-1)}\right) - n\sum_{x\le x'}\frac{m^{(n)}(x,x')}{n}\log\frac{m^{(n)}(x,x')}{n} - n\sum_{x\le x'}\frac{m^{(n)}(x,x')}{n}\log n \\
&\quad + \|\vec{m}^{(n)}\|_1 \log\left[n^2\left(\frac{n-1}{2n} - \frac{\|\vec{m}^{(n)}\|_1}{n^2}\right)\right] + o(n).
\end{aligned}
$$

Using the facts that $2\|\vec{m}^{(n)}\|_1/n(n-1) \to 0$ and $\log(1-x) = -x + O(x^2)$, this simplifies to

$$
\begin{aligned}
&= -\frac{n(n-1)}{2}\left[-\frac{2\|\vec{m}^{(n)}\|_1}{n(n-1)} + O\left(\frac{4\|\vec{m}^{(n)}\|_1^2}{(n(n-1))^2}\right)\right] - n\left(\sum_{x<x'} d_{x,x'}\log d_{x,x'} + \sum_x \frac{d_{x,x}}{2}\log\frac{d_{x,x}}{2}\right) \\
&\quad - \|\vec{m}^{(n)}\|_1\log n + 2\|\vec{m}^{(n)}\|_1\log n + n\frac{\|\vec{m}^{(n)}\|_1}{n}\log\left(\frac{n-1}{2n} - \frac{\|\vec{m}^{(n)}\|_1}{n^2}\right) + o(n).
\end{aligned}
$$

Since, by assumption, $\|\vec{m}^{(n)}\|_1/n \to \sum_{x<x'} d_{x,x'} + \sum_x d_{x,x}/2 = \sum_{x,x'} d_{x,x'}/2$, this simplifies to

$$
\begin{aligned}
&= n\frac{\|\vec{m}^{(n)}\|_1}{n} + \underbrace{O\left(\frac{\|\vec{m}^{(n)}\|_1^2}{n(n-1)}\right)}_{O(1)} - n\left(\sum_{x<x'} d_{x,x'}\log d_{x,x'} + \sum_x \frac{d_{x,x}}{2}\log\frac{d_{x,x}}{2}\right) \\
&\quad + \|\vec{m}^{(n)}\|_1\log n + n\left(\sum_{x,x'}\frac{d_{x,x'}}{2}\log\frac{1}{2} + o(1)\right) + o(n) \\
&= \|\vec{m}^{(n)}\|_1\log n + n\sum_{x,x'} s(d_{x,x'}) - n\sum_{x<x'} d_{x,x'}\log 2 + o(n).
\end{aligned}
$$

Using this together with (B.1) and (B.2), we get

$$
\begin{aligned}
\log|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}| &= nH(Q) + \|\vec{m}^{(n)}\|_1\log n + n\sum_{x,x'} s(d_{x,x'}) - n\sum_{x<x'} d_{x,x'}\log 2 \\
&\quad + n\sum_{x<x'} d_{x,x'}\log 2 + o(n) \\
&= \|\vec{m}^{(n)}\|_1\log n + nH(Q) + n\sum_{x,x'} s(d_{x,x'}) + o(n),
\end{aligned}
$$

which is precisely what was stated in (3.2).

Appendix C

Proofs for Chapter 4

C.1 Proofs for Section 4.2

Proof of Lemma 4.1. Let $\mathcal{A}$ be the set of $1 \le i \le n$ such that $[G, i]_h = [G', \pi(i)]_h$. Then, for any Borel set $B \subset \bar{\mathcal{G}}_*$, we have

$$U(G)(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left[[G, i] \in B\right]$$
$$\le \frac{1}{n} \sum_{i \in \mathcal{A}} \mathbb{1}\left[[G, i] \in B\right] + 1 - \frac{L}{n}.$$

Note that if for some $i \in \mathcal{A}$ we have $[G, i] \in B$ then, since $(G, i)_h \equiv (G', \pi(i))_h$, we have $d_*([G, i], [G', \pi(i)]) \le \frac{1}{1+h}$. This means that, for such i, $[G', \pi(i)] \in B^{\delta + 1/(1+h)}$ for arbitrary $\delta > 0$. Continuing the chain of inequalities, we have

$$U(G)(B) \le \frac{1}{n} \sum_{i \in \mathcal{A}} \mathbb{1}\left[[G', \pi(i)] \in B^{\delta + 1/(1+h)}\right] + 1 - \frac{L}{n}$$
$$\le \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left[[G', i] \in B^{\delta + 1/(1+h)}\right] + 1 - \frac{L}{n}$$
$$= U(G')(B^{\delta + 1/(1+h)}) + 1 - \frac{L}{n}.$$

Changing the order of G and G', we have

$$d_{\mathrm{LP}}(U(G), U(G')) \le \max\left\{\frac{1}{1+h} + \delta, 1 - \frac{L}{n}\right\}.$$

We get the desired result by sending δ to zero. □

Next, we prove Lemma 4.2. Before that, we state and prove the following lemmas which will be useful in our proof. For a marked graph G on a finite or countably infinite vertex set, let $\mathsf{UM}(G)$ denote the unmarked graph which has the same set of vertices and edges as in G, but is obtained from G by removing all the vertex and edge marks. Given a probability distribution $\mu \in \mathcal{P}(\mathcal{G}_*)$ on the space of isomorphism classes of rooted unmarked graphs, for $\epsilon > 0$ and integers n and m, let $\mathcal{G}_{n,m}(\mu, \epsilon)$ denote the set of unmarked graphs G on the vertex set $\{1, \dots, n\}$ with m edges such that $d_{\mathrm{LP}}(U(G), \mu) < \epsilon$.

Lemma C.1. *For $[G, o]$ and $[G', o']$ in $\bar{\mathcal{G}}_*$, we have*

$$d_*([\mathsf{UM}(G), o], [\mathsf{UM}(G'), o']) \le \bar{d}_*([G, o], [G', o']).$$

Proof. By definition, for $\epsilon > 0$, the condition $\bar{d}_*([G, o], [G', o']) < \epsilon$ means that for some k with $1/(1 + k) < \epsilon$, we have $[G, o]_k \equiv [G', o']_k$. This implies $[\mathsf{UM}(G), o]_k \equiv [\mathsf{UM}(G'), o']_k$, which in particular means that $d_*([\mathsf{UM}(G), o], [\mathsf{UM}(G'), o']) \le 1/(1 + k) < \epsilon$. □

Lemma C.2. *Assume $\mu \in \mathcal{P}(\bar{\mathcal{G}}_*)$ is given. Let $\tilde{\mu} \in \mathcal{P}(\mathcal{G}_*)$ be the law of $[\mathsf{UM}(G), o]$ when $[G, o]$ has law μ. Then, given an integer n, edge and vertex mark count vectors $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ respectively, and $\epsilon > 0$, for all $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon)$, we have $\mathsf{UM}(G) \in \mathcal{G}_{n, m_n}(\tilde{\mu}, \epsilon)$ where $m_n := \|\vec{m}^{(n)}\|_1$.*

Proof. Fix $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon)$. Note that $\mathsf{UM}(G)$ has m_n edges, and we only need to show that $d_{\mathrm{LP}}(U(\mathsf{UM}(G)), \tilde{\mu}) < \epsilon$. Let $\delta := d_{\mathrm{LP}}(U(G), \mu)$. This means that for all $\delta' > \delta$, and for all Borel sets A in $\bar{\mathcal{G}}_*$, we have $(U(G))(A) \le \mu(A^{\delta'}) + \delta'$ and $\mu(A) \le (U(G))(A^{\delta'}) + \delta'$, where, $A^{\delta'}$ denotes the δ'–extension of A. Define $T : \bar{\mathcal{G}}_* \to \mathcal{G}_*$ that maps $[G, o] \in \bar{\mathcal{G}}_*$ to $[\mathsf{UM}(G), o] \in \mathcal{G}_*$. Lemma C.1 above implies that T is continuous and in fact 1–Lipschitz. It is easy to see that $U(\mathsf{UM}(G))$ is the pushforward of $U(G)$ under the mapping T. Also, $\tilde{\mu}$ is the pushforward of μ under T. Using the fact that T is 1–Lipschitz, it is easy to see that for any Borel set B in $\mathcal{G}_*$, and any $\zeta > 0$, we have $(T^{-1}(B))^{\zeta} \subset T^{-1}(B^{\zeta})$. Putting these together, for $\delta' > \delta$ and a Borel set B in $\mathcal{G}_*$, we have

$$\begin{aligned} U(\mathsf{UM}(G))(B) &= U(G)(T^{-1}(B)) \le \mu((T^{-1}(B))^{\delta'}) + \delta' \\ &\le \mu(T^{-1}(B^{\delta'})) + \delta' = \tilde{\mu}(B^{\delta'}) + \delta'. \end{aligned}$$

Similarly,

$$\begin{aligned} \tilde{\mu}(B) = \mu(T^{-1}(B)) &\le (U(G))((T^{-1}(B))^{\delta'}) + \delta' \\ &\le (U(G))(T^{-1}(B^{\delta'})) + \delta' \\ &= (U(\mathsf{UM}(G)))(B^{\delta'}) + \delta'. \end{aligned}$$

Since this holds for any $\delta' > \delta$ and any Borel set B in $\mathcal{G}_*$, we have $d_{\mathrm{LP}}(U(\mathsf{UM}(G)), \tilde{\mu}) \le \delta = d_{\mathrm{LP}}(U(G), \mu) < \epsilon$. Consequently, we have $\mathsf{UM}(G) \in \mathcal{G}_{n, m_n}(\tilde{\mu}, \epsilon)$ and the proof is complete. □

Now, we are ready to prove Lemma 4.2.

Proof of Lemma 4.2. To simplify the notation, for $\epsilon > 0$ define

$$a_n(\epsilon) := \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)| - \|\vec{m}^{(n)}\| \log n}{n}.$$

Note that there exists a subsequence $\{n_k\}$ such that $\limsup_{n\to\infty} a_n(\epsilon_n) = \lim_{k\to\infty} a_{n_k}(\epsilon_{n_k})$. Moreover, there is a further subsequence n_{k_r} such that for all $x, x' \in \Xi$, there exists $\bar{d}_{x,x'} \in [0,\infty]$ where

$$\frac{m^{(n_{k_r})}(x,x')}{n_{k_r}} \to \bar{d}_{x,x'}, \qquad x \neq x'; \tag{C.1a}$$

$$\frac{2m^{(n_{k_r})}(x,x)}{n_{k_r}} \to \bar{d}_{x,x}. \tag{C.1b}$$

Observe that since $a_{n_k}(\epsilon_{n_k})$ is convergent, it suffices that we focus on the subsequence $\{n_{k_r}\}$ and show that $\lim a_{n_{k_r}}(\epsilon_{n_{k_r}}) \le \Sigma(\mu)$. Note that due to conditions (4.4a) and (4.4b), we have $\bar{d}_{x,x'} \ge \deg_{x,x'}(\mu)$ for all $x, x' \in \Xi$. Therefore, there are two possible cases: either $\bar{d}_{x,x'} = \deg_{x,x'}(\mu)$ for all $x, x' \in \Xi$, or there exist $x, x' \in \Xi$ such that $\bar{d}_{x,x'} > \deg_{x,x'}(\mu)$. To simplify the notation, without loss of generality, we may assume that the subsequence n_{k_r} is the whole sequence, i.e. $a_n(\epsilon_n)$ is convergent, and (C.1a) and (C.1b) hold for the whole sequence.

Case 1: $\bar{d}_{x,x'} = \deg_{x,x'}(\mu)$ for all $x, x' \in \Xi$. We define edge and vertex mark count vectors $\vec{\widetilde{m}}^{(n)} = (\widetilde{m}^{(n)}(x,x') : x, x' \in \Xi)$ and $\vec{\widetilde{u}}^{(n)} = (\widetilde{u}^{(n)}(\theta) : \theta \in \Theta)$ as follows. For $x, x' \in \Xi$, define $\widetilde{m}^{(n)}(x,x')$ to be $m^{(n)}(x,x')$ if $\deg_{x,x'}(\mu) > 0$ and 0 otherwise. Also, fix some $\theta_0 \in \Theta$ such that $\Pi_{\theta_0}(\mu) > 0$. For $\theta \in \Theta$, define

$$\widetilde{u}^{(n)}(\theta) := \begin{cases} 0, & \Pi_\theta(\mu) = 0, \\ u^{(n)}(\theta), & \Pi_\theta(\mu) > 0, \theta \neq \theta_0, \\ u^{(n)}(\theta_0) + \sum_{\theta' : \Pi_{\theta'}(\mu) = 0} u^{(n)}(\theta') & \theta = \theta_0. \end{cases}$$

Note that, by construction and from (C.1a) and (C.1b), the sequences $\vec{\widetilde{m}}^{(n)}$ and $\vec{\widetilde{u}}^{(n)}$ are adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. Also, $m^{(n)}(x,x') \ge \widetilde{m}^{(n)}(x,x')$ for all n and all $x, x' \in \Xi$.

Now, fix $\epsilon > 0$, and pick an integer h such that $1/(1+h) < \epsilon$. Define B to be the set of $[G,o] \in \bar{\mathcal{G}}_*$ such that either for some $x, x' \in \Xi$ with $\deg_{x,x'}(\mu) = 0$ there exists an edge in $[G,o]_h$ with pair of marks x, x', or, for some $\theta \in \Theta$ with $\Pi_\theta(\mu) = 0$, there exists a vertex in $[G,o]_h$ with mark θ. Then, from Lemma 2.2, we have $\mu(B) = 0$. On the other hand, for n large enough that $\epsilon_n < 1/(1+h)$, we have $B^{\epsilon_n} = B$. Hence, for large enough n and $G \in \mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$, we have $U(G)(B) \le \epsilon_n$. For such G, we construct a marked graph $\widetilde{G}$ which is obtained from G by removing all edges which have their pair of marks x, x' with $\deg_{x,x'}(\mu) = 0$. Moreover, if a vertex v in G has mark θ with $\Pi_\theta(\mu) = 0$, we change its mark to θ_0 in $\widetilde{G}$, with the θ_0 defined above. Note that $\widetilde{G} \in \mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)},\vec{\widetilde{u}}^{(n)}}$. Furthermore, since

$U(G)(B) \leq \epsilon_n$, the number of vertices v in G such that $(G, v)_h \equiv (\widetilde{G}, v)_h$ is at least $n(1-\epsilon_n)$. Consequently, using Lemma 4.1, when n is large enough that $\epsilon_n < \epsilon$, we have $d_{\text{LP}}(G, \widetilde{G}) \leq \max\{1/(1+h), \epsilon_n\} < \epsilon$. This means that $\widetilde{G} \in \mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)}, \vec{\widetilde{u}}^{(n)}}(\mu, \epsilon_n + \epsilon) \subset \mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)}, \vec{\widetilde{u}}^{(n)}}(\mu, 2\epsilon)$.

Motivated by this discussion, for n large enough, we have

$$
\begin{aligned}
|\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon_n)| \leq |\mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)}, \vec{\widetilde{u}}^{(n)}}(\mu, 2\epsilon)| & \\
\times \prod_{\substack{x \leq x' \in \Xi \\ \deg_{x,x'}(\mu) = 0}} & |\mathcal{G}_{n, m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')}| \times 2^{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')} \\
\times \prod_{\theta \in \Theta} & \binom{n}{|u^{(n)}(\theta) - \widetilde{u}^{(n)}(\theta)|}.
\end{aligned}
\tag{C.2}
$$

Here we have assumed that, since Ξ is finite, it is an ordered set. For $x \leq x' \in \Xi$ with $\deg_{x,x'}(\mu) = 0$, using Lemma 3.5, we have

$$
\begin{aligned}
\log\left(|\mathcal{G}_{n, m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')}| \times 2^{m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')}\right) \leq (m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')) \log n & \\
+ ns\left(\frac{2(m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x'))}{n}\right) & \\
+ (m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')) \log 2. &
\end{aligned}
$$

Note that, for all $x, x' \in \Xi$, $\frac{1}{n}(m^{(n)}(x,x') - \widetilde{m}^{(n)}(x,x')) \to 0$. Also, for all $\theta \in \Theta$, $\frac{1}{n}|u^{(n)}(\theta) - \widetilde{u}^{(n)}(\theta)| \to 0$. Additionally, $s(y) \to 0$ as $y \to 0$. Using these in (C.2) and simplifying, we get

$$
\begin{aligned}
\limsup_{n \to \infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon_n)| - \|\vec{m}^{(n)}\|_1 \log n}{n} &\leq \limsup_{n \to \infty} \frac{\log |\mathcal{G}^{(n)}_{\vec{\widetilde{m}}^{(n)}, \vec{\widetilde{u}}^{(n)}}(\mu, 2\epsilon)| - \|\vec{\widetilde{m}}^{(n)}\|_1 \log n}{n} \\
&\leq \overline{\Sigma}_{\vec{\deg}(\mu), \vec{\Pi}(\mu)}(\mu, 2\epsilon)|_{\vec{\widetilde{m}}^{(n)}, \vec{\widetilde{u}}^{(n)}},
\end{aligned}
$$

where the last inequality employs the fact that, by construction, $\vec{\widetilde{m}}^{(n)}$ and $\vec{\widetilde{u}}^{(n)}$ are adapted to $(\vec{\deg}(\mu), \vec{\Pi}(\mu))$. The above inequality holds for all $\epsilon > 0$. Therefore, from Theorem 3.2, as $\epsilon \to 0$, the right hand side converges to $\Sigma(\mu)$. This completes the proof for this case.

Case 2: $\bar{d}_{x,x'} > \deg_{x,x'}(\mu)$ for some $x, x' \in \Xi$. Let $\bar{d} := \sum_{x,x' \in \Xi} \bar{d}_{x,x'}$. Note that $\bar{d} > \deg(\mu)$. First, assume that $\bar{d} = \infty$. Observe that

$$
\begin{aligned}
|\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}(\mu, \epsilon)| &\leq |\mathcal{G}^{(n)}_{\vec{m}^{(n)}, \vec{u}^{(n)}}| \\
&\leq |\Theta|^n \prod_{x \leq x' \in \Xi} |\mathcal{G}_{n, m^{(n)}(x,x')}| 2^{m^{(n)}(x,x')}.
\end{aligned}
$$

Using Lemma 3.5,

$$\begin{aligned}|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon)| &\le n\log|\Theta| + \|\vec{m}^{(n)}\|_1 \log n + n\sum_{x\le x'\in\Xi}\left(s\left(\frac{2m^{(n)}(x,x')}{n}\right) + m^{(n)}(x,x')\log 2\right)\\ &= n\log|\Theta| + \|\vec{m}^{(n)}\|_1\log n + 2n\sum_{x\le x'\in\Xi} s\left(\frac{m^{(n)}(x,x')}{n}\right).\end{aligned} \tag{C.3}$$

Since we have assumed $\bar{d}=\infty$, there exist $\bar{x}\le\bar{x}'\in\Xi$ such that $\bar{d}_{\bar{x},\bar{x}'}=\infty$. Therefore, (C.1a) and (C.1b) imply that $m^{(n)}(\bar{x},\bar{x}')/n\to\infty$. On the other hand, $s(y)\to-\infty$ as $y\to\infty$. Using these in (C.3), we get $\limsup_{n\to\infty} a_n(\epsilon_n) = -\infty$ which completes the proof. Therefore, it remains to consider the case $\bar{d}<\infty$.

Let $\tilde{\mu}\in\mathcal{P}(\bar{\mathcal{T}}_*)$ be the law of $[\mathsf{UM}(T),o]$ when $[T,o]$ has law μ, and let $m_n := \|\vec{m}^{(n)}\|_1$. From Lemma C.2, if $G\in\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)$, we have $\mathsf{UM}(G)\in\mathcal{G}_{n,m_n}(\tilde{\mu},\epsilon_n)$. Moreover, by finding an upper bound on the number of possible ways to mark vertices and edges for an unmarked graph in $\mathcal{G}_{n,m_n}$, we have

$$|\mathcal{G}^{(n)}_{\vec{m}^{(n)},\vec{u}^{(n)}}(\mu,\epsilon_n)| \le |\mathcal{G}_{n,m_n}(\tilde{\mu},\epsilon_n)| \times |\Theta|^n \times \frac{m_n!}{\prod_{x\le x'} m^{(n)}(x,x')!}\times 2^{m_n}. \tag{C.4}$$

Note that $m_n/n\to\bar{d}/2<\infty$, and $m^{(n)}(x,x')/n$ converges to $\bar{d}_{x,x'}/2$ when $x=x'$, and $\bar{d}_{x,x'}$ when $x\neq x'$. Hence,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\left(|\Theta|^n\times\frac{m_n!}{\prod_{x\le x'}m^{(n)}(x,x')!}\times 2^{m_n}\right) &= \log|\Theta| + \sum_{x<x'\in\Xi}\bar{d}_{x,x'}\log\frac{\bar{d}}{\bar{d}_{x,x'}}\\ &\quad+\sum_{x\in\Xi}\frac{\bar{d}_{x,x}}{2}\log\frac{2\bar{d}}{\bar{d}_{x,x}} =: \alpha.\end{aligned}$$

Note that, as $\bar{d}<\infty$, α is a bounded real number. Also, since $\epsilon_n\to 0$, for each $\epsilon>0$ fixed we have $\epsilon_n<\epsilon$ for n large enough. Putting these in (C.4), we get

$$\limsup_{n\to\infty} a_n(\epsilon_n) \le \alpha + \limsup_{n\to\infty}\frac{\log|\mathcal{G}_{n,m_n}(\tilde{\mu},\epsilon)| - m_n\log n}{n}. \tag{C.5}$$

Note that $\deg(\tilde{\mu}) = \deg(\mu)$. Moreover, $m_n/n\to\bar{d}>\deg(\mu)=\deg(\tilde{\mu})>0$. Furthermore, our notion of marked BC entropy reduces to the unmarked BC entropy of [BC15] when Θ and Ξ have cardinality one. Therefore, since $\bar{d}\neq\deg(\tilde{\mu})$, from part 3 of Theorem 3.1, (or equivalently from part 3 of Theorem 1.2 in [BC15]), the right hand side of (C.5) goes to $-\infty$ as $\epsilon\to 0$. Therefore, $\limsup_{n\to\infty}a_n(\epsilon_n)=-\infty$, which completes the proof. □

Proof of Lemma 4.3. Let μ_n and $\tilde{\mu}_n$ denote $U(G^{(n)})$ and $U((G^{(n)})^{\Delta_n})$ respectively. For an integer $k\ge 0$ and a marked rooted tree (T,i) with depth at most k, define

$$A^k_{(T,i)} := \{[G,o]\in\bar{\mathcal{G}}_* : (G,o)_k \equiv (T,i)\},$$

as in (2.4). From Lemma 2.1 in Section 2.4, in order to show $\tilde{\mu}_n \Rightarrow \mu$, it suffices to show that $\tilde{\mu}_n(A^k_{(T,i)}) \to \mu(A^k_{(T,i)})$ for all such k and (T,i). We will now do this. Fix some integer $k \geq 0$ throughout the rest of the discussion. For an integer Δ, define

$$B^\Delta := \{[G,o] \in \bar{\mathcal{G}}_* : \deg_G(j) > \Delta \text{ for some } j \text{ with distance at most } k+1 \text{ from } o\}.$$

