2021

Exponential Random Graphs and a Generalization of Parking Functions

Ryan DeMuse
University of Denver

Exponential Random Graphs and a Generalization of Parking Functions

Abstract

Random graphs are a powerful tool in the analysis of modern networks. Exponential random graph models provide a framework that allows one to encode desirable subgraph features directly into the probability measure. Using the theory of graph limits pioneered by Borgs et. al. as a foundation, we build upon the work of Chatterjee & Diaconis and Radin & Yin. We add complexity to the previously studied models by considering exponential random graph models with edge-weights coming from a generic distribution satisfying mild assumptions. In particular, we show that a large family of two-parameter, edge-weighted exponential random graphs display a phase transtion and identify the limiting behavior of such graphs in the dual space provided by the Legendre-Fenchel transform.

For finite systems, we analyze the mixing time of exponential random graph models. The mixing time of unweighted exponential random graphs was studied by Bhamidi, Bresler, and Sly. We extend upon the work of Levin, Luczak, and Peres by studying the Glauber dynamics of a certain vertex-weighted exponential random graph model on the complete graph. Specifically, we identify regions of the parameter space where the mixing time is $\Theta(n \log n)$ and where it is exponentially slow.

Toward the end of this work, we take a drastic turn in a different direction by studying a generalization of parking functions that we call interval parking functions. Parking functions are a classical combinatorial object dating back to the work of Konheim and Weiss in the 1960s. Among other things, we explore the connections that bioutcomes of interval parking functions have to various partial orders on the symmetric group on n letters including the (left) weak order, (strong) Bruhat order, and the bubble-sorting order.

Document Type

Dissertation

Degree Name

Ph.D.

First Advisor

Mei Yin

Second Advisor

Paul Rullkoetter

Third Advisor

Ronnie Pavlov

Keywords

Bruhat order, Exponential random graph model, Interval parking functions, Parking functions, Random graphs, Sorting order

Subject Categories

Mathematics | Other Mathematics | Physics | Statistical, Nonlinear, and Soft Matter Physics

Exponential Random Graphs and a Generalization of Parking Functions

---

A Dissertation

Presented to

the Faculty of the College of Natural Sciences and Mathematics

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Ryan DeMuse

June 2021

Advisor: Dr. Mei Yin

Author: Ryan DeMuse
Title: Exponential Random Graphs and a Generalization of Parking Functions
Advisor: Dr. Mei Yin
Degree Date: June 2021

ABSTRACT

Random graphs are a powerful tool in the analysis of modern networks. Exponential random graph models provide a framework that allows one to encode desirable subgraph features directly into the probability measure. Using the theory of graph limits pioneered by Borgs et. al. as a foundation, we build upon the work of Chatterjee & Diaconis and Radin & Yin. We add complexity to the previously studied models by considering exponential random graph models with edge-weights coming from a generic distribution satisfying mild assumptions. In particular, we show that a large family of two-parameter, edge-weighted exponential random graphs display a phase transtion and identify the limiting behavior of such graphs in the dual space provided by the Legendre-Fenchel transform.

For finite systems, we analyze the mixing time of exponential random graph models. The mixing time of unweighted exponential random graphs was studied by Bhamidi, Bresler, and Sly. We extend upon the work of Levin, Luczak, and Peres by studying the Glauber dynamics of a certain vertex-weighted exponential random graph model on the complete graph. Specifically, we identify regions of the parameter space where the mixing time is $\Theta(n \log n)$ and where it is exponentially slow.

Toward the end of this work, we take a drastic turn in a different direction by studying a generalization of parking functions that we call interval parking functions. Parking functions are a classical combinatorial object dating back to the work of Konheim and Weiss in the 1960s. Among other things, we explore the connections that bioutcomes of interval parking functions have to various partial orders on the symmetric group on n letters including the (left) weak order, (strong) Bruhat order, and the bubble-sorting order.