With this, we have

$$\begin{aligned}
\lim_{\Delta\to\infty} \mu((B^\Delta)^c) &= \mu\left(\bigcup_{\Delta=1}^{\infty} (B^\Delta)^c\right) \\
&= \mu\Big(\deg_G(j) < \infty \text{ for all } j \text{ with distanceat most } k+1 \text{ from } o\Big) \\
&= 1,
\end{aligned} \tag{C.6}$$

where the last equality comes from the fact that all graphs in $\bar{\mathcal{G}}_*$ are locally finite. Next, define

$$\begin{aligned}
C_n := \{i \in V(G^{(n)}) : &\deg_{G^{(n)}}(j) \leq \Delta_n \text{ for all nodes} \\
&j \text{ in } G^{(n)} \text{ with distance at most } k+1 \text{ from } i\} \\
= \{i \in V(&G^{(n)}) : [G^{(n)}, i] \in (B^{\Delta_n})^c\}.
\end{aligned}$$

Now, since $V(G^{(n)}) = \{1, \dots, n\}$, we have

$$\begin{aligned}
\tilde{\mu}_n(A_{(T,i)}) &= \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\left[((G^{(n)})^{\Delta_n}, j)_k \equiv (T,i)\right] \\
&= \frac{1}{n}\sum_{j\in C_n} \mathbb{1}\left[((G^{(n)})^{\Delta_n}, j)_k \equiv (T,i)\right] + \frac{1}{n}\sum_{j\in C_n^c} \mathbb{1}\left[((G^{(n)})^{\Delta_n}, j)_k \equiv (T,i)\right] \\
&= \frac{1}{n}\sum_{j\in C_n} \mathbb{1}\left[(G^{(n)}, j)_k \equiv (T,i)\right] + \frac{1}{n}\sum_{j\in C_n^c} \mathbb{1}\left[((G^{(n)})^{\Delta_n}, j)_k \equiv (T,i)\right].
\end{aligned}$$

Comparing this to

$$\mu_n(A_{(T,i)}) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\left[(G^{(n)}, j)_k \equiv (T,i)\right],$$

we realize that

$$|\tilde{\mu}_n(A_{(T,i)}) - \mu_n(A_{(T,i)})| \leq \frac{1}{n}|C_n^c| = \mu_n(B^{\Delta_n}).$$

Now, fix an integer $\Delta > 0$. Since $\Delta_n \to \infty$, we have $\Delta < \Delta_n$ for n large enough. Moreover, as B^Δ is closed,

$$\begin{aligned}
\limsup_{n\to\infty} |\tilde{\mu}_n(A_{(T,i)}) - \mu_n(A_{(T,i)})| &\leq \limsup_{n\to\infty} \mu_n(B^\Delta) \\
&\leq \mu(B^\Delta).
\end{aligned}$$

This is true for all $\Delta > 0$; therefore, sending Δ to infinity and using (C.6), we have $|\tilde{\mu}_n(A_{(T,i)}) - \mu_n(A_{(T,i)})| \to 0$. On the other hand, we have assumed that $\mu_n \Rightarrow \mu$. Therefore, Lemma 2.1 implies that $\mu_n(A_{(T,i)}) \to \mu(A_{(T,i)})$. This means that $\tilde{\mu}_n(A_{(T,i)}) \to \mu(A_{(T,i)})$. Since this is true for all k and (T, i), Lemma 2.1 implies that $\tilde{\mu}_n \Rightarrow \mu$, which completes the proof. □

Proof of Lemma 4.4. In order to count $|\mathcal{A}_{k_n,\Delta_n}|$, note that, for $\Delta_n \geq 2$, a rooted graph of depth at most k_n and maximum degree at most Δ_n has at most

$$1 + \Delta_n + \Delta_n^2 + \cdots + \Delta_n^{k_n} \leq \Delta_n^{k_n+1},$$

many vertices, each of which has $|\Theta|$ many choices for the vertex mark. On the other hand, such a graph can have at most $\Delta_n^{2(k_n+1)}$ many edges, each of which can be present or not, and, if present, has $|\Xi|^2$ many choices for the edge mark. Consequently,

$$|\mathcal{A}_{k_n,\Delta_n}| \leq (1 + |\Xi|^2)^{\Delta_n^{2(k_n+1)}} |\Theta|^{\Delta_n^{k_n+1}}.$$

Therefore,

$$\begin{aligned}
\log |\mathcal{A}_{k_n,\Delta_n}| &\leq \Delta_n^{2(1+k_n)} \log(1 + |\Xi|^2) + \Delta_n^{1+k_n} \log |\Theta| \\
&\leq \Delta_n^{2(1+k_n)} \log(|\Theta|(1 + |\Xi|^2)) \\
&\leq \Delta_n^{4k_n} \log(|\Theta|(1 + |\Xi|^2)),
\end{aligned}$$

where the last inequality holds for n large enough that $k_n \geq 1$. Note that in order to show $|\mathcal{A}_{k_n,\Delta}| = o(n/\log n)$, it suffices to show that $\log |\mathcal{A}_{k_n,\Delta_n}| - \log(n/\log n) \to -\infty$. Motivated by the above inequality, we observe that this is satisfied if $\Delta_n^{4k_n} = O(\sqrt{\log n})$. Suppose now that $\Delta_n \leq \log \log n$ and $k_n \leq \sqrt{\log \log n}$. For n large enough, we have

$$\begin{aligned}
\log(\Delta_n^{4k_n}) &\leq 4\sqrt{\log \log n} \log \log \log n \\
&\leq \frac{1}{2} \sqrt{\log \log n} \sqrt{\log \log n} \\
&= \frac{1}{2} \log \log n.
\end{aligned}$$

This means that for n large enough we have $\Delta_n^{4k_n} \leq \sqrt{\log n}$. This completes the proof. □

Proof of Lemma 4.5. For $\Delta > 0$, define $B_\Delta \subset \bar{\mathcal{G}}_*$ as

$$B_\Delta := \{[G, o] \in \bar{\mathcal{G}}_* : \deg_G(o) \leq \Delta \text{ and } \deg_G(i) \leq \Delta \text{ for all } i \sim_G o\}.$$

Since all graphs in $\bar{\mathcal{G}}_*$ are locally finite, we have $\mu(B_\Delta) \to 1$ as $\Delta \to \infty$. On the other hand,

$$\frac{|R_n|}{n} = U(G^{(n)})(B^c_{\Delta_n}).$$

Since $\Delta_n \to \infty$, we have $B_\Delta \subseteq B_{\Delta_n}$ for n large enough, for any value of Δ. Moreover, B_Δ is both open and closed. Therefore,

$$\frac{|R_n|}{n} = U(G^{(n)})(B^c_{\Delta_n}) \leq U(G^{(n)})(B^c_\Delta) \to \mu(B^c_\Delta).$$

But this is true for all Δ, and $\mu(B_\Delta) \to 1$ as $\Delta \to \infty$. Consequently, $|R_n|/n \to 0$, and the proof is complete. □

Appendix D

Proofs for Chapter 5

D.1 Proof of Lemma 5.1

Throughout this section, we treat $0 \log 0$ as equal to 0. Consider the first part of Lemma 5.1. Since $a_n/n \to a > 0$ as $n \to \infty$, using Stirling's approximation we have $\log a_n! = a_n \log a_n - a_n + o(n)$. Similarly, from the assumption that $b_i^n/n \to b_i \geq 0$ as $n \to \infty$ for $1 \leq i \leq k$, we have $\log b_i^n! = b_i^n \log b_i^n - b_i^n + o(n)$, which holds irrespective of whether $b_i > 0$ or $b_i = 0$. Hence we have

$$\log \binom{a_n}{\{b_i^n\}_{1 \leq i \leq k}} = a_n \log a_n - a_n - \sum_{i=1}^k b_i^n \log b_i^n + \sum_{i=1}^k b_i^n + o(n)$$
$$= a_n \log \frac{a_n}{n} - \sum_{i=1}^k b_i^n \log \frac{b_i^n}{n} + o(n),$$

where we have used $a_n = \sum_{i=1}^k b_i^n$. This gives

$$\lim_{n \to \infty} \frac{1}{n} \log \binom{a_n}{\{b_i^n\}_{1 \leq i \leq k}} = a \log a - \sum_{i=1}^k b_i \log b_i$$
$$= aH\left(\left\{ \frac{b_i}{a} \right\}_{1 \leq i \leq k} \right).$$

Next, consider the second part of Lemma 5.1. Since $a_n / \binom{n}{2} \to 1$ as $n \to \infty$, using Stirling's approximation we have $\log a_n! = a_n \log a_n - a_n + o(n)$. As noted earlier, since $b_i^n/n \to b_i \geq 0$ as $n \to \infty$ for $1 \leq i \leq k$, we have $\log b_i^n! = b_i^n \log b_i^n - b_i^n + o(n)$, which holds irrespective of whether $b_i > 0$ or $b_i = 0$. Moreover, with $b_n := \sum_{i=1}^k b_i^n$, we have $\log(a_n - b_n)! = (a_n - b_n)\log(a_n - b_n) - (a_n - b_n) + o(n)$. Therefore, we have

$$\log \binom{a_n}{\{b_i^n\}_{1 \leq i \leq k}} = a_n \log a_n - a_n - \sum_{i=1}^k b_i^n \log b_i^n + \sum_{i=1}^k b_i^n - (a_n - b_n)\log(a_n - b_n)$$

$$+ (a_n - b_n) + o(n)$$
$$= a_n \log \frac{a_n}{n} - \sum_{i=1}^{k} b_i^n \log \frac{b_i^n}{n} - (a_n - b_n) \log \frac{a_n - b_n}{n} + o(n),$$

where we have used $a_n = b_n + (a_n - b_n)$. This gives

$$\frac{1}{n} \log \binom{a_n}{\{b_i^n\}_{1\le i\le k}} = \frac{a_n}{n} \log \frac{a_n}{n} - \sum_{i=1}^{k} \frac{b_i^n}{n} \log \frac{b_i^n}{n} - \frac{a_n - b_n}{n} \log \frac{a_n - b_n}{n} + o(1)$$
$$= -\frac{a_n}{n} \log \left(1 - \frac{b_n}{a_n}\right) + \frac{b_n}{n} \log \frac{a_n - b_n}{\frac{n^2}{2}} + \frac{b_n}{n} \log \frac{n}{2} - \sum_{i=1}^{k} \frac{b_i^n}{n} \log \frac{b_i^n}{n} + o(1). \tag{D.1}$$

Since $b_n/a_n \to 0$ as $n \to \infty$, we write $\log(1 - b_n/a_n) = -b_n/a_n + O(b_n^2/a_n^2)$. Consequently, we have

$$-\frac{a_n}{n} \log \left(1 - \frac{b_n}{a_n}\right) = \frac{b_n}{n} + O\left(\frac{b_n^2}{na_n}\right),$$

and since $b_n^2/(na_n) \to 0$ we have

$$\lim_{n\to\infty} -\frac{a_n}{n} \log \left(1 - \frac{b_n}{a_n}\right) = b. \tag{D.2}$$

Further, since $(a_n - b_n)/(n^2/2) \to 1$ we have

$$\lim_{n\to\infty} \frac{b_n}{n} \log \frac{a_n - b_n}{\frac{n^2}{2}} = 0. \tag{D.3}$$

Using (D.2) and (D.3) in (D.1), we get

$$\lim_{n\to\infty} \frac{\log \binom{a_n}{\{b_i^n\}_{1\le i\le k}} - b_n \log n}{n} = b - b\log 2 - \sum_{i=1}^{k} b_i \log b_i$$
$$= \sum_{i=1}^{k} s(2b_i),$$

which completes the proof.

D.2 Proof of Lemma 5.3

The assumptions of the lemma imply that $b_n/n \to d/2 > 0$ and, in particular, $b_n \to \infty$ as $n \to \infty$. Therefore, Theorem 4.6 in [McK85] implies that

$$\lim_{n\to\infty} \frac{|\mathcal{G}_{\vec{a}^{(n)}}^{(n)}|}{\alpha_n \frac{(b_n - 1)!!}{\prod_{i=1}^n a^{(n)}(i)!}} = 1,$$

where

$$\alpha_n := \exp\left(-\lambda_n - \lambda_n^2\right), \qquad \lambda_n := \frac{1}{2b_n}\sum_{i=1}^n a^{(n)}(i)(a^{(n)}(i)-1),$$

and

$$(b_n-1)!! := (b_n-1)\times(b_n-3)\times\cdots\times 3\times 1 = \frac{b_n!}{2^{b_n/2}(b_n/2)!}.$$

Under the assumptions of the lemma, we have $b_n/n \to d/2$ as $n\to\infty$. Therefore, using Stirling's approximation, we have $\log(b_n-1)!! = \frac{b_n}{2}\log n - ns(d) + o(n)$. Moreover, since $c_k(\vec{a}^{(n)})/n \to \mathbb{P}(Y=k)$ as $n\to\infty$ for all $0\le k\le\Delta$, we have

$$\frac{1}{n}\log\prod_{i=1}^n a^{(n)}(i)! = \frac{1}{n}\sum_{k=0}^{\Delta} c_k(\vec{a}^{(n)})\log k! = \mathbb{E}\left[\log Y!\right] + o(1).$$

On the other hand, we have

$$\begin{aligned}\lim_{n\to\infty}\lambda_n &= \lim_{n\to\infty}\frac{1}{2b_n/n}\frac{1}{n}\sum_{i=1}^n a^{(n)}(i)(a^{(n)}(i)-1)\\ &= \frac{1}{d}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{\Delta} c_k(\vec{a}^{(n)})k(k-1)\\ &= \frac{1}{d}\mathbb{E}\left[Y(Y-1)\right] =: \lambda.\end{aligned}$$

This implies that, as $n\to\infty$, $\alpha_n \to \alpha := \exp(-\lambda-\lambda^2) > 0$. Therefore, $\frac{1}{n}\log\alpha_n \to 0$ as $n\to\infty$. Putting these together, we get the desired result.

D.3 Asymptotic behavior of the entropy of the configuration model

Here, we prove (5.14a)–(5.14c). Let X be a random variable with law $\vec{r}$, and let X_1 and X_2 be defined as in (5.13). Let $\Gamma = (\Gamma_1, \Gamma_2)$ and $Q = (Q_1, Q_2)$ denote random variables with laws $\vec{\gamma}$ and $\vec{q}$, respectively. Let $\beta_1 := \mathbb{P}(\Gamma_1 \neq \circ_1)$ and let $\tilde{\Gamma}_1$ be a random variable on Ξ_1 with the law of Γ_1 conditioned on $\Gamma_1 \neq \circ_1$.

As in Section 5.3.2, we let $\mathcal{D}^{(n)}$ denote the set of degree sequences $\vec{d} = (d(1),\dots,d(n))$ with entries bounded by Δ such that $c_k(\vec{d}) = c_k(\vec{d}^{(n)})$ for all $0\le k\le\Delta$. Let $F^{(n)}_{1,2}$ be a simple unmarked graph chosen uniformly at random from the set $\cup_{\vec{d}\in\mathcal{D}^{(n)}}\mathcal{G}^{(n)}_{\vec{d}}$, where we recall that $\mathcal{G}^{(n)}_{\vec{d}}$ denotes the set of simple unmarked graphs G on the vertex set $[n]$ such that $\deg_G(i) = d(i)$ for $1\le i\le n$. By definition, $G^{(n)}_{1,2} \sim \mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$ is obtained from $F^{(n)}_{1,2}$ by adding independent edge and vertex marks according to the laws of $\vec{\gamma}$ and $\vec{q}$ respectively.

If we first create $G_{1,2}^{(n)}$ from $F_{1,2}^{(n)}$, and then drop the edges with the first domain mark $\circ_1$, if $F_1^{(n)}$ denotes the unmarked version of the resulting marked graph, then $F_1^{(n)}$ is effectively obtained from $F_{1,2}^{(n)}$ by independently removing each edge with probability $1 - \beta_1$. Also, the corresponding first domain marked graph, i.e. $G_1^{(n)}$, obtained from $G_{1,2}^{(n)}$ in this way is effectively obtained from $F_1^{(n)}$ by adding independent vertex and edge marks to $F_1^{(n)}$ with the laws of Q_1 and $\tilde{\Gamma}_1$, respectively. With this viewpoint, we may consider $G_{1,2}^{(n)}$, $F_{1,2}^{(n)}$, $G_1^{(n)}$ and $F_1^{(n)}$ as being defined on a joint probability space.

As in Section 5.3.2, we let $\mathcal{W}^{(n)}$ denote the set of graphs $H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$ such that: (i) $\overrightarrow{\mathrm{dg}}_{H_{1,2}^{(n)}} \in \mathcal{D}^{(n)}$, (ii) $\vec{m}_{H_{1,2}^{(n)}} \in \mathcal{M}^{(n)}$, (iii) $\vec{u}_{H_{1,2}^{(n)}} \in \mathcal{U}^{(n)}$, (iv) for all $0 \le l \le k \le \Delta$, recalling the notation in (5.2), we have

$$|c_{k,l}(\overrightarrow{\mathrm{dg}}_{H_{1,2}^{(n)}}, \overrightarrow{\mathrm{dg}}_{H_1^{(n)}}) - n\mathbb{P}\left(X = k, X_1 = l\right)| \le n^{2/3},$$

and (v) for all $0 \le l \le k \le \Delta$ we have

$$|c_{k,l}(\overrightarrow{\mathrm{dg}}_{H_{1,2}^{(n)}}, \overrightarrow{\mathrm{dg}}_{H_2^{(n)}}) - n\mathbb{P}\left(X = k, X_2 = l\right)| \le n^{2/3}.$$

Here, as in Section 5.3.2, $\mathcal{M}^{(n)}$ denotes the set of mark count vectors $\vec{m}$ such that $\sum_{x \in \Xi_{1,2}} m(x) = m_n$ and $\sum_{x\in\Xi_{1,2}} |m(x) - m_n\gamma_x| \le n^{2/3}$, where we recall that $m_n := (\sum_{i=1}^n d^{(n)}(i))/2$, while, as in Section 5.3.2, $\mathcal{U}^{(n)}$ denotes the set of vertex mark count vectors $\vec{u}$ such that $\sum_{\theta\in\Theta_{1,2}} |u(\theta) - nq_\theta| \le n^{2/3}$.

We can now prove the following lemma.

Lemma D.1. *If $G_{1,2}^{(n)} \sim \mathcal{G}(n; \vec{d}^{(n)}, \vec{\gamma}, \vec{q}, \vec{r})$, we have $\mathbb{P}(G_{1,2}^{(n)} \notin \mathcal{W}^{(n)}) \le \kappa n^{-1/3}$ for some constant $\kappa > 0$.*

Proof. Condition (i) in the definition of $\mathcal{W}^{(n)}$ holds for every realization of $G_{1,2}^{(n)}$. Chebyshev's inequality implies that conditions (ii) and (iii) hold with probability at least $1 - \kappa_1 n^{-1/3}$, for some $\kappa_1 > 0$. To show (iv), fix $0 \le l \le k \le \Delta$ and, for $1 \le i \le n$, let Y_i be the indicator of the event that $\deg_{G_{1,2}^{(n)}}(i) = k$ and $\deg_{G_1^{(n)}}(i) = l$. With $Y := \sum_{i=1}^n Y_i$, we have $c_{k,l}(\overrightarrow{\mathrm{dg}}_{G_{1,2}^{(n)}}, \overrightarrow{\mathrm{dg}}_{G_1^{(n)}}) = Y$. Note that an edge of $G_{1,2}^{(n)}$ exists in $G_1^{(n)}$ if its mark is not of the form $(\circ_1, x_2)$, which happens with probability β_1. Therefore,

$$\mathbb{E}\left[Y_i | F_{1,2}^{(n)}\right] = \mathbb{1}\left[\deg_{F_{1,2}^{(n)}}(i) = k\right] \binom{\deg_{F_{1,2}^{(n)}}(i)}{l} \beta_1^l (1-\beta_1)^{k-l}.$$

Consequently,

$$\mathbb{E}\left[Y | F_{1,2}^{(n)}\right] = c_k(\vec{d}^{(n)}) \binom{k}{l} \beta_1^l (1-\beta_1)^{k-l}.$$

Since this is a constant, it is also equal to $\mathbb{E}[Y]$. Now, if $s_{k,l} := \mathbb{P}(X = k, X_1 = l)$, we have $s_{k,l} = r_k \binom{k}{l} \beta_1^l (1-\beta_1)^{k-l}$. Hence the assumption in (5.10) implies that

$$|\mathbb{E}[Y] - n s_{k,l}| \le K n^{1/2} \binom{k}{l} \beta_1^l (1-\beta_1)^{k-l}. \tag{D.4}$$

Furthermore, since edge marks are chosen independently conditioned on $F_{1,2}^{(n)}$, if i and j are nonadjacent vertices in $F_{1,2}^{(n)}$, then Y_i are Y_j are conditionally independent, conditioned on $F_{1,2}^{(n)}$. As a result, if $\mathcal{I}$ denotes the set of (i,j) with $1 \le i \ne j \le n$ such that i and j are not adjacent in $F_{1,2}^{(n)}$, we have

$$\begin{aligned}
\mathbb{E}\left[Y^2 | F_{1,2}^{(n)}\right] &= \sum_{i=1}^n \mathbb{E}\left[Y_i^2 | F_{1,2}^{(n)}\right] + \sum_{1 \le i \ne j \le n} \mathbb{E}\left[Y_i Y_j | F_{1,2}^{(n)}\right] \\
&\le n + \sum_{(i,j)\notin \mathcal{I}} \mathbb{E}\left[Y_i Y_j | F_{1,2}^{(n)}\right] + \sum_{(i,j)\in \mathcal{I}} \mathbb{E}\left[Y_i Y_j | F_{1,2}^{(n)}\right] \\
&\le n + 2m_n + \sum_{(i,j)\in \mathcal{I}} \mathbb{E}\left[Y_i Y_j | F_{1,2}^{(n)}\right] \\
&\stackrel{(a)}{=} n + 2m_n + \sum_{(i,j)\in \mathcal{I}} \mathbb{E}\left[Y_i | F_{1,2}^{(n)}\right] \mathbb{E}\left[Y_j | F_{1,2}^{(n)}\right] \\
&\le n + 2m_n + \sum_{1 \le i \ne j \le n} \mathbb{E}\left[Y_i | F_{1,2}^{(n)}\right] \mathbb{E}\left[Y_j | F_{1,2}^{(n)}\right] \\
&\le n + 2m_n + \mathbb{E}\left[Y | F_{1,2}^{(n)}\right]^2,
\end{aligned}$$

where (a) uses the fact that, conditioned on $F_{1,2}^{(n)}$, the random variables Y_i and Y_j are conditionally independent for $(i,j) \in \mathcal{I}$. From (5.10), we have $|m_n - n d_{1,2}^{\mathrm{CM}}/2| \le \kappa_2 K n^{1/2}$, where $\kappa_2 := (\Delta+1)/2$ and $d_{1,2}^{\mathrm{CM}} := \deg(\mu_{1,2}^{\mathrm{CM}}) = \sum_{k=0}^{\Delta} k r_k$. As a consequence of the above discussion, we have $\mathrm{Var}(Y|F_{1,2}^{(n)}) \le \kappa_3 n$ for some $\kappa_3 > 0$. On the other hand, as we saw above, $\mathbb{E}\left[Y|F_{1,2}^{(n)}\right] = \mathbb{E}[Y]$. Therefore, using the law of total variance, we have $\mathrm{Var}(Y) \le \kappa_3 n$. This, together with (D.4) and Chebyshev's inequality, implies that the condition (iv) holds with probability at least $1 - \kappa_4 n^{-1/3}$, for some $\kappa_4 > 0$. Similarly, the same statement holds for condition (v). □

Let $B_{1,2}^{(n)}$ be the set of pairs of degree sequences $\vec{d}$ and $\vec{\delta}$ with n elements bounded by Δ such that for all $0 \le k,l \le \Delta$, $|c_{k,l}(\vec{d},\vec{\delta}) - n\mathbb{P}(X_1 = k, X - X_1 = l)| \le n^{2/3}$. Moreover, let $B_1^{(n)}$ be the set of $\vec{d}$ such that for some $\vec{\delta}$, we have $(\vec{d},\vec{\delta}) \in B_{1,2}^{(n)}$. For $\vec{d} \in B_1^{(n)}$, let $B_{2|1}^{(n)}(\vec{d})$ be the set of degree sequences $\vec{\delta}$ such that $(\vec{d},\vec{\delta}) \in B_{1,2}^{(n)}$.

In order to show (5.14a), note that since $G_{1,2}^{(n)}$ is formed by adding independent vertex and edge marks to $F_{1,2}^{(n)}$, we have

$$H(G_{1,2}^{(n)}) = \log |\mathcal{D}^{(n)}| + \log |\mathcal{G}_{\vec{d}^{(n)}}^{(n)}| + m_n H(\Gamma) + nH(Q).$$

From (5.10), we have $|m_n - nd_{1,2}^{\text{CM}}/2| \leq \frac{(\Delta+1)K}{2} n^{1/2}$. Moreover, we have $\mathbb{E}[X] > 0$. Consequently, using Lemma 5.3 and the fact that $\frac{1}{n} \log |\mathcal{D}^{(n)}| \to H(X)$, we get (5.14a).

We now turn to showing (5.14b). Since the expected number of the edges in $F_1^{(n)}$ is $nd_1^{\text{CM}}/2$, we have

$$H(G_1^{(n)}) = H(F_1^{(n)}) + n\frac{d_1^{\text{CM}}}{2} H(\Gamma_1 | \Gamma_1 \neq \circ_1) + nH(Q_1). \tag{D.5}$$

With this, we focus on $H(F_1^{(n)})$. With E_n being the indicator of the event that $G_{1,2}^{(n)} \notin \mathcal{W}^{(n)}$, we have

$$\begin{aligned} H(F_1^{(n)}) \leq H(F_1^{(n)}, E_n) &\leq 1 + H(F_1^{(n)} | E_n) \\ &= 1 + H(F_1^{(n)} | E_n = 0)\mathbb{P}(E_n = 0) \\ &\quad + H(F_1^{(n)} | E_n = 1)\mathbb{P}(E_n = 1). \end{aligned}$$

Note that $F_1^{(n)}$ is obtained from $F_{1,2}^{(n)}$ by removing some edges. Hence, we may write

$$\begin{aligned} H(F_1^{(n)} | E_n = 1) \leq H(F_1^{(n)}) &\leq \log |\mathcal{D}^{(n)}| + \log |\mathcal{G}_{\vec{d}^{(n)}}^{(n)}| + m_n \log 2 \\ &\leq H(G_{1,2}^{(n)}) + m_n \log 2 \\ &\leq \kappa' n \log n, \end{aligned} \tag{D.6}$$

where in the last line, $\kappa' > 0$ is obtained from (5.14a). Putting this together with Lemma D.1, we have

$$H(F_1^{(n)} | E_n = 1)\mathbb{P}(E_n = 1) \leq \kappa' n \log n \kappa n^{-1/3}. \tag{D.7}$$

Note that the right hand side of (D.7) above is $o(n)$. On the other hand, by the definition of $\mathcal{W}^{(n)}$, if $E = 0$, we have $\overrightarrow{\text{dg}}_{F_1^{(n)}} \in B_1^{(n)}$. Therefore, $H(F_1^{(n)} | E_n = 0) \leq \log |B_1^{(n)}| + \max_{\vec{d} \in B_1^{(n)}} \log |\mathcal{G}_{\vec{d}}^{(n)}|$. The assumption $r_0 < 1$ together with (5.8) imply that $d_1^{\text{CM}} > 0$. Additionally note that, by definition, for $\vec{d} \in B_1^{(n)}$, we have $|c_k(\vec{d}) - n\mathbb{P}(X_1 = k)| \leq n^{2/3}$ for all $0 \leq k \leq \Delta$. Thereby, we have

$$\limsup_{n \to \infty} \frac{\log |B_1^{(n)}|}{n} \leq H(X_1).$$

Putting the above together with Lemma 5.3 and (D.7), we have

$$\begin{aligned}\limsup_{n\to\infty} \frac{H(F_1^{(n)}) - n\frac{d_1^{\text{CM}}}{2}\log n}{n} &\le \limsup_{n\to\infty} \frac{\log|B_1^{(n)}| + \max_{\vec{d}\in B_1^{(n)}} \log|\mathcal{G}_{\vec{d}}^{(n)}| - n\frac{d_1^{\text{CM}}}{2}\log n}{n} \\ &\le \limsup_{n\to\infty} \frac{\log|B_1^{(n)}|}{n} + \max_{\vec{d}\in B_1^{(n)}} \frac{\log|\mathcal{G}_{\vec{d}}^{(n)}| - \frac{\sum_{i=1}^n d(i)}{2}\log n}{n} \\ &\quad + \max_{\vec{d}\in B_1^{(n)}} \frac{\frac{\sum_{i=1}^n d(i)}{2}\log n - n\frac{d_1^{\text{CM}}}{2}\log n}{n} \\ &\le -s(d_1^{\text{CM}}) + H(X_1) - \mathbb{E}\left[\log X_1!\right],\end{aligned} \tag{D.8}$$

where in the last line, have used the fact that due to the definition of $B_1^{(n)}$, for $\vec{d} \in B_1^{(n)}$, we have $|\sum_{i=1}^n d(i) - d_1^{\text{CM}}| \le \Delta n^{2/3} = o(n/\log n)$. Now, let $\tilde{F}_1^{(n)}$ be the unmarked graph consisting of the edges removed from $F_{1,2}^{(n)}$ to obtain $F_1^{(n)}$, and note that

$$\begin{aligned}H(F_1^{(n)}) &= H(F_1^{(n)}, \tilde{F}_1^{(n)}) - H(\tilde{F}_1^{(n)}|F_1^{(n)}) \\ &= H(F_{1,2}^{(n)}) + m_n H(\beta_1) - H(\tilde{F}_1^{(n)}|F_1^{(n)}).\end{aligned} \tag{D.9}$$

Furthermore, conditioned on $E_n = 0$, we have $\overrightarrow{\text{dg}}_{\tilde{F}_1^{(n)}} \in B_{2|1}^{(n)}(\overrightarrow{\text{dg}}_{F_1^{(n)}})$. Moreover, the assumption (5.8), for $x_1 = \circ_1$, together with $r_0 < 1$, implies that $d_{1,2}^{\text{CM}} - d_1^{\text{CM}} > 0$. Hence, using a similar method to that used in proving (D.8), we have

$$\begin{aligned}\limsup_{n\to\infty} \frac{H(\tilde{F}_1^{(n)}|F_1^{(n)}) - n\frac{d_{1,2}^{\text{CM}} - d_1^{\text{CM}}}{2}\log n}{n} &\le -s(d_{1,2}^{\text{CM}} - d_1^{\text{CM}}) \\ &\quad + H(X - X_1|X_1) - \mathbb{E}\left[\log(X - X_1)!\right].\end{aligned} \tag{D.10}$$

To see this, with E_n as defined previously, we may write

$$H(\tilde{F}_1^{(n)}|F_1^{(n)}) \le 1 + H(\tilde{F}_1^{(n)}|F_1^{(n)}, E_n = 0)\mathbb{P}(E_n = 0) + H(\tilde{F}_1^{(n)}|F_1^{(n)}, E_n = 1)\mathbb{P}(E_n = 1).$$

Since $\tilde{F}_1^{(n)}$ is obtained from $F_{1,2}^{(n)}$ by removing some edges, similar to (D.6), we have

$$H(\tilde{F}_1^{(n)}|F_1^{(n)}, E_n = 1)\mathbb{P}(E_n = 1) = o(n).$$

Moreover, conditioned on $E_n = 0$, we have $\overrightarrow{\text{dg}}_{\tilde{F}_1^{(n)}} \in B_{2|1}^{(n)}(\overrightarrow{\text{dg}}_{F_1^{(n)}})$. This implies that when $E_n = 0$, for any realization $f_1^{(n)}$ of $F_1^{(n)}$, we have

$$H(\tilde{F}_1^{(n)}|F_1^{(n)} = f_1^{(n)}, E_n = 0) \le \log|B_{2|1}^{(n)}(\overrightarrow{\text{dg}}_{f_1^{(n)}})| + \max_{\vec{\delta}\in B_{2|1}^{(n)}(\overrightarrow{\text{dg}}_{f_1^{(n)}})} \log|\mathcal{G}_{\vec{\delta}}^{(n)}|.$$

Note that, conditioned on $E_n = 0$, we have $\overrightarrow{\mathrm{dg}}_{f_1^{(n)}} \in B_1^{(n)}$. Hence, we have $\log |B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{f_1^{(n)}})| = nH(X - X_1|X_1) + o(n)$. Furthermore, using Lemma 5.3 and the fact that for $\vec{\delta} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{f_1^{(n)}})$, we have $|\sum_{i=1}^n \delta(i) - (d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}})/2| \le \Delta n^{2/3} = o(n/\log n)$, we have

$$\max_{\vec{\delta} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{f_1^{(n)}})} \log |\mathcal{G}_{\vec{\delta}}^{(n)}| = n(-s(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) - \mathbb{E}\left[\log(X - X_1)!\right]) + n\frac{d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}}{2} \log n + o(n).$$

Putting the above together, we arrive at (D.10).

On the other hand, using the definition of $F_{1,2}^{(n)}$, we have $H(F_{1,2}^{(n)}) = \log |\mathcal{D}^{(n)}| + \log |\mathcal{G}_{\vec{d}^{(n)}}^{(n)}|$. Employing Lemma 5.3 and using the assumption (5.10), we have

$$\lim_{n\to\infty} \frac{\log |\mathcal{G}_{\vec{d}^{(n)}}^{(n)}| - n\frac{d_{1,2}^{\mathrm{CM}}}{2} \log n}{n} = -s(d_{1,2}^{\mathrm{CM}}) - \mathbb{E}\left[\log X!\right].$$

Furthermore, we have $\frac{1}{n} \log |\mathcal{D}^{(n)}| \to H(X)$. Therefore, we have

$$\lim_{n\to\infty} \frac{H(F_{1,2}^{(n)}) - n\frac{d_{1,2}^{\mathrm{CM}}}{2} \log n}{n} = -s(d_{1,2}^{\mathrm{CM}}) + H(X) - \mathbb{E}\left[\log X!\right]. \tag{D.11}$$

Using (D.10) and (D.11) back in (D.9), followed by a simplification using Lemma 5.2, we get

$$\begin{aligned}
\liminf_{n\to\infty} \frac{H(F_1^{(n)}) - n\frac{d_1^{\mathrm{CM}}}{2}\log n}{n} &= \liminf_{n\to\infty} \frac{1}{n}\Bigg(H(F_{1,2}^{(n)}) - n\frac{d_{1,2}^{\mathrm{CM}}}{2}\log n + m_n H(\beta_1) \\
&\qquad - H(\tilde{F}_1^{(n)}|F_1^{(n)}) + n\frac{d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}}{2}\log n\Bigg) \\
&\geq \liminf_{n\to\infty} \frac{H(F_{1,2}^{(n)}) - n\frac{d_{1,2}^{\mathrm{CM}}}{2}\log n}{n} + \frac{d_{1,2}^{\mathrm{CM}}}{2}H(\beta_1) \\
&\qquad - \limsup_{n\to\infty} \frac{H(\tilde{F}_1^{(n)}|F_1^{(n)}) - n\frac{d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}}{2}\log n}{n} \\
&\geq -s(d_{1,2}^{\mathrm{CM}}) + H(X) - \mathbb{E}\left[\log X!\right] + \frac{d_{1,2}^{\mathrm{CM}}}{2}H(\beta_1) \\
&\qquad - \big(-s(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) + H(X - X_1|X_1) - \mathbb{E}\left[\log(X - X_1)!\right]\big) \\
&= H(X) + d_{1,2}^{\mathrm{CM}}H(\beta_1) - \mathbb{E}\left[\log\binom{X}{X_1}\right] - \mathbb{E}\left[\log X_1!\right] \\
&\qquad - \frac{d_{1,2}^{\mathrm{CM}}}{2}H(\beta_1) - s(d_{1,2}^{\mathrm{CM}}) + s(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) - H(X - X_1|X_1) \\
&\overset{(a)}{=} H(X_1, X - X_1) - H(X - X_1|X_1) - \mathbb{E}\left[\log X_1!\right] \\
&\qquad - s(d_{1,2}^{\mathrm{CM}}) + s(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) - \frac{d_{1,2}^{\mathrm{CM}}}{2}H(\beta_1) \\
&= H(X_1) - \mathbb{E}\left[\log X_1!\right] - s(d_{1,2}^{\mathrm{CM}}) + s(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) - \frac{d_{1,2}^{\mathrm{CM}}}{2}H(\beta_1),
\end{aligned} \tag{D.12}$$

where in (a), we have used Lemma 5.2. Since $\beta_1 = d_1^{\mathrm{CM}}/d_{1,2}^{\mathrm{CM}}$, we may write

$$\begin{aligned}
-s(d_{1,2}^{\mathrm{CM}}) + s(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) - \frac{d_{1,2}^{\mathrm{CM}}}{2}H(\beta_1) &= \frac{d_{1,2}^{\mathrm{CM}}}{2}\log d_{1,2}^{\mathrm{CM}} - \frac{d_{1,2}^{\mathrm{CM}}}{2} + \frac{d_{1,2}^{\mathrm{CM}}}{2} - \frac{d_1^{\mathrm{CM}}}{2} \\
&\quad - \frac{d_{1,2}^{\mathrm{CM}}}{2}\log(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) + \frac{d_1^{\mathrm{CM}}}{2}\log(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) \\
&\quad + \frac{d_1^{\mathrm{CM}}}{2}\log d_1^{\mathrm{CM}} - \frac{d_1^{\mathrm{CM}}}{2}\log d_{1,2}^{\mathrm{CM}} \\
&\quad + \frac{d_{1,2}^{\mathrm{CM}}}{2}\log(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) \\
&\quad - \frac{d_1^{\mathrm{CM}}}{2}\log(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) - \frac{d_{1,2}^{\mathrm{CM}}}{2}\log d_{1,2}^{\mathrm{CM}} \\
&\quad + \frac{d_1^{\mathrm{CM}}}{2}\log d_{1,2}^{\mathrm{CM}}
\end{aligned}$$

$$= -\frac{d_1^{\text{CM}}}{2} + \frac{d_1^{\text{CM}}}{2} \log d_1^{\text{CM}}$$
$$= -s(d_1^{\text{CM}}).$$

Substituting this into (D.12), we arrive at

$$\liminf \frac{H(F_1^{(n)}) - n\frac{d_1^{\text{CM}}}{2}\log n}{n} \geq -s(d_1^{\text{CM}}) + H(X_1) - \mathbb{E}\left[\log X_1!\right]. \tag{D.13}$$

This, together with (D.8) and (D.5), completes the proof of (5.14b). The proof of (5.14c) is similar.

D.4 Bounding $|S_2^{(n)}(H_1^{(n)})|$ for $H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$ in the Erdős–Rényi case

Note that for $H_{1,2}^{(n)} \in \mathcal{G}_{1,2}^{(n)}$ and $G_2^{(n)} \in \mathcal{G}_2^{(n)}$, if $H_1^{(n)} \oplus G_2^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)}$, we have $\vec{m}_{H_1^{(n)} \oplus G_2^{(n)}} \in \mathcal{M}^{(n)}$ and $\vec{u}_{H_1^{(n)} \oplus G_2^{(n)}} \in \mathcal{U}^{(n)}$. On the other hand, for fixed $\vec{m} \in \mathcal{M}^{(n)}$ and $\vec{u} \in \mathcal{U}^{(n)}$, the number of $G_2^{(n)}$ such that $\vec{m}_{H_1^{(n)} \oplus G_2^{(n)}} = \vec{m}$ and $\vec{u}_{H_1^{(n)} \oplus G_2^{(n)}} = \vec{u}$ is at most

$$A_2(\vec{m},\vec{u}) := \left(\prod_{x_1 \in \Xi_1} \binom{m(x_1)}{\{m(x_1,x_2)\}_{x_2 \in \Xi_2 \cup \{\circ_2\}}}\right) \times \binom{\binom{n}{2} - \sum_{x_1 \in \Xi_1} m(x_1)}{\{m(\circ_1, x_2)\}_{x_2 \in \Xi_2}}$$
$$\times \left(\prod_{\theta_1 \in \Theta_1} \binom{u(\theta_1)}{\{u(\theta_1,\theta_2)\}_{\theta_2 \in \Theta_2}}\right),$$

where we have used the notational conventions in (5.3) and (5.4). Consequently, we have

$$\begin{aligned}\max_{H_{1,2}^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)}} |S_2^{(n)}(H_1^{(n)})| &\leq |\mathcal{M}^{(n)}||\mathcal{U}^{(n)}| \max_{\substack{\vec{m} \in \mathcal{M}^{(n)} \\ \vec{u} \in \mathcal{U}^{(n)}}} A_2(\vec{m},\vec{u}) \\ &\leq (2n^{2/3}+1)^{(|\Xi_{1,2}|+|\Theta_{1,2}|)} \max_{\substack{\vec{m} \in \mathcal{M}^{(n)} \\ \vec{u} \in \mathcal{U}^{(n)}}} A_2(\vec{m},\vec{u}).\end{aligned} \tag{D.14}$$

Now, if $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$ are sequences in $\mathcal{M}^{(n)}$ and $\mathcal{U}^{(n)}$, respectively, then for all $x \in \Xi_{1,2}$ we have $m^{(n)}(x)/n \to p_x/2$. Furthermore, for all $x_1 \in \Xi_1$ and $\theta_1 \in \Theta_1$, we have $m^{(n)}(x_1)/n \to p_{x_1}/2$ and $u^{(n)}(\theta_1)/n \to q_{\theta_1}$. As a result, using Lemma 5.1, for any such sequences $\vec{m}^{(n)}$ and $\vec{u}^{(n)}$, with $Q = (Q_1, Q_2)$ having law $\vec{q}$, we have

$$\lim_{n\to\infty} \frac{\log A_2(\vec{m}^{(n)},\vec{u}^{(n)}) - (\sum_{x_2 \in \Xi_2} m^{(n)}(\circ_1, x_2))\log n}{n}$$
$$= \sum_{x_2 \in \Xi_2} s(p_{\circ_1,x_2}) + \sum_{x_1 \in \Xi_1} \frac{p_{x_1}}{2} H\left(\left\{\frac{p_{(x_1,x_2)}}{p_{x_1}}\right\}_{x_2 \in \Xi_2 \cup \{\circ_2\}}\right)$$

$$
\begin{aligned}
&+ \sum_{\theta_1 \in \Theta_1} q_{\theta_1} H\left(\left\{\frac{q_{\theta_1,\theta_2}}{q_{\theta_1}}\right\}_{\theta_2\in\Theta_2}\right)\\
&= H(Q_2|Q_1) + \sum_{x \in \Xi_{1,2}} s(p_x) - \sum_{x_1\in\Xi_1} s(p_{x_1})\\
&= \Sigma(\mu_2^{\mathrm{ER}}|\mu_1^{\mathrm{ER}}),
\end{aligned}
$$

where the second equality follows by rearranging the terms and using the definition of $s(.)$. Using the fact that $|m^{(n)}(\circ_1, x_2) - np_{\circ_1,x_2}/2| \le n^{2/3}$, we have

$$
\lim_{n\to\infty} \frac{\log A_2(\vec{m}^{(n)}, \vec{u}^{(n)}) - n\frac{d_{1,2}^{\mathrm{ER}} - d_1^{\mathrm{ER}}}{2}\log n}{n} = \Sigma(\mu_2^{\mathrm{ER}}|\mu_1^{\mathrm{ER}}).
$$

This together with (D.14) implies (5.21).

D.5 Bounding $|S_2^{(n)}(H_1^{(n)})|$ for $H_{1,2}^{(n)} \in \mathcal{W}^{(n)}$ in the configuration model

Here, we find an upper bound for $\max_{H_{1,2}^{(n)}\in\mathcal{W}^{(n)}} |S_2^{(n)}(H_1^{(n)})|$, where $\mathcal{W}^{(n)}$ is defined in Section 5.3.2, and use it to show (5.27). Take $H_{1,2}^{(n)} \in \mathcal{W}^{(n)}$ and assume $\hat{H}_2^{(n)} \in S_2^{(n)}(H_1^{(n)})$. With $\hat{H}_{1,2}^{(n)} := H_1^{(n)} \oplus \hat{H}_2^{(n)}$, let $\widetilde{H}_2^{(n)}$ be the subgraph of $\hat{H}_{1,2}^{(n)}$ consisting of the edges not present in $H_1^{(n)}$. Employing the notation of Appendix D.3, we have $\overrightarrow{\mathrm{dg}}_{\widetilde{H}_2^{(n)}} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})$, which follows from the definition of the set $\mathcal{W}^{(n)}$. Therefore, we can think of $\hat{H}_{1,2}^{(n)}$ as being constructed from $H_1^{(n)}$ by adding a graph to $H_1^{(n)}$ with degree sequence $\overrightarrow{\mathrm{dg}}_{\widetilde{H}_2^{(n)}} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})$, marking its edges, adding second domain marks to edges in $H_1^{(n)}$, and also adding second domain marks to vertices. Motivated by this, we have

$$
\begin{aligned}
\max_{H_{1,2}^{(n)}\in\mathcal{W}^{(n)}} \log |S_2^{(n)}(H_1^{(n)})| &\le \max_{H_{1,2}^{(n)}\in\mathcal{W}^{(n)}} \log |B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})| + \max_{H_{1,2}^{(n)}\in\mathcal{W}^{(n)}, \vec{\delta}\in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})} \log |\mathcal{G}_{\vec{\delta}}^{(n)}|\\
&+ \max_{\vec{m}\in\mathcal{M}^{(n)}} \log \binom{m_n - \sum_{x_1\in\Xi_1} m(x_1)}{\{m((\circ_1,x_2))\}_{x_2\in\Xi_2}} \prod_{x_1\in\Xi_1} \binom{m(x_1)}{\{m((x_1,x_2))\}_{x_2\in\Xi_2}}\\
&+ \max_{\vec{u}\in\mathcal{U}^{(n)}} \log \prod_{\theta_1\in\Theta_1} \binom{u(\theta_1)}{\{u((\theta_1,\theta_2))\}_{\theta_2\in\Theta_2}}.
\end{aligned}
\tag{D.15}
$$

We establish an upper bound for each term. The definition of $B_{2|1}^{(n)}$ implies that

$$
\lim_{n\to\infty} \frac{1}{n} \max_{H_{1,2}^{(n)}\in\mathcal{W}^{(n)}} \log |B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})| = H(X - X_1|X_1), \tag{D.16}
$$

where (X, X_1) are defined as in Section 5.3.2. Note that the assumption (5.8), for $x_1 = \circ_1$, together with $r_0 < 1$, implies that $d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}} > 0$. On the other hand, we have

$$\limsup_{n\to\infty} \max_{\substack{H_{1,2}^{(n)} \in \mathcal{W}^{(n)} \\ \vec{\delta} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})}} \frac{\log|\mathcal{G}_{\vec{\delta}}^{(n)}| - n\frac{d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}}{2}\log n}{n} \le \limsup_{n\to\infty} \max_{\substack{H_{1,2}^{(n)} \in \mathcal{W}^{(n)} \\ \vec{\delta} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})}} \frac{\log|\mathcal{G}_{\vec{\delta}}^{(n)}| - \frac{\sum_{i=1}^n \delta_i}{2}\log n}{n}$$
$$+ \limsup_{n\to\infty} \max_{\substack{H_{1,2}^{(n)} \in \mathcal{W}^{(n)} \\ \vec{\delta} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})}} \frac{1}{n}\left(\frac{\sum_{i=1}^n \delta_i}{2}\log n - n\frac{d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}}{2}\log n\right). \tag{D.17}$$

By definition, for $\vec{\delta} = (\delta_1, \dots, \delta_n) \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})$, we have

$$\begin{aligned}
\left|\left(\sum_{i=1}^n \delta_i\right) - n(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}})\right| &= \left|\left(\sum_{k=0}^{\Delta} k c_k(\vec{\delta})\right) - n\mathbb{E}\left[X - X_1\right]\right| \\
&\le \sum_{k=0}^{\Delta} k|c_k(\vec{\delta}) - n\mathbb{P}\left(X - X_1 = k\right)| \\
&\le \sum_{k=0}^{\Delta} k \sum_{j=0}^{\Delta} |c_{j,k}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}}, \vec{\delta}) - n\mathbb{P}\left(X_1 = j, X - X_1 = k\right)| \\
&\le \Delta^3 n^{2/3}.
\end{aligned}$$

This implies that the second term in the right hand side of (D.17) vanishes. Therefore, Lemma 5.3 implies that

$$\limsup_{n\to\infty} \max_{\substack{H_{1,2}^{(n)} \in \mathcal{W}^{(n)} \\ \vec{\delta} \in B_{2|1}^{(n)}(\overrightarrow{\mathrm{dg}}_{H_1^{(n)}})}} \frac{\log|\mathcal{G}_{\vec{\delta}}^{(n)}| - n\frac{d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}}{2}\log n}{n} \le -s(d_{1,2}^{\mathrm{CM}} - d_1^{\mathrm{CM}}) - \mathbb{E}\left[\log(X - X_1)!\right]. \tag{D.18}$$

Furthermore, if $\vec{m}^{(n)}$ is a sequence in $\mathcal{M}^{(n)}$, by definition we have $\sum_{x\in\Xi_{1,2}} |m^{(n)}(x) - m_n\gamma_x| \le n^{2/3}$. Therefore, we have

$$\lim_{n\to\infty} \frac{m_n - \sum_{x_1\in\Xi_1} m^{(n)}(x_1)}{n} = \frac{d_{1,2}^{\mathrm{CM}}}{2}\left(1 - \sum_{x_1\in\Xi_1}\gamma_{x_1}\right),$$

where γ_{x_1} for $x_1 \in \Xi_1$ is defined to be $\sum_{x_2\in\Xi_2\cup\{\circ_2\}} \gamma_{(x_1,x_2)}$. Similarly, for $x_2 \in \Xi_2$, we have

$$\lim_{n\to\infty} \frac{m^{(n)}((\circ_1, x_2))}{m_n - \sum_{x_1\in\Xi_1} m^{(n)}(x_1)} = \frac{\gamma_{(\circ_1,x_2)}}{1 - \sum_{x_1\in\Xi_1}\gamma_{x_1}} = \frac{\gamma_{(\circ_1,x_2)}}{\sum_{x_2'\in\Xi_2}\gamma_{(\circ_1,x_2')}}.$$

Consequently, using Lemma 5.1, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\binom{m_n-\sum_{x_1\in\Xi_1}m^{(n)}(x_1)}{\{m^{(n)}((\circ_1,x_2))\}_{x_2\in\Xi_2}} = \frac{d^{\text{CM}}_{1,2}}{2}\left(1-\sum_{x_1\in\Xi_1}\gamma_{x_1}\right)H\left(\left\{\frac{\gamma_{(\circ_1,x_2)}}{\sum_{x'_2\in\Xi_2}\gamma_{(\circ_1,x'_2)}}\right\}_{x_2\in\Xi_2}\right)$$
$$= \frac{d^{\text{CM}}_{1,2}}{2}\mathbb{P}\left(\Gamma_1=\circ_1\right)H(\Gamma_2|\Gamma_1=\circ_1). \tag{D.19}$$

Here, $\Gamma = (\Gamma_1, \Gamma_2)$ has law $\vec{\gamma}$. On the other hand, for $x_1 \in \Xi_1$ and $x_2 \in \Xi_2$, we have $m^{(n)}(x_1)/n \to \frac{d^{\text{CM}}_{1,2}}{2}\gamma_{x_1}$, and $\frac{m^{(n)}((x_1,x_2))}{m^{(n)}(x_1)} \to \frac{\gamma_{(x_1,x_2)}}{\gamma_{x_1}}$. Consequently, another use of Lemma 5.1 implies that for all $x_1 \in \Xi_1$, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\binom{m^{(n)}(x_1)}{\{m^{(n)}((x_1,x_2))\}_{x_2\in\Xi_2}} = \frac{d^{\text{CM}}_{1,2}}{2}\gamma_{x_1}H\left(\left\{\frac{\gamma_{(x_1,x_2)}}{\gamma_{x_1}}\right\}_{x_2\in\Xi_2}\right)$$
$$= \frac{d^{\text{CM}}_{1,2}}{2}\mathbb{P}\left(\Gamma_1=x_1\right)H(\Gamma_2|\Gamma_1=x_1). \tag{D.20}$$

Putting together (D.19) and (D.20), we realize that

$$\lim_{n\to\infty}\max_{\vec{m}\in\mathcal{M}^{(n)}}\log\binom{m_n-\sum_{x_1\in\Xi_1}m(x_1)}{\{m((\circ_1,x_2))\}_{x_2\in\Xi_2}}\prod_{x_1\in\Xi_1}\binom{m(x_1)}{\{m((x_1,x_2))\}_{x_2\in\Xi_2}}$$
$$= \frac{d^{\text{CM}}_{1,2}}{2}\Bigg(\mathbb{P}\left(\Gamma_1=\circ_1\right)H(\Gamma_2|\Gamma_1=\circ_1)$$
$$+\sum_{x_1\in\Xi_1}\mathbb{P}\left(\Gamma_1=x_1\right)H(\Gamma_2|\Gamma_1=x_1)\Bigg)$$
$$= \frac{d^{\text{CM}}_{1,2}}{2}H(\Gamma_2|\Gamma_1). \tag{D.21}$$

Using a similar technique, if $\vec{u}^{(n)}$ is a sequence in $\mathcal{U}^{(n)}$, for all $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$, we have $\frac{u^{(n)}(\theta_1)}{n} \to q_{\theta_1}$ and $\frac{u^{(n)}((\theta_1,\theta_2))}{u^{(n)}(\theta_1)} \to \frac{q_{(\theta_1,\theta_2)}}{q_{\theta_1}}$. Thereby, using Lemma 5.1, for all $\theta_1 \in \Theta_1$, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\binom{u^{(n)}(\theta_1)}{\{u^{(n)}((\theta_1,\theta_2))\}_{\theta_2\in\Theta_2\}}} = q_{\theta_1}H\left(\left\{\frac{q_{(\theta_1,\theta_2)}}{q_{\theta_1}}\right\}_{\theta_2\in\Theta_2}\right) = \mathbb{P}\left(Q_1=\theta_1\right)H(Q_2|Q_1=\theta_1),$$

where $Q = (Q_1, Q_2)$ has law $\vec{q}$. Consequently, we have

$$\lim_{n\to\infty}\frac{1}{n}\max_{\vec{u}\in\mathcal{U}^{(n)}}\log\prod_{\theta_1\in\Theta_1}\binom{u(\theta_1)}{\{u((\theta_1,\theta_2))\}_{\theta_2\in\Theta_2}} = \sum_{\theta_1\in\Theta_1}\mathbb{P}\left(Q_1=\theta_1\right)H(Q_2|Q_1=\theta_1) = H(Q_2|Q_1). \tag{D.22}$$

Putting (D.16), (D.18), (D.21), and (D.22) back into (D.15), we get

$$\lim_{n\to\infty} \frac{\max_{H^{(n)}_{1,2}\in\mathcal{W}^{(n)}} \log |S^{(n)}_2(H^{(n)}_1)| - n\frac{d^{\text{CM}}_{1,2}-d^{\text{CM}}_1}{2}\log n}{n} = -s(d^{\text{CM}}_{1,2} - d^{\text{CM}}_1) + H(X - X_1|X_1)$$
$$-\mathbb{E}\left[\log(X - X_1)!\right] + \frac{d^{\text{CM}}_{1,2}}{2} H(\Gamma_2|\Gamma_1) + H(Q_2|Q_1).$$

Using Lemma 5.2 and rearranging, this is precisely equal to $\Sigma(\mu^{\text{CM}}_2|\mu^{\text{CM}}_1)$, which completes the proof of (5.27).

D.6 Proof of Theorem 5.2: generalization to multiple sources

The proof of Theorem 5.2 is similar to that of Theorem 5.1 which was given in Section 5.3.

It is easy to verify that if $G^{(n)}_{[k]}$ is distributed according to either the multi-source Erdős–Rényi ensembles or the multi-source configuration model ensembles discussed in Section 5.3.5, then given any nonempty $A \subset [k]$, $A \neq [k]$, the joint distribution of $(G^{(n)}_A, G^{(n)}_{A^c})$ is similar to that of a two–source ensemble as in Section 5.1 with the following mark sets:

$$\tilde{\Xi}_1 := \left\{ x_A \in \Xi_A : \sum_{(x'_j:j\in[k]):(x'_j:j\in A)=x_A} p_{(x'_j:j\in[k])} > 0 \right\}$$
$$\tilde{\Xi}_2 := \left\{ x_{A^c} \in \Xi_{A^c} : \sum_{(x'_j:j\in[k]):(x'_j:j\in A^c)=x_{A^c}} p_{(x'_j:j\in[k])} > 0 \right\}$$
$$\tilde{\Xi}_{1,2} := \Xi_{[k]}$$
$$\tilde{\Theta}_1 := \left\{ \theta_A \in \Theta_A : \sum_{(\theta'_j:j\in[k]):(\theta'_j:j\in A)=\theta_A} q_{(\theta'_j:j\in[k])} > 0 \right\}$$
$$\tilde{\Theta}_2 := \left\{ \theta_{A^c} \in \Theta_{A^c} : \sum_{(\theta'_j:j\in[k]):(\theta'_j:j\in A^c)=\theta_{A^c}} q_{(\theta'_j:j\in[k])} > 0 \right\}$$
$$\tilde{\Theta}_{1,2} := \Theta_{[k]}$$

Moreover, we set $\tilde{\circ}_1 := \circ_A$ and $\tilde{\circ}_2 := \circ_{A^c}$. To establish the analogy, for the Erdős–Rényi ensemble, we define $\vec{\tilde{p}} = \{\tilde{p}_x\}_{x\in\tilde{\Xi}_{1,2}}$ such that $\tilde{p}_x = p_x$ for $x \in \tilde{\Xi}_{1,2}$. Furthermore, we define $\vec{\tilde{q}} = \{\tilde{q}_\theta\}_{\theta\in\tilde{\Theta}_{1,2}}$ such that $\tilde{q}_\theta = q_\theta$ for $\theta \in \tilde{\Theta}_{1,2}$. Similarly, for the configuration model ensemble, we let $\vec{\tilde{\gamma}} = \{\tilde{\gamma}_x\}_{x\in\tilde{\Xi}_{1,2}}$ such that $\tilde{\gamma}_x = \gamma_x$ for $x \in \tilde{\Xi}_{1,2}$, and define $\vec{\tilde{q}} = \{\tilde{q}_\theta\}_{\theta\in\tilde{\Theta}_{1,2}}$ where $\tilde{q}_\theta = q_\theta$

for $\theta \in \tilde{\Theta}_{1,2}$. It can be easily verified that (5.6) and (5.7) follow from the assumptions (5.49) and (5.50). Likewise, (5.8) and (5.9) follow from (5.51), (5.52) and (5.53).

Using this observation together with (5.12a)–(5.12c), we realize that for the multi-source Erdős–Rényi ensemble and nonempty $A \subseteq [k]$, we have

$$H(G_A^{(n)}) = \frac{d_A^{\text{ER}}}{2} n \log n + n \left(H(Q_A) + \sum_{x \in \Xi_A} s(p_x) \right) + o(n), \tag{D.23}$$

where $d_A^{\text{ER}} := \deg(\mu_A^{\text{ER}})$, and with $Q = (Q_i : i \in [k])$ having law $\vec{q}$, we let $Q_A := (Q_i : i \in A)$. In fact, the coefficient of n in the above expression is $\Sigma(\mu_A^{\text{ER}})$. Similarly, the above observation together with (5.14a)–(5.14c) establishes that for the multi-source configuration model ensemble and for nonempty $A \subseteq [k]$,

$$\begin{aligned} H(G_A^{(n)}) = \frac{d_A^{\text{CM}}}{2} n \log n + n \Big(&- s(d_A^{\text{CM}}) + H(X_A) - \mathbb{E}\left[\log X_A!\right] \\ &+ H(Q_A) + \frac{d_A^{\text{CM}}}{2} H(\Gamma_A | \Gamma_A \neq \circ_A) \Big) + o(n) \end{aligned}$$

where $d_A^{\text{CM}} := \deg(\mu_A^{\text{CM}})$. In the above expression, with $X \sim \vec{r}$ and $\Gamma^i = (\Gamma_j^i : j \in [k])$ for $1 \leq i \leq \Delta$ which are i.i.d. with law $\vec{\gamma}$, we define $X_A := \sum_{i=1}^{X} \mathbb{1}\left[\Gamma_j^i \neq \circ_j \text{ for some } j \in A\right]$. Here, if $X = 0$, then $X_A := 0$. Moreover, $Q = (Q_i : i \in [k])$ has law $\vec{q}$ and $Q_A := (Q_i : i \in A)$. Furthermore, $\Gamma = (\Gamma_i : i \in [k])$ has law $\vec{\gamma}$ and $\Gamma_A := (\Gamma_i : i \in A)$. It can be seen that the coefficient of n in the above expression is $\Sigma(\mu_A^{\text{CM}})$.

D.6.1 Proof of converse

Observe that for both the Erdős–Rényi and the configuration model ensembles, for $A \subset [k]$ nonempty, $A \neq [k]$, even if all the encoders in the set A as well as all the encoders in the set A^c can cooperate, since the distribution of $(G_A^{(n)}, G_{A^c}^{(n)})$ is identical to a two–source ensemble as was discussed above, using the converse result corresponding to the two–source case (i.e. Sections 5.3.3 and 5.3.4), with $\alpha_B := \sum_{i \in B} \alpha_i$ and $R_B := \sum_{i \in B} R_i$ for $B \subset [k]$, for $((\alpha_i, R_i) : i \in [k]) \in \mathcal{R}$, we must have

$$\begin{aligned} (\alpha_A, R_A) &\succeq ((d_{[k]} - d_{A^c})/2, \Sigma(\mu_A | \mu_{A^c})) \\ (\alpha_{A^c}, R_{A^c}) &\succeq ((d_{[k]} - d_A)/2, \Sigma(\mu_{A^c} | \mu_A)) \\ (\alpha_{[k]}, R_{[k]}) &\succeq (d_{[k]}/2, \Sigma(\mu_{[k]})). \end{aligned}$$

Here, μ denotes μ^{ER} or μ^{CM}, depending on the ensemble. Repeating this for all nonempty $A \subset [k]$, $A \neq [k]$, recovers all the necessary inequalities and completes the converse proof.

D.6.2 Proof of achievability for the Erdős–Rényi ensemble

Similar to Section 5.3.1, we employ a random binning codebook construction with $L_i^{(n)} = \lfloor \exp(\alpha_i n \log n + R_i n) \rfloor$ for $i \in [k]$. More precisely, For $i \in [k]$ and $H_i^{(n)} \in \mathcal{G}_i^{(n)}$, we generate

$f_i^{(n)}(H_i^{(n)})$ uniformly in $[L_i^{(n)}]$. The choice of $f_i^{(n)}(H_i^{(n)})$ is made independently for each $H_i^{(n)} \in \mathcal{G}_i^{(n)}$ and also for each domain $i \in [k]$. To explain the decoding procedure, similar to Section 5.3.1, let $\mathcal{M}^{(n)}$ be the set of $\vec{m} = \{m(x)\}_{x \in \Xi_{[k]}}$ such that $\sum_{x \in \Xi_{[k]}} |m(x) - np_x/2| \le n^{2/3}$. Furthermore, let $\mathcal{U}^{(n)}$ be the set of $\vec{u} = \{u(\theta)\}_{\theta \in \Theta_{[k]}}$ such that $\sum_{\theta \in \Theta_{[k]}} |u(\theta) - nq_\theta| \le n^{2/3}$. With these, let $\mathcal{G}_{\vec{p},\vec{q}}^{(n)}$ be the set of $H_{[k]}^{(n)} \in \mathcal{G}_{[k]}^{(n)}$ such that $\vec{m}_{H_{[k]}^{(n)}} \in \mathcal{M}^{(n)}$ and $\vec{u}_{H_{[k]}^{(n)}} \in \mathcal{U}^{(n)}$. At the receiver, upon receiving bin indices $i_j, 1 \le j \le k$, we form the set of $H_{[k]}^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)}$ such that $f_j^{(n)}(H_j^{(n)}) = i_j$ for $j \in [k]$. If there is only one graph in this set, the decoder outputs that graph; otherwise, it reports an error. It can be easily seen that the error events are as follows:

$$\mathcal{E}_1^{(n)} = \{G_{[k]}^{(n)} \notin \mathcal{G}_{\vec{p},\vec{q}}^{(n)}\},$$

and, for each nonempty $A \subset [k]$,

$$\begin{aligned}\mathcal{E}_A^{(n)} = \{\exists H_{[k]}^{(n)} \in \mathcal{G}_{\vec{p},\vec{q}}^{(n)} : H_i^{(n)} = G_i^{(n)} \text{ for } i \notin A, \\ H_i^{(n)} \neq G_i^{(n)}, f_i^{(n)}(H_i^{(n)}) = f_i^{(n)}(G_i^{(n)}) \text{ for } i \in A\}.\end{aligned}$$

For nonempty $A \subset [k]$ and $H_A^{(n)} \in \mathcal{G}_A^{(n)}$, we denote $(f_i^{(n)}(H_i^{(n)}) : i \in A)$ by $f_A^{(n)}(H_A^{(n)})$. Note that we may treat $f_A^{(n)}(H_A^{(n)})$ as an integer in the range $\prod_{i \in A} L_i^{(n)} \approx \lfloor \exp(\alpha_A n \log n + R_A n) \rfloor$. Recall that $\alpha_A = \sum_{i \in A} \alpha_i$ and $R_A = \sum_{i \in A} R_i$. Observe that due to our random binning procedure, $f_A^{(n)}(H_A^{(n)})$ is uniformly distributed in the range $\prod_{i \in A} L_i^{(n)}$. Moreover, for $H_{[k]}^{(n)}$ such that $H_i^{(n)} \neq G_i^{(n)}$ for $i \in A$, $f_A^{(n)}(H_A^{(n)})$ is independent from $f_A^{(n)}(G_A^{(n)})$. Thereby, for nonempty $A \subset [k]$, $A \neq [k]$, using the previously discussed fact that $(G_A^{(n)}, G_{A^c}^{(n)})$ is distributed according to a two–source ensemble, and using the analysis of Section 5.3.1, we realize that the probabilities of the error events $\mathcal{E}_A^{(n)}$, $\mathcal{E}_{A^c}^{(n)}$, $\mathcal{E}_{[k]}^{(n)}$, and $\mathcal{E}_1^{(n)}$ vanish as $n \to \infty$ given that $(\alpha_A, R_A) \succeq ((d_{[k]} - d_{A^c})/2, \Sigma(\mu_A^{\text{ER}} | \mu_{A^c}^{\text{ER}}))$, $(\alpha_{A^c}, R_{A^c}) \succeq ((d_{[k]} - d_A)/2, \Sigma(\mu_{A^c}^{\text{ER}} | \mu_A^{\text{ER}}))$, and $(\alpha_{[k]}, R_{[k]}) \succeq (d_{[k]}/2, \Sigma(\mu_{[k]}^{\text{ER}}))$. Repeating this argument for all nonempty $A \subset [k]$, $A \neq [k]$, we realize that the probabilities of all error events vanish, which completes the proof of achievability.

D.6.3 Proof of achievability for the configuration model ensemble

We again employ a random binning procedure as in the above, where, for $i \in [k]$ and $H_i^{(n)} \in \mathcal{G}_i^{(n)}$, we choose $f_i^{(n)}(H_i^{(n)})$ uniformly in the set $[L_i^{(n)}]$ with $L_i^{(n)} = \lfloor \exp(\alpha_i n \log n + R_i n) \rfloor$. To explain the decoding procedure, similar to the setup in Section 5.3.2, we define $\mathcal{D}^{(n)}$ be the set of degree sequences $\vec{d}$ such that $c_i(\vec{d}) = c_i(\vec{d}^{(n)})$ for all $0 \le i \le \Delta$. Moreover, let $\mathcal{M}^{(n)}$ be the set of $\vec{m} = (m(x) : x \in \Xi_{[k]})$ such that $\sum_{x \in \Xi_{[k]}} m(x) = m_n$, where $m_n := (\sum_{i=1}^n d^{(n)}(i))/2$, and $\sum_{x \in \Xi_{[k]}} |m(x) - m_n \gamma_x| \le n^{2/3}$. Also, let $\mathcal{U}^{(n)}$ be the set of $\vec{u} = (u(\theta) : \theta \in \Theta_{[k]})$ such that $\sum_{\theta \in \Theta_{[k]}} |u(\theta) - nq_\theta| \le n^{2/3}$. Let the random variables X and X_A for $A \subset [k]$ nonempty be defined as above, i.e. $X \sim \vec{r}$ and with $\Gamma^i = (\Gamma_j^i : j \in [k])$ for $1 \le i \le \Delta$ being i.i.d. with law

$\vec{\gamma}$, we define $X_A := \sum_{i=1}^{X} \mathbb{1}\left[\Gamma^i_j \neq \circ_j \text{ for some } j \in A\right]$ if $X > 0$, and $X_A := 0$ if $X = 0$. With this, let $\mathcal{W}^{(n)}$ be the set of $H^{(n)}_{[k]} \in \mathcal{G}^{(n)}_{[k]}$ such that (i) $\overrightarrow{\text{dg}}_{H^{(n)}_{[k]}} \in \mathcal{D}^{(n)}$, (ii) $\vec{m}_{H^{(n)}_{[k]}} \in \mathcal{M}^{(n)}$, (iii) $\vec{u}_{H^{(n)}_{[k]}} \in \mathcal{U}^{(n)}$, and (iv) for all $A \subset [k]$ nonempty and $0 \le j \le i \le \Delta$, we have

$$|c_{i,j}(\overrightarrow{\text{dg}}_{H^{(n)}_{[k]}}, \overrightarrow{\text{dg}}_{H^{(n)}_{A}}) - n\mathbb{P}\left(X = i, X_A = j\right)| \le n^{2/3}.$$

At the decoder, upon receiving $i_j : 1 \le j \le k$, we form the set of graphs $H^{(n)}_{[k]} \in \mathcal{W}^{(n)}$ such that $f^{(n)}(H^{(n)}_j) = i_j$ for $1 \le j \le k$. If there is only one graph in this set, the decoder outputs this graph; otherwise, it reports an error. It can be easily seen that the error events are as follows:

$$\mathcal{E}^{(n)}_1 = \{\mathcal{G}^{(n)}_{[k]} \notin \mathcal{W}^{(n)}\},$$

and for nonempty $A \subset [k]$,

$$\begin{aligned}\mathcal{E}^{(n)}_A = \{\exists H^{(n)}_{[k]} \in \mathcal{W}^{(n)} : H^{(n)}_i &= G^{(n)}_i \text{ for } i \notin A \\ H^{(n)}_i &\neq G^{(n)}_i, f^{(n)}(H^{(n)}_i) = f^{(n)}(G^{(n)}_i) \text{ for } i \in A\}.\end{aligned}$$

Similar to the above discussion in Section D.6.2, since for $A \subset [k]$ nonempty, $A \neq [k]$, the distribution of $(G^{(n)}_A, G^{(n)}_{A^c})$ is identical to a two–source configuration model ensemble, using the analysis in Section 5.3.2, we realize that the probabilities of the error events $\mathcal{E}^{(n)}_A$, $\mathcal{E}^{(n)}_{A^c}$, $\mathcal{E}^{(n)}_{[k]}$, and $\mathcal{E}^{(n)}_1$ vanish as $n \to \infty$ given that $(\alpha_A, R_A) \succeq ((d_{[k]} - d_{A^c})/2, \Sigma(\mu^{\text{CM}}_A | \mu^{\text{CM}}_{A^c}))$, $(\alpha_{A^c}, R_{A^c}) \succeq ((d_{[k]} - d_A)/2, \Sigma(\mu^{\text{CM}}_{A^c} | \mu^{\text{CM}}_A))$, and $(\alpha_{[k]}, R_{[k]}) \succeq (d_{[k]}/2, \Sigma(\mu^{\text{CM}}_{[k]}))$. Repeating this argument for all nonempty $A \subset [k]$, $A \neq [k]$, we realized that the probabilities of all error events vanish, which completes the proof of achievability.

Appendix E

Proofs for Chapter 6

E.1 Weak uniqueness of balanced allocations

Proof of Proposition 6.2. For a fixed $\delta > 0$, define the set

$$A_\delta := \{i \in V(H) : \partial_b\theta(i) - \partial_b\theta'(i) > \delta\}.$$

By assumption, we have $\sum_{i\in V(H)} |\partial_b\theta(i) - \partial_b\theta'(i)| < \infty$. Hence, A_δ is a finite set. Moreover

$$\sum_{i\in A_\delta} \partial_b\theta(i) - \partial_b\theta'(i) = \sum_{i\in A_\delta} \sum_{e\ni i, e\not\subseteq A_\delta} \theta(e,i) - \theta'(e,i). \tag{E.1}$$

Now, fix some $e \in E(H)$ such that $e \cap A_\delta \neq \emptyset$ and $e \not\subseteq A_\delta$. For $i \in e \cap A_\delta$ and $j \in e \setminus A_\delta$, we have

$$\partial_b\theta(j) - \partial_b\theta'(j) \le \delta < \partial_b\theta(i) - \partial_b\theta(i'),$$

which means that

$$\partial_b\theta(j) - \partial_b\theta(i) < \partial_b\theta'(j) - \partial_b\theta'(i).$$

Hence it is either the case that $\partial_b\theta'(j) > \partial_b\theta'(i)$ or $\partial_b\theta(j) < \partial_b\theta(i)$.

If $\theta(e,i) = 0$ for all $i \in e \cap A_\delta$, then $\sum_{i\in e\cap A_\delta} \theta(e,i) - \theta'(e,i) \le 0$. If $\theta(e,i^*) \neq 0$ for some $i^* \in e \cap A_\delta$, then $\partial_b\theta(i^*) \le \partial_b\theta(j)$ for all $j \in e \setminus A_\delta$. Consequently, $\partial_b\theta'(j) > \partial_b\theta'(i^*)$ for all $j \in e \setminus A_\delta$; thereby, $\theta'(e,j) = 0$ for all $j \in e \setminus A_\delta$. This means that $\sum_{i\in e\cap A_\delta} \theta'(e,i) = 1 \ge \sum_{i\in e\cap A_\delta} \theta(e,i)$. Hence, we have observed that, in either case, we have $\sum_{i\in e\cap A_\delta} \theta(e,i) - \theta'(e,i) \le 0$. Since this is true for all e such that $e \cap A_\delta \neq \emptyset$ and $e \not\subseteq A_\delta$, substituting into (E.1) we realize that

$$\sum_{i\in A_\delta} \partial_b\theta(i) - \partial_b\theta'(i) \le 0.$$

On the other hand, $\sum_{i\in A_\delta} \partial_b\theta(i) - \partial_b\theta'(i) \ge \delta|A_\delta|$. Combining these two we conclude that $A_\delta = \emptyset$. Symmetrically, the set $B_\delta := \{i \in V(H) : \partial_b\theta'(i) - \partial_b\theta(i) > \delta\}$ should be empty. Since δ is arbitrary, we conclude that $\partial\theta \equiv \partial\theta'$, i.e. $\partial_b\theta \equiv \partial_b\theta'$, which completes the proof. ☐

E.2 $\bar{\mathcal{H}}_*(\Xi)$ and $\bar{\mathcal{H}}_{**}(\Xi)$ are Polish spaces

In this section, we prove that $\bar{\mathcal{H}}_*(\Xi)$ and $\bar{\mathcal{H}}_{**}(\Xi)$ are Polish spaces when Ξ is a Polish space. In particular, by setting $\Xi = \{\emptyset\}$, this means that $\mathcal{H}_*$ and $\mathcal{H}_{**}$ are Polish spaces.

Proposition E.1. *Assume Ξ is a Polish space. Then, $\bar{\mathcal{H}}_*(\Xi)$ and $\bar{\mathcal{H}}_{**}(\Xi)$ are Polish spaces.*

Proof. We give the proof for $\bar{\mathcal{H}}_*(\Xi)$ here. The proof for $\bar{\mathcal{H}}_{**}(\Xi)$ is similar, and is therefore omitted.

First, we show $\bar{\mathcal{H}}_*(\Xi)$ is separable. Since Ξ is separable, it has a countable dense subset $X = \{\zeta_1, \zeta_2, \dots\} \subseteq \Xi$. Define A_n to be the set of all hypergraphs with n vertices with marks taking values in X, i.e.

$$A_n := \{[\bar{H}, i] \in \bar{\mathcal{H}}_*(\Xi) : |V(\bar{H})| = n, \xi_{\bar{H}}(e, i) \in X \, \forall (e, i) \in \Psi(\bar{H})\}.$$

Since there are finitely many hypergraphs on n vertices and X is countable, we see that A_n is countable. Now, define $A = \cup_n A_n$, which is countable. We claim that A is dense in $\bar{\mathcal{H}}_*(\Xi)$. To see this, for $[\bar{H}, i] \in \bar{\mathcal{H}}_*(\Xi)$ and $\epsilon > 0$ given, pick $(\bar{H}, i) \in [\bar{H}, i]$. Then take n large enough such that $\frac{1}{1+n} < \epsilon$. With H being the underlying unmarked hypergraph associated to $\bar{H}$, we now define a marked rooted hypergraph $(\bar{H}', i')$ where the underlying unmarked hypergraph H' has the property that (H', i') is the truncation of (H, i) up to depth n and the mark function $\xi_{\bar{H}'}$ is defined as follows. For $(\tilde{e}, \tilde{i}) \in \Psi(H')$, define $\xi_{\bar{H}'}(\tilde{e}, \tilde{i}) \in X$ such that $d_\Xi(\xi_{H'}(\tilde{e}, \tilde{i}), \xi_{\bar{H}}(\tilde{e}, \tilde{i})) < 1/(n+1)$. In this way, we have

$$\bar{d}_*([\bar{H}, i], [\bar{H}', i']) \le \frac{1}{1+n} < \epsilon.$$

But, $(\bar{H}', i')$ is finite, and the edge marks are in X; hence, $[\bar{H}', i'] \in A$. Since ϵ was arbitrary, this shows that A is dense in $\bar{\mathcal{H}}_*(\Xi)$. Thus, $\bar{\mathcal{H}}_*(\Xi)$ is separable.

Now, we turn to showing that $\bar{\mathcal{H}}_*(\Xi)$ is complete. Take a Cauchy sequence $[\bar{H}_n, i_n]$ in $\bar{\mathcal{H}}_*(\Xi)$ and let $(\bar{H}_n, i_n)$ be an arbitrary member of $[\bar{H}_n, i_n]$. Without loss of generality, by taking a subsequence if needed, we can assume that for $m > n$ we have

$$\bar{d}_*([\bar{H}_n, i_n], [\bar{H}_m, i_m]) < \frac{1}{1+n}.$$

This means that, with H_k being the underlying unmarked hypergraph associated to $\bar{H}_k$ for $k \ge 1$, we have

$$(H_n, i_n) \equiv_n (H_m, i_m) \qquad \forall m > n, \tag{E.2}$$

and

$$d_\Xi(\xi_{\bar{H}_n}(e', i'), \xi_{\bar{H}_m}(\phi_{n,m}(e'), \phi_{n,m}(i'))) < \frac{1}{1+n}, \tag{E.3}$$

for all $e' \in E_{H_n}(V^{H_n}_{i_n,n})$ and $i' \in e'$, where $\phi_{n,m}$ is the depth n isomorphism between H_n and H_m. Note that we can choose $\phi_{n,m}$ for $n > m$ so that

$$\phi_{n,m} = \phi_{m-1,m} \circ \dots \circ \phi_{n,n+1}. \tag{E.4}$$

In fact, since the RHS of (E.4) defines a depth n isomorphism from (H_n, i_n) to (H_m, i_m), one can define $\phi_{n,m}$ in this way.

In view of (E.2), we can construct a rooted hypergraph (H, i) so that $(H, i) \equiv_n (H_n, i_n)$ for all n. Further, there are depth n isomorphisms, ϕ_n, from (H, i) to (H_n, i_n), which satisfy the consistency condition

$$\phi_m = \phi_{n,m} \circ \phi_n \qquad \forall m > n. \tag{E.5}$$

So far we have constructed a rooted hypergraph (H, i) such that $[H_n, i_n] \to [H, i]$. Now, we construct a marked rooted hypergraph $(\bar{H}, i)$, where its underlying unmarked rooted hypergraph is (H, i), and the mark function $\xi_{\bar{H}} : \Psi(H) \to \Xi$ is defined as follows. Take $(e', i') \in \Psi(H)$ and choose d such that $e' \in E_H(V_{i,d}^H)$. We claim that the sequence

$$\{\xi_{\bar{H}_n}(\phi_n(e'), \phi_n(i'))\}_{n \geq d},$$

is Cauchy in Ξ. Indeed, for $m > n$, using (E.3), we have

$$d_\Xi(\xi_{\bar{H}_n}(\phi_n(e'), \phi_n(i')), \xi_{\bar{H}_m}(\phi_{n,m} \circ \phi_n(e'), \phi_{n,m} \circ \phi_n(i'))) < \frac{1}{1+n}.$$

Using (E.5), this means

$$d_\Xi(\xi_{\bar{H}_n}(\phi_n(e'), \phi_n(i')), \xi_{\bar{H}_m}(\phi_m(e'), \phi_m(i'))) < \frac{1}{1+n},$$

which means that the sequence is Cauchy in Ξ. Since Ξ is complete, we can define $\xi_{\bar{H}}(e', i')$ to be the limit of this sequence.

Now, we show that $[\bar{H}_n, i_n] \to [\bar{H}, i]$. For a given $d \in \mathbb{N}$, define

$$A_d := \{(e', i') : e' \in E_H(V_{i,d}^H), i' \in e'\}.$$

Since H_n are locally finite, H is also locally finite, and thus A_d is finite. On the other hand, since $\xi_{\bar{H}_n}(\phi_n(e'), \phi_n(i')) \to \xi_{\bar{H}}(e', i')$ for all $(e', i') \in A_d$, there exists a $N_d > d$ such that for all $n > N$, we have

$$d_\Xi(\xi_{\bar{H}_n}(\phi_n(e'), \phi_n(i')), \xi_{\bar{H}}(e', i')) < \frac{1}{1+d} \qquad \forall (e', i') \in A_d.$$

Moreover, since $n > N_d > d$, $[H_n, i_n] \equiv_d [H, i]$. Hence,

$$\bar{d}_*([\bar{H}_n, i_n], [\bar{H}, i]) < \frac{1}{1+d} \qquad \forall n > N_d.$$

Since d was arbitrary, this means that $[\bar{H}_n, i_n] \to [\bar{H}, i]$ and $\bar{\mathcal{H}}_*(\Xi)$ is complete. □

As was mentioned earlier, by setting $\Xi = \{\emptyset\}$, we conclude that $\mathcal{H}_*$ and $\mathcal{H}_{**}$ are Polish spaces. This is explicitly stated below as a corollary.

Corollary E.1. *The spaces $\mathcal{H}_*$ and $\mathcal{H}_{**}$ are Polish spaces.*

E.3 Some properties of measures on $\mathcal{H}_*$

Proof of Lemma 6.1. Part (i): We have

$$\vec{\mu}(\tilde{A}) = \int \mathbb{1}_{\tilde{A}} d\vec{\mu} = \int \partial \mathbb{1}_{\tilde{A}} d\mu.$$

But,

$$\partial \mathbb{1}_{\tilde{A}}(H, i) = \sum_{e \ni i} \mathbb{1}_{\tilde{A}}(H, e, i) = \sum_{e \ni i} \mathbb{1}_A(H, i) = \deg_H(i) \mathbb{1}_A(H, i).$$

Hence

$$\vec{\mu}(\tilde{A}) = \int \mathbb{1}_A(H, i) \deg_H(i) d\mu = \int \deg_H(i) d\mu = \int d\vec{\mu} = \vec{\mu}(\mathcal{H}_{**}),$$

which shows that $\tilde{A}$ happens $\vec{\mu}$ almost everywhere and the proof is complete.

Part (ii): Note that

$$\vec{\mu}(B) = \int_{\mathcal{H}_{**}} \mathbb{1}\left[[H, e, i] \in B\right] d\vec{\mu}([H, e, i]) = \int_{\mathcal{H}_*} \sum_{e \ni i} \mathbb{1}\left[[H, e, i] \in B\right] d\mu([H, i]).$$

On the other hand, $\vec{\mu}(B) = \vec{\mu}(\mathcal{H}_{**}) = \deg(\mu) = \int \deg_H(i) d\mu([H, i])$. Moreover, for all $[H, i] \in \mathcal{H}_*$, $\sum_{e \ni i} \mathbb{1}\left[[H, e, i] \in B\right] \leq \deg_H(i)$. Consequently, it must be the case that for μ–almost all $[H, i] \in \mathcal{H}_*$, $\sum_{e \ni i} \mathbb{1}\left[[H, e, i] \in B\right] = \deg_H(i)$, or equivalently $[H, e, i] \in B$ for all $e \ni i$. □

Proof of Lemma 6.2. If we define

$$A := \{[H, i] \in \mathcal{H}_* : f_k([H, i]) \to f_0([H, i])\},$$

and

$$\tilde{A} := \{[H, e, i] \in \mathcal{H}_{**} : \tilde{f}_k([H, e, i]) \to \tilde{f}_0([H, e, i])\},$$

then we have

$$\tilde{A} = \{[H, e, i] \in \mathcal{H}_{**} : [H, i] \in A\}.$$

Then the proof is an immediate consequence of Lemma 6.1. □

Proof of Lemma 6.3. Define $B := \{[H, e, i] \in \mathcal{H}_{**} : f_k(H, e, i) \to f_0(H, e, i)\}$. As $\vec{\mu}(B) = \vec{\mu}(\mathcal{H}_{**})$, from part (ii) of Lemma 6.1, for μ–almost all $[H, i] \in \mathcal{H}_*$, for all $e \ni i$, $f_k(H, e, i) \to f_0(H, e, i)$. This in particular implies that for μ–almost all $[H, i] \in \mathcal{H}_*$, $\partial f_k(H, i) \to \partial f_0(H, i)$. □

E.4 Proof of Lemma 6.4

We first prove that if the condition mentioned in Lemma 6.4 is satisfied, then $\mu_n \Rightarrow \mu$. Fix $\epsilon > 0$. Let $f : \mathcal{H}_* \to \mathbb{R}$ be a uniformly continuous and bounded function. There is some $\delta > 0$ such that $|f([H,i]) - f([H',i'])| < \epsilon$ when $d_{\mathcal{H}_*}([H,i],[H',i']) < \delta$. Now choose d such that $1/(1+d) < \delta$. For all rooted trees $[H,i] \in \mathcal{T}_*$, $[H,i] \in A_{(H,i)_d}$. Hence, one can find countably many rooted trees $\{T_j, i_j\}_{j=1}^\infty$ with depth at most d such that $A_{(T_j,i_j)}$ partitions $\mathcal{T}_*$; hence, one can find finitely many (T_j,i_j), $1 \le j \le m$ such that $\sum_{j=1}^m \mu(A_{(T_j,i_j)}) \ge 1-\epsilon$. If $\mathcal{A}$ denotes $\cup_{j=1}^m A_{(T_j,i_j)}$, then we have

$$\begin{aligned}\left|\int f d\mu - \sum_{j=1}^m f([T_j,i_j])\mu(A_{(T_j,i_j)})\right| &\le \sum_{j=1}^m \left|\int_{A_{(T_j,i_j)}} f d\mu - f([T_j,i_j])\mu(A_{(T_j,i_j)})\right| \\ &\quad + \|f\|_\infty \mu(\mathcal{A}^c) \\ &\le \epsilon(1+\|f\|_\infty),\end{aligned}$$

where the last inequality uses the facts that $\mu(\mathcal{A}^c) < \epsilon$ and $|f([H,i]) - f([T_j,i_j])| < \epsilon$ for $[H,i] \in A_{(T_j,i_j)}$, $1 \le j \le m$ since $1/(1+d) < \epsilon$. Similarly, we have

$$\begin{aligned}\left|\int f d\mu_n - \sum_{j=1}^m f([T_j,i_j])\mu(A_{(T_j,i_j)})\right| &\le \left|\int f d\mu_n - \sum_{j=1}^m f([T_j,i_j])\mu_n(A_{(T_j,i_j)})\right| \\ &\quad + \sum_{j=1}^m |f(T_j,i_j)||\mu_n(A_{(T_j,i_j)}) - \mu(A_{(T_j,i_j)})| \\ &\le \|f\|_\infty \left(1 - \sum_{j=1}^m \mu_n(A_{(T_j,i_j)})\right) + \epsilon \\ &\quad + \|f\|_\infty \sum_{j=1}^m |\mu_n(A_{(T_j,i_j)}) - \mu(A_{(T_j,i_j)})|.\end{aligned}$$

Combining these two, we have

$$\begin{aligned}\left|\int f d\mu_n - \int f d\mu\right| \le \|f\|_\infty \left(1 - \sum_{j=1}^m \mu_n(A_{(T_j,i_j)})\right) + \|f\|_\infty \left|\sum_{j=1}^m \mu_n(A_{(T_j,i_j)}) - \mu(A_{(T_j,i_j)})\right| \\ + \epsilon(2+\|f\|_\infty).\end{aligned}$$

Now, as n goes to infinity, $\mu_n(A_{(T_j,i_j)}) \to \mu(A_{(T_j,i_j)})$ by assumption. Also, $\mu(\mathcal{A}^c) < \epsilon$. Thus, we have

$$\limsup_{n\to\infty} \left|\int f d\mu_n - \int f d\mu\right| \le \epsilon(2 + 2\|f\|_\infty).$$

Since $\|f\|_\infty < \infty$ and this is true for any $\epsilon > 0$, we get $\int f d\mu_n \to \int f d\mu$; hence, $\mu_n \Rightarrow \mu$.

For the converse, note that $\mathbb{1}_{A_{(T,j)}}$ is a continuous function since $\mathbb{1}_{A_{(T,j)}}([H,i]) = \mathbb{1}_{A_{(T,i)}}([H',j'])$ for $d_{\mathcal{H}_*}([H,i],[H',i']) < 1/(1+d)$.

E.5 Some properties of Unimodular Measures

First, we prove Proposition 6.3. Our proof depends on the following lemma:

Lemma E.1. *Assume $\tau : \mathcal{H}_{**} \to \mathbb{R}$ is a measurable function and $\mu \in \mathcal{P}(\mathcal{H}_*)$ is a unimodular measure such that $\tau = 1$, $\vec{\mu}$–almost everywhere. Then, we have*

1. *With $\tau_1(H, e, i) := \mathbb{1}[\tau(H, e', i) = 1, \forall e' \ni i]$, it holds that $\tau_1 = 1$ $\vec{\mu}$–almost everywhere.*
2. *With $\tau_2(H, e, i) := \mathbb{1}[\tau(H, e, i') = 1, \forall i' \in e]$, it holds that $\tau_2 = 1$ $\vec{\mu}$–almost everywhere.*

Proof. In order to prove the first part, note that from Lemma 6.1 part (ii), we have that for μ–almost all $[H, i] \in \mathcal{H}_*$, $\tau(H, e', i) = 1$ for all $e' \ni i$. Now, part (i) of Lemma 6.1 implies that for $\vec{\mu}$–almost all $[H, e, i] \in \mathcal{H}_{**}$, $\tau(H, e, i) = 1$ for all $e' \ni i$, which is precisely what we need to prove.

For the second part, since μ is unimodular, we have

$$\int \mathbb{1}_{\tau=1} d\vec{\mu} = \int \nabla \mathbb{1}_{\tau=1} d\vec{\mu} = \int \frac{1}{|e|} \sum_{i' \in e} \mathbb{1}_{\tau(H,e,i')=1} d\vec{\mu}(H, e, i).$$

But since $\mathbb{1}_{\tau=1} = 1$ holds $\vec{\mu}$–almost everywhere and $0 \le \frac{1}{|e|} \sum_{i' \in e} \mathbb{1}_{\tau(H,e,i')=1} \le 1$, we conclude that

$$\frac{1}{|e|} \sum_{i' \in e} \mathbb{1}_{\tau(H,e,i')=1} = 1, \qquad \vec{\mu}\text{–a.e.}.$$

Since the summands are either zero or one, this means that $\tau(H, e, i') = 1$ for all $i' \in e$, $\vec{\mu}$–almost everywhere, which is what we wanted to show. □

Proof of Proposition 6.3. Define

$$A_k := \{[H, e, i] \in \mathcal{H}_{**} : \tau(H, e', i') = 1 \ \forall i' \in V(H) : d_H(i, i') \le k, \ \forall e' \ni i'\}. \tag{E.6}$$

Note that $A_0 = \{[H, e, i] : \tau(H, e, i) = 1\}$, for which it is known from the assumption that $\vec{\mu}(A_0) = \vec{\mu}(\mathcal{H}_{**})$. Now, we want to show that $\vec{\mu}(A_k) = \vec{\mu}(\mathcal{H}_{**})$ for all $k \ge 0$. We will do this by induction on k. Assume that $\vec{\mu}(A_k) = \vec{\mu}(\mathcal{H}_{**})$. Hence, with $\phi_k(H, e, i) := \mathbb{1}[(H, e, i) \in A_k]$, we have $\phi_k = 1$ $\vec{\mu}$–almost everywhere. Now, we will use Lemma E.1 to prove that $\phi_{k+1} = 1$ $\vec{\mu}$–almost everywhere.

To do so, using part 2 of Lemma E.1, if we define

$$B_1^k := \{[H, e, i] : \forall i' \in e, \ \phi_k(H, e, i') = 1\},$$

then we know $\vec{\mu}(B_1^k) = \vec{\mu}(\mathcal{H}_{**})$. Then, applying part 1 of Lemma E.1 for the function $\mathbb{1}_{B_1^k}$, we get that

$$B_2^k := \{[H, e, i] : \forall e' \ni i, [H, e', i] \in B_1^k\},$$

has the property that $\vec{\mu}(B_2^k) = \vec{\mu}(\mathcal{H}_{**})$. On the other hand,

$$B_2^k = \{[H, e, i] : \forall e' \ni i, \ \forall i' \in e', [H, e', i'] \in A_k\} = A_{k+1}.$$

Hence, we have proved that $\vec{\mu}(A_{k+1}) = \vec{\mu}(\mathcal{H}_{**})$, which is the inductive step.

Thus, $\vec{\mu}(\cap_{k\in\mathbb{N}} A_k) = \vec{\mu}(\mathcal{H}_{**})$. Using the fact that the vertex set is countable, and that the hypergraphs corresponding to the elements of $\mathcal{H}_*$ are connected, the property τ holds for all the directed edges $\vec{\mu}$–almost everywhere, in $A := \cap_{k\in\mathbb{N}} A_k$. Hence, the proof is complete. ☐

Now, we prove that the local weak limit of finite marked hypergraphs is a unimodular probability distribution on $\bar{\mathcal{H}}_*(\Xi)$. By setting $\Xi = \{\emptyset\}$ in the following proposition, we can conclude that the local weak limit of finite simple hypergraphs is unimodular, as claimed in Section 6.2.8.

Proposition E.2. *Assume $\{\bar{H}_n\}$ is a sequence of finite marked hypergraphs, with $\xi_{\bar{H}_n} = \xi_n$ the associated edge mark functions, taking values in some metric space Ξ. Now, if*

$$\bar{\mu}_n := u_{\bar{H}_n} = \frac{1}{|V(\bar{H}_n)|} \sum_{i\in V(\bar{H}_n)} \delta_{[(\bar{H}_n, i), i]},$$

then $\bar{\mu}_n \in \mathcal{P}(\bar{\mathcal{H}}_*(\Xi))$ is unimodular for each n. Moreover, if $\bar{\mu}_n$ converge weakly to some limit $\bar{\mu} \in \mathcal{P}(\bar{\mathcal{H}}_*(\Xi))$ such that $\deg(\bar{\mu}) < \infty$, $\bar{\mu}$ is also unimodular.

Proof. First, we show that $\bar{\mu}_n$ is unimodular for each n. For this, take a Borel function $f : \bar{\mathcal{H}}_{**}(\Xi) \to [0, \infty)$ and note that

$$\begin{aligned}
\int f(\bar{H}, e, i) d\vec{\bar{\mu}}_n &= \int \partial f(\bar{H}, i) d\bar{\mu}_n \\
&= \frac{1}{|V(\bar{H}_n)|} \sum_{i\in V(\bar{H}_n)} \partial f(\bar{H}_n, i) \\
&= \frac{1}{|V(\bar{H}_n)|} \sum_{i\in V(\bar{H}_n)} \sum_{e\ni i} f(\bar{H}_n, e, i) \\
&= \frac{1}{|V(\bar{H}_n)|} \sum_{e\in E(\bar{H}_n)} \sum_{i\in e} f(\bar{H}_n, e, i) \\
&= \frac{1}{|V(\bar{H}_n)|} \sum_{e\in E(\bar{H}_n)} \sum_{i\in e} \nabla f(\bar{H}_n, e, i) \\
&= \int \nabla f(\bar{H}, e, i) d\vec{\bar{\mu}}_n.
\end{aligned}$$

Since this holds for all nonnegative Borel functions f, $\bar{\mu}_n$ is unimodular, by definition.

Now, for each n, define measures $\vec{\bar{\mu}}_n^{(1)}$ and $\vec{\bar{\mu}}_n^{(2)}$ on $\bar{\mathcal{H}}_{**}(\Xi)$ so that for any Borel function $f : \bar{\mathcal{H}}_{**}(\Xi) \to [0, \infty)$, we have

$$\int f d\vec{\bar{\mu}}_n^{(1)} := \int \sum_{e \ni i} f(\bar{H}, e, i) d\bar{\mu}_n([\bar{H}, i]), \tag{E.7}$$

and

$$\int f d\vec{\bar{\mu}}_n^{(2)} := \int \sum_{e \ni i} \frac{1}{|e|} \sum_{j \in e} f(\bar{H}, e, j) d\bar{\mu}_n([\bar{H}, i]). \tag{E.8}$$

We also define $\vec{\bar{\mu}}^{(1)}$ and $\vec{\bar{\mu}}^{(2)}$ for $\bar{\mu}$ in a similar fashion. Note that the RHS of (E.7) is $\int \partial f d\bar{\mu}_n = \int f d\vec{\bar{\mu}}_n$. Therefore, $\vec{\bar{\mu}}_n^{(1)} = \vec{\bar{\mu}}_n$. Also, the RHS of (E.8) is $\int \partial \nabla f d\bar{\mu}_n = \int \nabla f d\vec{\bar{\mu}}_n$. Hence, $\int f d\vec{\bar{\mu}}_n^{(2)} = \int \nabla f d\vec{\bar{\mu}}_n$. Since we have shown that $\bar{\mu}_n$ is unimodular, this implies that $\vec{\bar{\mu}}_n^{(1)} = \vec{\bar{\mu}}_n^{(2)}$.

Now, we claim that $\vec{\bar{\mu}}^{(1)} = \vec{\bar{\mu}}^{(2)}$. To show this, take a bounded continuous function $f : \bar{\mathcal{H}}_{**}(\Xi) \to \mathbb{R}$. For $k > 0$, define

$$f_k(\bar{H}, e, i) := \begin{cases} f(\bar{H}, e, i) & \text{if } \deg_H(i) \le k, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that f_k is continuous for each k, as f is continuous. Moreover, ∂f_k and $\partial \nabla f_k$ are bounded for each k. This, together with $\bar{\mu}_n \Rightarrow \bar{\mu}$, implies that

$$\int f_k d\vec{\bar{\mu}}_n^{(1)} = \int \partial f_k d\bar{\mu}_n \to \int \partial f_k d\bar{\mu} = \int f_k d\vec{\bar{\mu}}^{(1)},$$

and

$$\int f_k d\vec{\bar{\mu}}_n^{(2)} = \int \partial \nabla f_k d\bar{\mu}_n \to \int \partial \nabla f_k d\bar{\mu} = \int f_k d\vec{\bar{\mu}}^{(2)}.$$

This, together with the fact that $\vec{\bar{\mu}}_n^{(1)} = \vec{\bar{\mu}}_n^{(2)}$, implies that for all k

$$\int f_k d\vec{\bar{\mu}}^{(1)} = \int f_k d\vec{\bar{\mu}}^{(2)}.$$

Note that as all hypergraphs are locally finite, $f_k \to f$ pointwise. Thus, sending k to infinity and using the dominated convergence theorem, we have that for any bounded continuous function $f : \bar{\mathcal{H}}_{**}(\Xi) \to [0, \infty)$,

$$\int f d\vec{\bar{\mu}}^{(1)} = \int f d\vec{\bar{\mu}}^{(2)}.$$

Since f can be an arbitrary bounded continuous function, we have $\vec{\bar{\mu}}^{(1)} = \vec{\bar{\mu}}^{(2)}$. But the definition of $\vec{\bar{\mu}}^{(1)}$ and $\vec{\bar{\mu}}^{(2)}$ then implies that for any nonnegative Borel function f we have $\int f d\vec{\bar{\mu}} = \int \nabla f d\vec{\bar{\mu}}$, which means that $\bar{\mu}$ is unimodular. □

E.6 Proof of unimodularity of $\mathsf{UGWHT}(P)$

Here we show that $\mathsf{UGWHT}(P)$ is unimodular. Before that we prove the following lemma, which is useful in calculating $\int f d\mu$ when $\mu = \mathsf{UGWHT}(P)$.

Lemma E.2. *Let $P \in \mathcal{P}(\Lambda)$, and Γ a random variable with law P, where $\mathbb{E}[\Gamma(k)] < \infty$ for all $k \geq 2$. Moreover, let $\mu = \mathsf{UGWHT}(P)$. Then, for any Borel function $f : \mathcal{H}_{**} \to [0, \infty)$, we have*

$$\int f d\vec{\mu} = \sum_{k=2}^{\infty} \mathbb{E}[\Gamma(k)] \sum_{\gamma \in \Lambda} \hat{P}_k(\gamma) \mathbb{E}[f(T, (k, 1), \emptyset) | \Gamma_\emptyset = \gamma + \mathbb{e}_k],$$

where the expectation is with respect to the random rooted hypertree of Definition 6.28. Here, $\mathbb{e}_k \in \Lambda$ is such that $\mathbb{e}_k(k) = 1$ and $\mathbb{e}_k(j) = 0$ for $j \neq k$.

Proof. Due to the definition of $\vec{\mu}$, we have

$$\int f d\vec{\mu} = \int \partial f d\mu = \mathbb{E}\left[\sum_{k=2}^{h(\Gamma_\emptyset)} \sum_{i=1}^{\Gamma_\emptyset(k)} f(T, (k, i), \emptyset)\right].$$

Since $\Gamma_\emptyset$ has distribution P, we have

$$\mathbb{E}\left[\sum_{k=1}^{h(\Gamma_\emptyset)} \sum_{i=1}^{\Gamma_\emptyset(k)} f(T, (k, i), \emptyset)\right] = \sum_{\gamma \in \Lambda} P(\gamma) \mathbb{E}\left[\sum_{k=2}^{h(\gamma)} \sum_{i=1}^{\gamma(k)} f(T, (k, i), \emptyset) \middle| \Gamma_\emptyset = \gamma\right].$$

Now, due to symmetry, conditioned on $\Gamma_\emptyset = \gamma$, for a given $k \leq h(\gamma)$, all $f(T, (k, i), \emptyset)$ for $1 \leq i \leq \gamma(k)$ have the same distribution, hence

$$\begin{aligned}\int f d\vec{\mu} &= \sum_{\gamma \in \Lambda} P(\gamma) \sum_{k=2}^{h(\gamma)} \gamma(k) \mathbb{E}[f(T, (k, 1), \emptyset) | \Gamma_\emptyset = \gamma] \\ &= \sum_{\gamma \in \Lambda} P(\gamma) \sum_{k=2}^{\infty} \gamma(k) \mathbb{E}[f(T, (k, 1), \emptyset) | \Gamma_\emptyset = \gamma],\end{aligned}$$

where the second equality uses the fact that $\gamma(k) = 0$ for $k > h(\gamma)$. Now, since all the terms are nonnegative, employing Tonelli's theorem to switch the order of integrals we have

$$\int f d\vec{\mu} = \sum_{k=2}^{\infty} \sum_{\gamma \in \Lambda : \gamma(k) > 0} P(\gamma) \gamma(k) \mathbb{E}[f(T, (k, 1), \emptyset) | \Gamma_\emptyset = \gamma].$$

Using the definition of $\hat{P}_k$, $P(\gamma)\gamma(k)$ is equal to $\mathbb{E}[\Gamma(k)]\,\hat{P}_k(\gamma - \mathbb{e}_k)$ for $\gamma \in \Lambda$ such that $\gamma(k) > 0$, where $\mathbb{e}_k \in \Lambda$ is such that $\mathbb{e}_k(k) = 1$ and $\mathbb{e}_k(j) = 0$ for $j \neq k$. Hence, we have

$$\begin{aligned}\int f d\vec{\mu} &= \sum_{k=2}^{\infty} \mathbb{E}[\Gamma(k)] \sum_{\gamma\in\Lambda:\gamma(k)>0} \hat{P}_k(\gamma - \mathbb{e}_k)\mathbb{E}[f(T,(k,1),\emptyset)|\Gamma_\emptyset = \gamma] \\ &= \sum_{k=2}^{\infty} \mathbb{E}[\Gamma(k)] \sum_{\gamma\in\Lambda} \hat{P}_k(\gamma)\mathbb{E}[f(T,(k,1),\emptyset)|\Gamma_\emptyset = \gamma + \mathbb{e}_k]\,.\end{aligned}$$

We can interpret the last expression above as follows. In computing $\int f d\vec{\mu}$, when we consider $f([T, e, \emptyset])$, the edge e attached to the root $\emptyset$ of the tree T is of size k with probability $\Gamma(k)$. This explains the outer summation. Since the contribution to the integral is the same whichever edge of size k connected to the root is picked, suppose the edge picked is the edge $(k, 1)$. Then the type of the rest of the edges connected to the root is given by $\hat{P}_k$, and so $\Gamma_\emptyset$ will equal $\gamma + \mathbb{e}_k$ with probability $\hat{P}_k(\gamma)$. This explains the inner summation. □

Now we are ready to prove the unimodularity of $\mathsf{UGWHT}(P)$:

Proof of Proposition 6.4. We need to prove that for a nonnegative measurable function $f : \mathcal{H}_{**} \to [0,\infty)$ we have $\int f d\vec{\mu} = \int \nabla f d\vec{\mu}$. Using Lemma E.2, we have

$$\begin{aligned}\int \nabla f d\vec{\mu} &\overset{(a)}{=} \sum_{k=2}^{\infty} \mathbb{E}[\Gamma(k)]\,\mathbb{E}\left[\nabla f(\tilde{T}_k,(k,1),\emptyset)\right] \\ &= \sum_{k=2}^{\infty} \mathbb{E}[\Gamma(k)]\,\frac{1}{k}\left(\mathbb{E}\left[f(\tilde{T}_k,(k,1),\emptyset)\right] + \sum_{i=1}^{k-1}\mathbb{E}\left[f(\tilde{T}_k,(k,1),(k,1,i))\right]\right).\end{aligned}$$

Here, for $k \geq 2$, $\tilde{T}_k$ denotes a tree with root $\emptyset$ that has an edge $(k, 1)$ of size k connected to the root, with the type of the other edges connected to the root being $\hat{P}_k$, and with the subtrees at the other vertices of all the edges (including the edge $(k, 1)$) generated according to the rules of $\mathsf{UGWHT}(P)$. Step (a) is justified because, for each $k \geq 2$, we have

$$\mathbb{E}\left[\nabla f(\tilde{T}_k,(k,1),\emptyset)\right] = \sum_{k=2}^{\infty} \mathbb{E}[\Gamma(k)] \sum_{\gamma\in\Lambda} \hat{P}_k(\gamma)\mathbb{E}[\nabla f(T,(k,1),\emptyset)|\Gamma_\emptyset = \gamma + \mathbb{e}_k]\,.$$

Now, because of the symmetry in the construction of $\tilde{T}_k$, for each $i = 1,\dots,k-1$, $(\tilde{T}_k,(k,1),(k,1,i))$ has the same distribution as $(\tilde{T}_k,(k,1),\emptyset)$. Hence, $\mathbb{E}\left[f(\tilde{T}_k,(k,1),(k,1,i))\right] = \mathbb{E}\left[f(\tilde{T}_k,(k,1),\emptyset)\right]$. Substituting and simplifying we get

$$\int \nabla f d\vec{\mu} = \sum_{k=2}^{\infty} \mathbb{E}[\Gamma(k)]\,\mathbb{E}\left[f(\tilde{T},(k,1),\emptyset)\right] = \int f d\vec{\mu},$$

where the last equality again uses Lemma E.2. This completes the proof of the unimodularity of μ. □

E.7 Configuration model on hypergraphs and their local weak limit: Proof of Theorem 6.4

In this section, we prove Theorem 6.4. We prove this in two steps: first, in Section E.7.2, we prove that $\mathbb{E}\left[u_{H_n^e}\right]$ converges weakly to $\mathsf{UGWHT}(P)$, where the expectation in $\mathbb{E}\left[u_{H_n^e}\right]$ is taken with respect to the randomness in the construction of H_n^e. Later, in Section E.7.3, we conclude the almost sure convergence by a concentration argument. See [Bor14] for an argument on the local weak convergence of the configuration model in the graph regime.

Throughout this section, we employ the vertex and edge indexing notations $\mathbb{N}_{\text{vertex}}$ and $\mathbb{N}_{\text{edge}}$ defined in Section 6.2.9. By an abuse of notation, for $a = (s_1, e_1, i_1, \dots, s_k, e_k, i_k) \in \mathbb{N}_{\text{vertex}}$ where $s_j \geq 2$, $e_j \geq 1$, and $1 \leq i_j \leq s_j - 1$ for all $1 \leq j \leq k$, and integers $s \geq 2$, $e \geq 1$ and $1 \leq r \leq s-1$, we write (a, s, e, r) for $(s_1, e_1, i_1, \dots, s_k, e_k, i_k, s, e, r)$. For an edge $e \in \mathbb{N}_{\text{edge}}$ with size k, and an integer $1 \leq r \leq k-1$, (e, r) is defined similarly. Furthermore, $\mathcal{T}(\mathbb{N}_{\text{vertex}}, \mathbb{N}_{\text{edge}})$ in this section denotes the set of hypertrees with vertex set and edge set being subsets of $\mathbb{N}_{\text{vertex}}$ and $\mathbb{N}_{\text{edge}}$, respectively. Such a hypertree is treated to be rooted at $\emptyset$, unless otherwise stated. Moreover, for a sequence of types $\{\gamma_a\}_{a \in \mathbb{N}_{\text{vertex}}}$, $\mathcal{T}(\{\gamma_a\}_{a \in \mathbb{N}_{\text{vertex}}})$ denotes the tree in $\mathcal{T}(\mathbb{N}_{\text{vertex}}, \mathbb{N}_{\text{edge}})$ in which the type of each node $a \in \mathbb{N}_{\text{vertex}}$ in the hypertree below that node is equal to γ_a (recall Figure 6.4 from Section 6.2.9 as an example).

Before going through the proof, we need to define a procedure called the "exploration process" in Section E.7.1 below.

E.7.1 Exploration process

Assume that a random hypergraph H_n on the vertex set $\{1, \dots, n\}$ is obtained from a given type sequence $\gamma^{(n)} = (\gamma_1^{(n)}, \dots, \gamma_n^{(n)})$ satisfying (6.82a) and (6.82b). Note that, following our discussion in Section 6.9.1, H_n is identified by a set of random matchings $\sigma_2, \dots, \sigma_n$. Here, we introduce a procedure that choses a node v_0 uniformly at random in $\{1, \dots, n\}$ and explores the local neighborhood of that node in a breadth–first manner. This process at each step produces a hypertree in $\mathcal{T}(\mathbb{N}_{\text{vertex}}, \mathbb{N}_{\text{edge}})$, which turns out to be locally isomorphic to the neighborhood of v_0 given that the local neighborhood of v_0 in H_n is tree–like. A similar process in the graph regime is introduced in [Bor14].

Formally speaking, the exploration process starts at time $t = 0$ with choosing a vertex $v_0 \in \{1, \dots, n\}$ uniformly at random as the root. Then, at each time step, we explore one edge in the neighborhood of v_0. This is done in a breadth first manner, i.e. we first explore edges adjacent to v_0, then edges connected to neighbors of v_0 and so on. More precisely, at each time $t \geq 1$, we have a node indexing function $\phi_t : \mathbb{N}_{\text{vertex}} \to \{1, \dots, n\} \cup \{\times\}$. To begin with, we define $\phi_1(\emptyset) = v_0$ and $\phi_1(\mathbf{i}) = \times$ for $\emptyset \neq \mathbf{i} \in \mathbb{N}_{\text{vertex}}$. Also, at each time step t, we partition $\Delta^{(n)}$ into three sets: an "active set" A_t, a "connected set" C_t, and an "undiscovered set" U_t. These are initialized by setting $A_1 = \Delta_{v_0}^{(n)}$, $C_1 = \emptyset$ and $U_1 = \Delta^{(n)} \setminus A_1$. Moreover, at time t, $N_t \subset \{1, \dots, n\}$ contains the explored nodes at time t, and is initialized as $N_1 = \{v_0\}$.

At time t, given the sets A_t, U_t, C_t and ϕ_t, we first form the set $W_t := \{(\mathbf{i}, k, j) : \phi_t(\mathbf{i}) \neq \times, e^k_{\phi_t(\mathbf{i}),j} \in A_t\}$. If W_t is nonempty, we define $(\mathbf{i}_t, k_t, j_t)$ to be the lexicographically smallest element in W_t and let $e_t := e^{k_t}_{\phi_t(\mathbf{i}_t),j_t}$. In fact, e_t is the partial edge chosen at time t to be matched with other partial edges to form an edge. Now, define

$$e_{t,j} := \sigma^{(j-1)}_{k_t}(e_t) \qquad 1 \le j \le k_t - 1,$$

which are the $k_t - 1$ other partial edges matched with e_t. Also, for $1 \le j \le k_t - 1$, let $u_{t,j} := \nu(e_{t,j})$ be the node associated to $e_{t,j}$. Moreover, we update the sets C_t, A_t, U_t and N_t as follows:

$$C_{t+1} = C_t \cup \{e_t, e_{t,1}, \dots, e_{t,k_t-1}\}, \tag{E.9a}$$

$$A_{t+1} = A_t \setminus \{e_t, e_{t,1}, \dots, e_{t,k_t-1}\} \bigcup_{j=1}^{k_t-1} \left(\Delta^{(n)}_{u_{t,j}} \cap U_t\right), \tag{E.9b}$$

$$U_{t+1} = U_t \setminus \bigcup_{j=1}^{k_t-1} \Delta^{(n)}_{u_{t,j}}, \tag{E.9c}$$

$$N_{t+1} = N_t \cup \{u_{t,1}, \dots, u_{t,k_t-1}\}. \tag{E.9d}$$

In order to update ϕ_t, define $\tilde{j}_t$ to be the minimum j such that

$$\phi_t((\mathbf{i}_t, k_t, j, 1)) = \times.$$

With this, set ϕ_{t+1} to be equal to ϕ_t except for the following values:

$$\phi_{t+1}((\mathbf{i}_t, k_t, \tilde{j}_t, l)) = u_{t,l} \qquad 1 \le l \le k_t - 1.$$

This in particular means that the set of nodes in $\{1, \dots, n\}$ that appear in the range of ϕ_{t+1} is precisely N_{t+1}.

At each time t, we define a rooted hypertree formed by the exploration process, which we denote by R_t, which is a member of $\mathcal{T}(\mathbb{N}_{\text{vertex}}, \mathbb{N}_{\text{edge}})$. R_t is identified through the mapping ϕ_t, i.e. its vertex set is $\{\mathbf{i} \in \mathbb{N}_{\text{vertex}} : \phi_t(\mathbf{i}) \neq \times\}$, and its edge set is $\{(\mathbf{i}_r, k_r, \tilde{j}_r), 1 \le r \le t-1\}$. This process continues until $A_t = \emptyset$, which results in exploring the connected component of v_0. Note that since the permutations determining the configuration model are random, the exploration process is in fact a random process. Let $\mathcal{F}_t$ be the sigma field generated by all the random variables defined above up to time t. Let τ be the stopping time corresponding to $A_\tau = \emptyset$. For the sake of simplicity, for $t > \tau$, we define $R_t = R_\tau$.

E.7.2 Convergence of expectation

In this section, we provide the proof of the convergence of $\mathbb{E}\left[u_{H^e_n}\right]$. This is done in two parts. Loosely speaking, we first show that, for any integer $d \ge 1$, with high probability, H_n

rooted at a vertex chosen uniformly at random up to depth d is a simple hypertree. Then, we prove that the distribution of the limiting depth t is that of a Galton–Watson process. The former is proved in Proposition E.3 below and the latter in Proposition E.4.

More precisely, let $\bar{\mu}^e_n$ and μ denote $\mathbb{E}\left[u_{H^e_n}\right]$ and $\mathsf{UGWHT}(P)$, respectively. In order to show that $\bar{\mu}^e_n \Rightarrow \mu$, using Lemma 6.4, it suffices to prove that for any $d \geq 1$, and with (T, o) being a rooted hypertree of depth at most d, $\bar{\mu}^e_n(A_{(T,o)}) \to \mu(A_{(T,o)})$, where we recall that $A_{(T,o)} = \{[H, j] \in \mathcal{H}_* : (H, j)_d \equiv (T, o)\}$. Note that, with v_0 being chosen uniformly at random in $\{1, \dots, n\}$, we have

$$\begin{aligned} \bar{\mu}^e_n(A_{(T,o)}) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \delta_{[H^e_n(i),i]}(A_{(T,o)})\right] \\ &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[H^e_n(i),i]\in A_{(T,o)}}\right] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{P}\left((H^e_n, i)_d \equiv (T, o)\right) \\ &= \mathbb{P}\left((H^e_n, v_0)_d \equiv (T, o)\right). \end{aligned} \tag{E.10}$$

Thus, motivated by the above discussion, we need to show that for all $d \geq 1$ and (T, o) with depth at most d, $\mathbb{P}\left((H^e_n, v_0)_d \equiv (T, o)\right) \to \mu(A_{(T,o)})$. We prove this in two steps. First, we prove in Proposition E.3 that the probability of $(H_n, v_0)_d$ being a simple hypertree goes to one as n goes to infinity. Subsequently, in Proposition E.4, we show that $\mathbb{P}\left((H^e_n, v_0)_d \equiv (T, i)\right)$ converges to $\mu(A_{(T,i)})$.

In the following statement, $\gamma^{(n)} = (\gamma^{(n)}_1, \dots, \gamma^{(n)}_n)$ is a fixed type sequence and H_n is the random multihypergraph resulting from the configuration model. Moreover, for an integer $d \geq 1$ and vertex $v \in \{1, \dots, n\}$, $(H_n, v)_d$ is the multihypergraph rooted at vertex v containing nodes with distance at most d from v and edges in H_n with all endpoints among this set of vertices.

Proposition E.3. *Assume conditions (6.84a), (6.84b) and (6.84c) are satisfied with constants $c_1, c_2, c_3, \epsilon > 0$. Then, if v_0 is chosen uniformly at random from $\{1, \dots, n\}$,*

$$\lim_{n\to\infty} \mathbb{P}\left((H_n, v_0)_d \text{ is a simple hypertree}\right) = 1.$$

Proof. Note that (6.84a) and (6.84b) imply that the degrees of all vertices and the sizes of all edges in H_n are bounded by $\alpha_n := c_1(\log n)^{c_2}$. Therefore, there are at most $\beta_n := (\alpha_n)^{2d+5}$ edges in $(H_n, v_0)_{d+1}$. Since at each step of the exploration process we form one edge, $(H_n, v_0)_{d+1}$ is completely observed up to step β_n. On the other hand, if at step t, the partial edge to be matched at that step, which is e_t in our notation, is matched to partial edges outside A_t, all of the endpoints of the newly formed edge at step t are in N^c_t, i.e. they are not observed so far. If in addition to this property, all the $k_t - 1$ vertices of these partial edges that are matched with e_t are distinct, no improper edges or multiple edges are formed at step t. If these properties hold for all $1 \leq t \leq \beta_n$, $(H_n, v_0)_d$ will be a simple hypertree. Note that, in order to make sure that $(H_n, v_0)_d$ is a simple hypertree, we need to make sure that there is no improper edge in the exploration process up to depth $d+1$. This guarantees that even vertices at depth d do not get connected to each other.

To formalize this, fix $1 \leq t \leq \beta_n$ and assume that the exploration process is not terminated up to step t, and at step t, we need to match the partial edge e_t of size k_t. Let $E_{t,1}$ denote the event that $e_{t,1} = \sigma^{(1)}_{k_t}(e_t) \in A_t \cap \Delta^{(n)}(k_t)$. Moreover, for $2 \leq i \leq k_t - 1$, let $E_{t,i}$ be te event that

$$e_{t,i} = \sigma^{(i)}_{k_t}(e_t) \in (A_t \cap \Delta^{(n)}(k_t)) \cup \left(\bigcup_{j=1}^{i-1} \Delta^{(n)}_{\nu(e_{t,j})}(k_t) \right),$$

where $\Delta^{(n)}_{\nu(e_{t,j})}(k_t)$ denotes the set of partial edges of size k_t connected to the vertex associated to $e_{t,j}$. Note that having chosen $e_{t,1}, \dots, e_{t,i-1}$, there are $|\Delta^{(n)}(k_t)| - |\Delta^{(n)}(k_t) \cap C_t| - i$ many candidates for $\sigma^{(i)}_{k_t}(e_t)$, each having the same chance of being chosen. We claim that $|A_t \cap \Delta^{(n)}(k_t)| \leq t\alpha_n^2$. The reason is that at each step in the exploration process, at most α_n many new vertices are added to N_t, each of which having at most α_n many partial edges of size k_t. On the other hand, $|\Delta^{(n)}_{\nu(e_{t,j})}(k_t)| \leq \alpha_n$. Consequently,

$$\left| (A_t \cap \Delta^{(n)}(k_t)) \cup \left(\bigcup_{j=1}^{i-1} \Delta^{(n)}_{\nu(e_{t,j})}(k_t) \right) \right| \leq (t+1)\alpha_n^2.$$

Additionally, at each step, at most α_n partial edges are added to C_t to form an edge; therefore, $|C_t| \leq t\alpha_n$. This, together with (6.84c), implies that for $1 \leq i \leq k_t \leq \alpha_n$, we have $|\Delta^{(n)}(k_t)| - |\Delta^{(n)}(k_t) \cap C_t| - i \geq c_3 n^\epsilon - (t+1)\alpha_n$. Since each of these candidates have the same probability of being chosen, we have

$$\mathbb{P}(E_{t,i}) \leq \frac{(t+1)\alpha_n^2}{c_3 n^\epsilon - (t+1)\alpha_n} \qquad 1 \leq i \leq k_t - 1.$$

If E_t denotes $\cup_{i=1}^{k_t-1} E_{t,i}$, using $k_t \leq \alpha_n$ and the union bound, we have

$$\mathbb{P}\left(\cup_{t=1}^{\beta_n} E_t \right) \leq \frac{(\beta_n+1)^2 \alpha_n^3}{c_3 n^\epsilon - (\beta_n+1)\alpha_n}.$$

Note that, α_n and β_n scale logarithmically in n, and d is fixed. Hence, due to the $c_3 n^\epsilon$ term in the denominator, the above probability goes to zero as n goes to infinity. As was discussed above, outside the event $\cup_{t=1}^{\beta_n} E_t$, the rooted hypergraph $(H_n, v_0)_d$ is a simple hypertree. □

Proposition E.4. *With the assumptions of Theorem 6.4, for an integer $d \geq 1$ and a rooted hypertree (T, o) with depth at most d, if v_0 is a node chosen uniformly at random from $\{1, \dots, n\}$, we have*

$$\lim_{n \to \infty} \mathbb{P}\left((H^e_n, v_0)_d \equiv (T, o) \right) = \mu(A_{(T,o)}).$$

Note that, the above proposition, together with the discussion in (E.10), implies that $\mathbb{E}\left[u_{H_n^e}\right] \Rightarrow \mu$. In Section E.7.3 below, we show that $u_{H_n^e}$ is concentrated around its mean, which completes the proof of Theorem 6.4.

Proof. Given the rooted hypertree (T, o) with depth at most d, let $\mathcal{C}_{[T,o]}$ denote the set of hypertrees $\tilde{T} \in \mathcal{T}(\mathbb{N}_{\text{vertex}}, \mathbb{N}_{\text{edge}})$ such that $(\tilde{T}, \emptyset) \equiv (T, o)$. For $\tilde{T} \in \mathcal{C}_{[T,o]}$, let $e_1^{\tilde{T}}, \dots, e_r^{\tilde{T}}$ be the edges in $\tilde{T}$ with depth at most $d-1$, ordered lexicographically in $\mathbb{N}_{\text{edge}}$. Moreover, let $e_{r+1}^{\tilde{T}}, \dots, e_{r+l}^{\tilde{T}}$ denote the edges in $\tilde{T}$ with depth precisely d, ordered lexicographically (if there is no such edge, $l = 0$). For a node $\mathbf{i}$ in $\tilde{T}$, let $\gamma_{\mathbf{i}}^{\tilde{T}}$ denote the type of the vertex $\mathbf{i}$ in the subtree below $\mathbf{i}$. Furthermore, define

$$\pi_{\tilde{T}} := \mathbb{P}\left((\mathcal{T}(\{\Gamma_a\}_{a \in \mathbb{N}_{\text{vertex}}}), \emptyset)_d = (\tilde{T}, \emptyset) \right),$$

under the probability in Definition 6.27. With the above notation, we have

$$\pi_{\tilde{T}} := P(\gamma_{\emptyset}^{\tilde{T}}) \prod_{t=1}^{r} \prod_{j=1}^{|e_t^{\tilde{T}}|-1} \hat{P}_{|e_t^{\tilde{T}}|}(\gamma_{(e_t^{\tilde{T}}, j)}^{\tilde{T}}). \tag{E.11}$$

By the definition of the distribution $\mathsf{UGWHT}(P)$, we have

$$\mu(A_{[T,o]}) = \sum_{\tilde{T} \in \mathcal{C}_{[T,o]}} \pi_{\tilde{T}}.$$

Recall that R_t denotes the hypertree in $\mathcal{T}(\mathbb{N}_{\text{vertex}}, \mathbb{N}_{\text{edge}})$ which results from the exploration process at step t. Note that if $(H_n, v_0)_d$ is a simple hypertree, we have $(H_n, v_0)_d \equiv (R_{\beta_n}, \emptyset)_d$ where β_n is defined in Proposition E.3 above. With this, we have

$$\begin{aligned} \mathbb{P}\left((H_n^e, v_0)_d \equiv (T, o)\right) = &\, \mathbb{P}\left((H_n^e, v_0)_d \equiv (T, o) \text{ and } (H_n, v_0)_d \text{ is a simple hypertree}\right) \\ &+ \mathbb{P}\left((H_n^e, v_0)_d \equiv (T, o) \text{ and } (H_n, v_0)_d \text{ is not a simple hypertree}\right) \end{aligned}$$

As is shown in Proposition E.3, the second term converges to zero; therefore, we need to study only the first term. But, if $(H_n, v_0)_d$ is a simple hypertree, $(H_n^e, v_0)_d \equiv (T, o)$ if and only if $R_{r+l+1} = \tilde{T}$ for some $\tilde{T} \in \mathcal{C}_{[T,o]}$. Consequently, it suffices for us to show that

$$\lim_{n \to \infty} \mathbb{P}\left(R_{r+l+1} = \tilde{T} \text{ and } (H_n, v_0)_d \text{ is a simple hypertree} \right) = \pi_{\tilde{T}} \qquad \forall \tilde{T} \in \mathcal{C}_{[T,o]}. \tag{E.12}$$

For $1 \le t \le r$, let E_t be the event defined in Proposition E.3. Recall that E_t^c is the event that $u_{t,1}, \dots, u_{t,k_t-1}$ are all distinct and are not in N_t. From Proposition E.3, we know that the probabilities of both the events "$(H_n, v_0)_d$ is a simple hypertree" and $\cap_{t=1}^r E_t^c$ converge to 1 as $n \to \infty$. Therefore, it suffices to show that

$$\lim_{n \to \infty} \mathbb{P}\left((R_{r+l+1} = \tilde{T}) \cap (\cap_{t=1}^r E_t^c) \right) = \pi_{\tilde{T}} \qquad \forall \tilde{T} \in \mathcal{C}_{[T,o]}. \tag{E.13}$$

To prove this, fix some $\tilde{T} \in \mathcal{T}(\mathbb{N}_{\text{vertex}}, \mathbb{N}_{\text{edge}})$ with depth at most d and define S_0 to be the event that $\gamma^{(n)}_{v_0} = \gamma^{\tilde{T}}_{\emptyset}$. From (6.83b), $\mathbb{P}(S_0) \to P(\gamma^{\tilde{T}}_{\emptyset})$ as $n \to \infty$. Moreover, for $1 \leq t \leq r$, define $\tilde{S}_t$ to be the event that

$$\gamma^{(n)}_{u_{t,s}} = \gamma^{\tilde{T}}_{(\mathbf{i}_t, k_t, \tilde{j}_t, s)} + \mathbb{e}_{k_t} \qquad 1 \leq s \leq k_t - 1,$$

and let $S_t = \tilde{S}_t \cap E^c_t$, which is in fact the intersection of $\tilde{S}_t$ and the event

$$u_{t,j}, 1 \leq j \leq k_t - 1 \text{ are distinct and are not in } N_t.$$

With this, let $S^t = \cap_{i=0}^t S_i$. We claim that the event $(R_{r+l+1} = \tilde{T}) \cap (\cap_{t=1}^r E^c_t)$ coincides with S^r. The reason is that on the event S^r, for each $1 \leq t \leq r$, the type of each of the $k_t - 1$ subnodes of edge formed at step t matches with that of $e^{\tilde{T}}_t$. Moreover, on the event $\cap_{t=1}^r E^c_t$, there is no improper edges or cycles formed during the exploration process. In particular, those partial edges connected to the vertex $u_{t,s}$ which are added to the active set are not used until the process goes to vertex $u_{t,s}$ itself. Also, note that as $\tilde{T}$ has depth at most d, its structure is determined by the type of the vertices of depth at most $d-1$, which are subnodes of edges of depth at most $d-1$ in $\tilde{T}$, which are precisely $e^{\tilde{T}}_1, \dots, e^{\tilde{T}}_r$.

Now, we prove by induction that for $0 \leq t \leq r$,

$$\mathbb{P}\left(S^t\right) \to \pi_{\tilde{T}}(t) := P(\gamma^{\tilde{T}}_{\emptyset}) \prod_{t'=1}^{t} \prod_{j=1}^{|e^{\tilde{T}}_{t'}|-1} \hat{P}_{|e^{\tilde{T}}_{t'}|}(\gamma^{\tilde{T}}_{(e^{\tilde{T}}_{t'}, j)}).$$

If $\pi_{\tilde{T}}(t-1) = 0$, then $\mathbb{P}(S^t) \leq \mathbb{P}(S^{t-1}) \to 0 = \pi_{\tilde{T}}(t)$. If $\pi_{\tilde{T}}(t-1) \neq 0$, we have $\mathbb{P}(S^{t-1}) > 0$ for n large enough. Note that $\mathbb{P}(S^t) = \mathbb{P}\left(S^{t-1} \cap \tilde{S}_t \cap E^c_t\right)$. Thereby,

$$\mathbb{P}\left(S^{t-1}\right) \mathbb{P}\left(\tilde{S}_t | S^{t-1}\right) - \mathbb{P}\left(E_t\right) \leq \mathbb{P}\left(S^t\right) \leq \mathbb{P}\left(S^{t-1}\right) \mathbb{P}\left(\tilde{S}_t | S^{t-1}\right).$$

But, we know that $\mathbb{P}(E_t) \to 0$. Consequently, it suffices to prove that

$$\mathbb{P}\left(\tilde{S}_t | S^{t-1}\right) \to \prod_{j=1}^{|e^{\tilde{T}}_t|-1} \hat{P}_{|e^{\tilde{T}}_t|}(\gamma^{\tilde{T}}_{(e^{\tilde{T}}_t, j)}). \tag{E.14}$$

Since we construct one edge at a time in the exploration process, conditioned on S^{t-1}, the first $t-1$ edges are constructed in a way consistent with $\tilde{T}$. Therefore, it is easy to see that

$$\mathbb{P}\left(\tilde{S}_t | S^{t-1}\right) = \mathbb{P}\left(\gamma^{(n)}_{u_{t,j}} = \gamma^{\tilde{T}}_{(e^{\tilde{T}}_t, j)} + \mathbb{e}_{|e^{\tilde{T}}_t|} \text{ for } 1 \leq j \leq |e^{\tilde{T}}_t| - 1\right).$$

For $1 \leq j \leq |e^{\tilde{T}}_t| - 1$, let $\tilde{S}_{t,j}$ denote the event that $\gamma^{(n)}_{u_{t,j}} = \gamma^{\tilde{T}}_{(e^{\tilde{T}}_t, j)} + \mathbb{e}_{|e^{\tilde{T}}_t|}$. Now, we study the probability of $\tilde{S}_{t,j}$ conditioned on S^{t-1} and $\tilde{S}_{t,1}, \dots, \tilde{S}_{t,j-1}$. Note that, having chosen

$e_{t,1},\dots,e_{t,j-1}$, there are $|\Delta^{(n)}(k_t)| - |\Delta^{(n)}(k_t)\cap C_t| - j$ many candidates for $e_{t,j}$, each having the same chance. With $B_{t,j} := \{e \in \Delta^{(n)}(|e_t^{\tilde{T}}|) : \gamma^{(n)}_{v(e)} = \gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)} + \mathbb{e}_{|e_t^{\tilde{T}}|}\}$, the event $\tilde{S}_{t,j}$ happens iff $e_{t,j}$ is chosen among the set $B_{t,j} \setminus (C_t \cup \{e_{t,1},\dots,e_{t,j-1}\})$. Therefore,

$$\mathbb{P}\left(\tilde{S}_{t,j}|S^{t-1},\tilde{S}_{t,1},\dots,\tilde{S}_{t,j-1}\right) = \frac{|B_{t,j}\setminus(C_t\cup\{e_{t,1},\dots,e_{t,j-1}\})|}{|\Delta^{(n)}(|e_t^{\tilde{T}}|)| - |\Delta^{(n)}(|e_t^{\tilde{T}}|)\cap C_t| - j}.$$

Note that

$$\frac{1}{n}|B_{t,j}| = \frac{1}{n}\sum_{i=1}^n (\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)}(|e_t^{\tilde{T}}|)+1)\mathbb{1}\left[\gamma^{(n)}_i = \gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)} + \mathbb{e}_{|e_t^{\tilde{T}}|}\right].$$

Using (6.83b), we have

$$\frac{1}{n}|B_{t,j}| \to (\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)}(|e_t^{\tilde{T}}|)+1)P(\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)} + \mathbb{e}_{|e_t^{\tilde{T}}|}).$$

On the other hand, conditioned on S^{t-1}, $|C_t| = \sum_{j=1}^{t-1}|e_j^{\tilde{T}}|$, which is a constant. Consequently,

$$\frac{1}{n}|B_{t,j}\setminus(C_t\cup\{e_{t,1},\dots,e_{t,j-1}\})| \to (\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)}(|e_t^{\tilde{T}}|)+1)P(\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)} + \mathbb{e}_{|e_t^{\tilde{T}}|}). \tag{E.15}$$

Moreover, using (6.85),

$$\frac{1}{n}|\Delta^{(n)}(|e_t^{\tilde{T}}|)| = \frac{1}{n}\sum_{i=1}^n \gamma^{(n)}_i(|e_t^{\tilde{T}}|) \to \mathbb{E}\left[\Gamma(|e_t^{\tilde{T}}|)\right]. \tag{E.16}$$

Note that we are conditioning on S^{t-1} and assuming that $\mathbb{P}(S^{t-1}) \neq 0$. On the other hand, $|e_t^{\tilde{T}}|$ is equal to the size of the partial edge e_t which is a member of $\Delta^{(n)}$. Using the assumption (6.83a), we have $|e_t^{\tilde{T}}| \in I$ and hence $\mathbb{E}\left[\Gamma(|e_t^{\tilde{T}}|)\right] > 0$. Putting (E.15) and (E.16) together, we have

$$\mathbb{P}\left(\tilde{S}_{t,j}|S^{t-1},\tilde{S}_{t,1},\dots,\tilde{S}_{t,j-1}\right) \to \frac{(\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)}(|e_t^{\tilde{T}}|)+1)P(\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)} + \mathbb{e}_{|e_t^{\tilde{T}}|})}{\mathbb{E}\left[\Gamma(|e_t^{\tilde{T}}|)\right]} = \hat{P}_{|e_t^{\tilde{T}}|}(\gamma^{\tilde{T}}_{(e_t^{\tilde{T}},j)})$$

Multiplying for $1 \le j \le |e_t^{\tilde{T}}| - 1$, we get (E.14) which completes the proof. □

E.7.3 Almost sure convergence

In this section we prove that, with the assumptions of Theorem 6.4, $u_{H_n^e} \Rightarrow \mathsf{UGWHT}(P)$ almost surely.

For a fixed n, Let $\Delta^{(n)}(k_1)\dots\Delta^{(n)}(k_L)$ be the nonempty sets among $\Delta^{(n)}(2),\dots,\Delta^{(n)}(n)$. From (6.84b) we know that $L \le c_1(\log n)^{c_2}$ and also $k_i \le c_1(\log n)^{c_2}$ for $1 \le i \le L$. For the sake of simplicity, write σ for $(\sigma_{k_1}, \dots, \sigma_{k_L})$ and M_i for $\mathcal{M}_{k_i}(\Delta^{(n)}(k_i))$, $1 \le i \le L$. From our construction, we know that σ_{k_i} is drawn uniformly at random from M_i and is independent from σ_{k_j}, $j \neq i$. With H_n^e being the simple hypergraph constructed by the configuration model, for a fixed $d > 0$ and a rooted tree (T, o) of depth at most d, define

$$F(\sigma) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(H_n^e, i)_d \equiv (T, i)}.$$

From Proposition E.4 we know that $\lim_{n\to\infty} \mathbb{E}[F(\sigma)] = \mu(A_{(T,o)})$. We will show that F is concentrated around its mean via a bounded difference argument.

Now, for each $1 \le j \le L$, fix a permutation $\pi_{k_j} \in M_j$ and define $\pi = (\pi_{k_1}, \dots, \pi_{k_L})$. Moreover, fix $1 \le i \le L$ and $e, e' \in \Delta^{(n)}(k_i)$. With this, define $\pi'_{k_i} := \text{swap}_{e,e'} \circ \pi_{k_i} \circ \text{swap}_{e,e'}$, which is the conjugation of π_{k_i} with the permutation that swaps e and e'. In fact, the cycle representation of π'_{k_i} is obtained by swapping e and e' in the cycle representation of π_{k_i}. Moreover, let $\pi' = (\pi_{k_1}, \dots, \pi'_{k_i}, \dots, \pi_{k_L})$ which differs from π only on the i^{th} coordinate. With this, the hypergraph obtained from π and the hypergraph obtained from π' differ only in at most two edges. Since all edge sizes and degrees in the graph are bounded to $c_1(\log n)^{c_2}$, there are at most $2(c_1(\log n)^{c_2})^{2d+1}$ many vertices in the hypergraph which have distance at most d to a vertex in any of these two edges. Consequently,

$$|F(\pi) - F(\pi')| \le \frac{2(c_1(\log n)^{c_2})^{2d+1}}{n}. \tag{E.17}$$

Now, fix $1 \le i \le L$ and $\pi_{k_j} \in M_j$ for $j \neq i$. Let σ_{k_i} being chosen uniformly at random in M_i and define $F_i(\sigma_{k_i}) = F(\pi_{k_1}, \dots, \sigma_{k_i}, \dots, \pi_{k_L})$. Since $\Delta^{(n)}(k_i)$ is finite, we can equip it with an arbitrary total order. Let X_1 be the smallest element in $\Delta^{(n)}(k_i)$ and define $Y_1 = (X_1, \sigma_{k_i}(X_1), \dots, \sigma_{k_i}^{(k_i-1)}(X_1))$, which is in fact the orbit of X_1, or in the configuration model language, the edge containing the partial edge X_1. Let X_2 be the smallest element that does not appear in Y_1 and let $Y_2 = (X_2, \sigma_{k_i}(X_2), \dots, \sigma_{k_i}^{(k_i-1)}(X_2))$. We continue this process inductively, i.e. let X_j be the smallest element that has not appeared in $Y_1, \dots, Y_{j-1}$ and let $Y_j = (X_j, \sigma_{k_i}(X_j), \dots, \sigma_{k_i}^{(k_i-1)}(X_j))$. This process yields $Y_1, \dots, Y_{|\Delta^{(n)}(k_i)|/k_i}$. For $1 \le j \le |\Delta^{(n)}(k_i)|/k_i$, let $\mathcal{F}_j$ be the sigma field generated by $Y_1, \dots, Y_j$. Moreover, let $Z_j = \mathbb{E}[F_i(\sigma_{k_i})|\mathcal{F}_j]$ for $1 \le j \le |\Delta^{(n)}(k_i)|/k_i$ and let $Z_0 = \mathbb{E}[F_i(\sigma_{k_i})]$. Note that π_{k_j}, $j \neq i$ are fixed; therefore, the randomness in the expression is with respect to σ_{k_i} only. Indeed, $(Z_j, 0 \le j \le |\Delta^{(n)}(k_i)|/k_i)$ is a martingale. We claim that, almost surely, we have

$$|Z_{j+1} - Z_j| \le k_i \frac{2(c_1(\log n)^{c_2})^{2d+1}}{n}.$$

The reason is that changing the value of the k_i variables in Y_{j+1} can change the value of F_i by at most $k_i \frac{2(c_1(\log n)^{c_2})^{2d+1}}{n}$ and the above inequality results from (E.17). Using Azuma's inequality and the fact that $k_i \le c_1(\log n)^{c_2}$, we have

$$\mathbb{P}(|F_i(\sigma_{k_i}) - \mathbb{E}[F_i(\sigma_{k_i})]| > \delta) < 2\exp\left(-\frac{\delta^2 n^2}{4|\Delta^{(n)}(k_i)|(c_1(\log n)^{c_2})^{4d+3}}\right). \tag{E.18}$$

To obtain an upper bound for $|\Delta^{(n)}(k_i)|$, note that

$$|\Delta^{(n)}(k_i)| = \left(\sum_{j=1}^n \gamma_j^{(n)}(k_i) \right) \leq n \left(\frac{1}{n} \sum_{j=1}^n \|\gamma_j^{(n)}\|_1^2 \right).$$

From (6.83c), there is a constant α independent of n and i that $\sum \|\gamma_j^{(n)}\|_1^2 < \alpha n$. Hence, $|\Delta^{(n)}(k_i)| < \alpha n$. Incorporating this into (E.18), we have, for $1 \leq i \leq L$,

$$\mathbb{P}\left(|F_i(\sigma_{k_i}) - \mathbb{E}\left[F_i(\sigma_{k_i})\right]| > \delta\right) < 2 \exp\left(- \frac{\delta^2 n}{4\alpha (c_1 (\log n)^{c_2})^{4d+3}} \right). \tag{E.19}$$

Since this is true for all i and π_{k_j}, $j \neq i$ and also the σ_{k_j} are independent, using the above inequality L times and using the fact that $L \leq c_1 (\log n)^{c_2}$, we have

$$\mathbb{P}\left(|F(\sigma) - \mathbb{E}\left[F(\sigma)\right]| > \delta\right) \leq 2 c_1 (\log n)^{c_2} \exp\left(- \frac{\delta^2 n}{4\alpha (c_1 (\log n)^{c_2})^{4d+3}} \right).$$

As the sum of the RHS over n is finite, using the Borel–Cantelli lemma and the fact that $\mathbb{E}\left[F(\sigma)\right] \to \mu(A_{(T,o)})$, we have $F(\sigma) \to \mu(A_{(T,o)})$ almost surely. But there are countably many choices for d and the rooted hypertree (T, o). Thus, outside a measure zero set, $u_{H_n^e}(A_{(T,o)}) \to \mu(A_{(T,o)})$ for all rooted tree (T, o) with finite depth. The proof is complete, using Lemma 6.4.

E.8 Proof of Proposition 6.9

Proof of Proposition 6.9. Since Θ'_ϵ is ϵ–balanced, from Definition 6.25, for $\vec{\mu}$–almost every $[H, e, i] \in \mathcal{H}_{**}$, we have

$$\Theta'_\epsilon(H, e, i) = \frac{\exp\left(- \frac{\partial \Theta'_\epsilon(H, i)}{\epsilon} \right)}{\sum_{j \in e} \exp\left(- \frac{\partial \Theta'_\epsilon(H, j)}{\epsilon} \right)}. \tag{E.20}$$

Using Proposition 6.3, there exists a $A \subset \mathcal{H}_{**}$ such that $\vec{\mu}(A^c) = 0$ and, for all $[H, e, i] \in A$, we have

$$\Theta'_\epsilon(H, e', i') = \frac{\exp\left(- \frac{\partial \Theta'_\epsilon(H, i')}{\epsilon} \right)}{\sum_{j \in e'} \exp\left(- \frac{\partial \Theta'_\epsilon(H, j)}{\epsilon} \right)} \qquad \forall \, (e', i') \in \Psi(H).$$

Now, fix some $[H, e, i] \in A$ and take an arbitrary element of this equivalence class $(H, e, i) \in [H, e, i]$. The above equation guarantees that if we define the allocation θ'^H_ϵ on H as $\theta'^H_\epsilon(e', i') := \Theta'_\epsilon(H, e', i')$ for $(e', i') \in \Psi(H)$, then θ'^H_ϵ is an ϵ–balanced allocation on H. Now, assume that $\theta^{H^\Delta}_\epsilon$ is the (unique) ϵ–balanced allocation on the truncated hypergraph H^Δ defined in Section 6.4.3 (uniqueness comes from boundedness of H^Δ). Proposition 6.5

then implies that $\partial\theta_\epsilon^{H^\Delta}(i') \leq \partial\theta'^H_\epsilon(i')$ for all $i' \in V(H)$. Sending Δ to infinity, this means that $\partial\theta^H_\epsilon(i') \leq \partial\theta'^H_\epsilon(i')$ for all $i' \in V(H)$, with θ^H_ϵ being the canonical ϵ–balanced allocation on H. Using Remark 6.12 and the definition of θ'^H_ϵ, this means that for $\vec{\mu}$–almost all $[H, e, i] \in \mathcal{H}_{**}$, $\partial\Theta_\epsilon(H, i) \leq \partial\Theta'_\epsilon(H, i)$. From part (ii) of Lemma 6.1, μ–almost surely we have

$$\partial\Theta_\epsilon \leq \partial\Theta'_\epsilon. \tag{E.21}$$

On the other hand, using unimodularity of μ and the fact that Θ_ϵ is a Borel allocation, we have

$$\int \partial\Theta_\epsilon d\mu = \int \Theta_\epsilon d\vec{\mu} = \int \nabla\Theta_\epsilon d\vec{\mu} = \int \frac{1}{|e|} d\vec{\mu}([H, e, i]).$$

Using the same logic, $\int \partial\Theta'_\epsilon d\mu = \int \frac{1}{|e|} d\vec{\mu}([H, e, i])$. This means that $\int \partial\Theta_\epsilon d\mu = \int \partial\Theta'_\epsilon d\mu$. As $\deg(\mu) < \infty$, this common value is finite. This, together with (E.21), implies that $\partial\Theta_\epsilon = \partial\Theta'_\epsilon$, μ–almost surely. Therefore, Proposition 6.3 implies that for μ–almost all $[H, i] \in \mathcal{H}_*$, we have $\partial\Theta_\epsilon(H, j) = \partial\Theta'_\epsilon(H, j)$ for all $j \in V(H)$. Then, part (i) of Lemma 6.1 implies that for $\vec{\mu}$–almost all $[H, e, i] \in \mathcal{H}_{**}$, $\partial\Theta_\epsilon(H, j) = \partial\Theta'_\epsilon(H, j)$ for all $j \in V(H)$. Thereby, using (E.20) for Θ_ϵ and Θ'_ϵ, we have $\Theta_\epsilon(H, e, i) = \Theta'_\epsilon(H, e, i)$ for $\vec{\mu}$–almost all $[H, e, i] \in \mathcal{H}_{**}$, which completes the proof. □

Bibliography

[Abb16] Emmanuel Abbe. Graph compression: The effect of clusters. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–8. IEEE, 2016.

[ADK12] Ferhat Ay, Michael Dang, and Tamer Kahveci. Metabolic network alignment in large scale by network compression. *BMC Bioinformatics*, 13(3):S2, 2012.

[AL07] David Aldous and Russell Lyons. Processes on unimodular random networks. *Electron. J. Probab*, 12(54):1454–1508, 2007.

[Ald01] David J Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures & Algorithms*, 18(4):381–418, 2001.

[AMO09] R.P. Agarwal, M. Meehan, and D. O'Regan. *Fixed Point Theory and Applications*. Cambridge Tracts in Mathematics. Cambridge University Press, 2009.

[Ang03] Omer Angel. Growth and percolation on the uniform infinite planar triangulation. *Geometric And Functional Analysis*, 13(5):935–974, 2003.

[AR14] David J Aldous and Nathan Ross. Entropy of some models of sparse random graphs with vertex-names. *Probability in the Engineering and Informational Sciences*, 28(02):145–168, 2014.

[AS03] Omer Angel and Oded Schramm. Uniform infinite planar triangulations. *Communications in mathematical physics*, 241(2-3):191–213, 2003.

[AS04] David Aldous and J Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

[AS16] Venkat Anantharam and Justin Salez. The densest subgraph problem in sparse random graphs. *The Annals of Applied Probability*, 26(1):305–327, 2016.

[BB12] François Baccelli and Pierre Brémaud. *Palm probabilities and stationary queues*, volume 41. Springer Science & Business Media, 2012.

[BBH+98] Krishna Bharat, Andrei Broder, Monika Henzinger, Puneet Kumar, and Suresh Venkatasubramanian. The connectivity server: Fast access to linkage information on the web. *Computer networks and ISDN Systems*, 30(1-7):469–477, 1998.

[BC15] Charles Bordenave and Pietro Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probability Theory and Related Fields*, 163(1-2):149–222, 2015.

[BCCZ19] Christian Borgs, Jennifer Chayes, Henry Cohn, and Yufei Zhao. An l^p theory of sparse graph convergence i: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, 372(5):3019–3062, 2019.

[BH18] Maciej Besta and Torsten Hoefler. Survey and taxonomy of lossless graph compression and space-efficient graph representations. *arXiv preprint arXiv:1806.01799*, 2018.

[BHMK20] François Baccelli, Mir-Omid Haji-Mirsadeghi, and Ali Khezeli. Unimodular hausdorff and minkowski dimensions. *arXiv preprint arXiv:1807.02980*, 2020.

[Bil71] Patrick Billingsley. Weak convergence of measures. *Philadelphia: SIAM*, 1971.

[Bil13] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[BKM+00] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer networks*, 33(1-6):309–320, 2000.

[BL10] Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures & Algorithms*, 37(3):332–352, 2010.

[Bol98] Béla Bollobás. Random graphs. In *Modern Graph Theory*, pages 215–252. Springer, 1998.

[Bor14] Charles Bordenave. Lecture notes on random graphs and probabilistic combinatorial optimization. 2014. http://www.math.univ-toulouse.fr/~bordenave/coursRG.pdf.

[BRSV11] Paolo Boldi, Marco Rosa, Massimo Santini, and Sebastiano Vigna. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *Proceedings of the 20th international conference on World wide web*, pages 587–596. ACM, 2011.

[BS01] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp. (electronic), 2001.

[BV04] Paolo Boldi and Sebastiano Vigna. The webgraph framework i: compression techniques. In *Proceedings of the 13th international conference on World Wide Web*, pages 595–602. ACM, 2004.

[BV17] Sourya Basu and Lav R Varshney. Universal source coding of deep neural networks. In *2017 Data Compression Conference (DCC)*, pages 310–319. IEEE, 2017.

[Cho86] S. A. Choudum. A simple proof of the Erdos-Gallai theorem on graph sequences. *Bulletin of the Australian Mathematical Society*, 33:67–70, 1986.

[CK11] Imre Csiszar and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[CKL$^+$09] F Chierichetti, R Kumar, S Lattanzi, M Mitzenmacher, A Panconesi, and P Raghavan. On compressing social networks in: Proc. of 15th conference on knowledge discovery and data mining (kdd'09), 2009.

[CS12] Yongwook Choi and Wojciech Szpankowski. Compression of graphical structures: Fundamental limits, algorithms, and experiments. *IEEE Transactions on Information Theory*, 58(2):620–638, 2012.

[CT12] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[DA17a] Payam Delgosha and Venkat Anantharam. Load balancing in hypergraphs. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 45–52. IEEE, 2017.

[DA17b] Payam Delgosha and Venkat Anantharam. Universal lossless compression of graphical data. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1578–1582. IEEE, 2017.

[DA18a] Payam Delgosha and Venkat Anantharam. Distributed compression of graphical data. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 2216–2220. IEEE, 2018.

[DA18b] Payam Delgosha and Venkat Anantharam. Load balancing in hypergraphs. *Journal of Statistical Physics*, 173(3-4):546–625, 2018.

[DA20] Payam Delgosha and Venkat Anantharam. Universal lossless compression of graphical data. *IEEE Transactions on Information Theory*, 2020.

[DWvW12] Kasper Dinkla, Michel A Westenberg, and Jarke J van Wijk. Compressed adjacency matrices: Untangling gene regulatory networks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2457–2466, 2012.

[EG60] Pál Erdös and Tibor Gallai. Gráfok elöírt fokú pontokkal. *Matematikai Lapok*, 11:264–274, (in Hungarian) 1960.

[Gam04] David Gamarnik. Linear phase transition in random linear constraint satisfaction problems. *Probability Theory and Related Fields*, 129(3):410–440, 2004.

[GHLB19] Moses Ganardi, Danny Hucke, Markus Lohrey, and Louisa Seelbach Benkner. Universal tree source coding using grammar-based compression. *IEEE Transactions on Information Theory*, 65(10):6399–6413, 2019.

[GNS06] David Gamarnik, Tomasz Nowicki, and Grzegorz Swirszcz. Maximum weight independent sets and matchings in sparse random graphs. exact results using the local weak convergence method. *Random Structures & Algorithms*, 28(1):76–106, 2006.

[Haj90] Bruce Hajek. Performance of global load balancing by local adjustment. *Information Theory, IEEE Transactions on*, 36(6):1398–1414, 1990.

[Haj96] Bruce Hajek. Balanced loads in infinite networks. *The Annals of Applied Probability*, 6(1):48–75, 1996.

[HPP16] Morteza Hosseini, Diogo Pratas, and Armando J Pinho. A survey on data compression methods for biological sequences. *Information*, 7(4):56, 2016.

[HY10] Siu-Wai Ho and Raymond W Yeung. The interplay between entropy and variational distance. *IEEE Transactions on Information Theory*, 56(12):5906–5929, 2010.

[KK14] M Kamal and Mohammad Ibrahim Khan. Memory optimization for global protein network alignment using pushdown automata and de bruijn graph based bloom filter. *Journal of Software*, 9(10):2622–2628, 2014.

[KYS09] John C Kieffer, E-H Yang, and Wojciech Szpankowski. Structural complexity of random binary trees. In *2009 IEEE International Symposium on Information Theory*, pages 635–639. IEEE, 2009.

[ŁMS19] Tomasz Łuczak, Abram Magner, and Wojciech Szpankowski. Asymmetry and structural information in preferential attachment graphs. *Random Structures & Algorithms*, 55(3):696–718, 2019.

[Lyo05] Russell Lyons. Asymptotic enumeration of spanning trees. *Combinatorics, Probability and Computing*, 14(4):491–522, 2005.

[Mas12] Hossein Maserrat. *Compression of social networks*. PhD thesis, Applied Science: School of Computing Science, 2012.

[McK85] Brendan D McKay. Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars Combin*, 19:15–25, 1985.

[MP10] Hossein Maserrat and Jian Pei. Neighbor query friendly compression of social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 533–542, 2010.

[MTS18] Abram Magner, Krzysztof Turowski, and Wojciech Szpankowski. Lossless compression of binary trees with correlated vertex names. *IEEE transactions on information theory*, 64(9):6070–6080, 2018.

[Mun00] James R Munkres. *Topology*. Prentice Hall, 2000.

[PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[SMHM99] Craig Silverstein, Hannes Marais, Monika Henzinger, and Michael Moricz. Analysis of a very large web search engine query log. In *ACm SIGIR Forum*, volume 33, pages 6–12. ACM New York, NY, USA, 1999.

[SS07] Moshe Shaked and J George Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007.

[SSA+16] Siavash Sheikhizadeh, M Eric Schranz, Mehmet Akdel, Dick de Ridder, and Sandra Smit. Pantools: representation, storage and exploration of pan-genomic data. *Bioinformatics*, 32(17):i487–i493, 2016.

[Ste02] J Michael Steele. Minimal spanning trees for graphs with random edge lengths. In *Mathematics and Computer Science II*, pages 223–245. Springer, 2002.

[Tao11] T. Tao. *An Introduction to Measure Theory*. Graduate studies in mathematics. American Mathematical Society, 2011.

[TMS20] Krzysztof Turowski, Abram Magner, and Wojciech Szpankowski. Compression of dynamic graphs generated by a duplication model. *Algorithmica*, 2020.

[VDH16] Remco Van Der Hofstad. *Random graphs and complex networks*, volume 1. Cambridge university press, 2016.

[Wel84] Terry A. Welch. A technique for high-performance data compression. *Computer*, (6):8–19, 1984.

[ZL77] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, 1977.

[ZL78] Jacob Ziv and Abraham Lempel. Compression of individual sequences via variable-rate coding. *IEEE transactions on Information Theory*, 24(5):530–536, 1978.

[ZYK13] Jie Zhang, En-Hui Yang, and John C Kieffer. A universal grammar-based code for lossless compression of binary trees. *IEEE Transactions on Information Theory*, 60(3):1373–1386, 2013.