ACKNOWLEDGEMENTS

This dissertation is dedicated to Danielle. She is always there for me. She is, and always will be, my best friend. There is no way that I can adequately express my gratitude to her for allowing me to dedicate six years of our life together to my academic pursuits, the support that she provides every single day, or the way that she completes me.

First and foremost, I bestow eternal gratitude upon my advisor, Mei Yin. It is a privilege to be Mei's first doctoral student and I count myself lucky to have her as my advisor. I especially want to thank her for guiding me in my research, no matter which direction it led. She is always open to indulge my mathematical curiosity and is equally content with chatting as a friend.

I would especially like to thank the members of my graduate committee. I realize how much of a commitment it is to be a doctoral committee chair and I would like to thank Paul Rullkoetter for being gracious enough to accept the role. I would like to thank Ronnie Pavlov, who has been my teaching mentor throughout my years at the University of Denver. He has provided me with valuable critiques and insights that have improved every aspect of my teaching. I would also like to thank Paul Horn. Paul mentored me in my application to the Graduate Research Workshop in Combinatorics at the University of Kansas in 2019. Without his invaluable guidance, I would not have attended the workshop and Chapters 4 and 5 would not have been possible. Lastly, I would like to thank Alvaro Arias. Alvaro was my professor for analysis and I had the pleasure of being his TA for several classes. Alvaro teaches with such joy and passion that I am inspired to savor every moment that I have in the mathematical community.

Lastly, I would like to thank my parents, Tom and Stacey, and my sister, Abby. My parents have always emphasized the importance of education and supported my journey in so many ways, I cannot thank them enough.

DECLARATIONS

Some chapters of this dissertation are based on work that has been published. These chapters are as follows:

1. Chapter 2 is based on published joint work with Danielle Larcomb and Mei Yin. The paper [26] was published in the *Journal of Statistical Physics*.

2. Chapter 3 is based on published joint work with Terry Easlick and Mei Yin. The paper [25] was published in the *Journal of Computational and Applied Mathematics*.

3. Chapters 4 and 5 is based on published joint work with Emma Colaric, Jeremy L. Martin, and Mei Yin. The paper [23] was published in *Advances in Applied Mathematics*.

None of the results herein have appeared in any other dissertation or thesis, and all coauthors have agreed to the inclusion of these joint works in this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOTATION

$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	natural numbers, integers, real numbers
$[\![a, b]\!]$	discrete interval, $[a, b] \cap \mathbb{N}$
$\mathfrak{S}_n$	symmetric group on n letters
$\mathcal{G}_n$	set of simple, labelled graphs on n vertices
$\overline{\mathcal{G}}_n$	set of edge-weighted, simple, labelled graphs on n vertices
$\dot{\mathcal{G}}_n$	set of vertex-weighted, simple, labelled graphs on n vertices
G_n	graph on n vertices
$E(G)$	set of edges in a graph G
$V(G)$	set of vertices in a graph G
$\mathbf{P}$, $\mathbf{Q}$	probability measure
$\mathbf{E_P}$	expectation with respect to the probability measure $\mathbf{P}$[†]
$\mathbf{Var_P}$	variance with respect to the probability measure $\mathbf{P}$[†]
iid	independent and identically distributed
$O(f)$	$g = O(f)$ if and only there exists a constant C and $M \in \mathbb{R}$ such that $\lvert g(x)\rvert \leq Cf(x)$ for all $x \geq M$
$\Omega(f)$	$g = \Omega(f)$ if and only if $f = O(g)$
$\Theta(f)$	$g = \Theta(f)$ if and only if $f = O(g)$ and $g = O(f)$
$f \asymp g$	$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$

[†]The associated probability measure is omitted when it is clear from the context.

Introduction

Organization

This dissertation is primarily concerned with two different topics: exponential random graph models and interval parking functions, and is therefore divided into two main parts. Chapters 1 to 3 pertains to the analysis of exponential random graph models. Chapters 4 and 5 introduces and studies interval parking functions. We provide a brief motivation for both of the topics covered herein.

Exponential random graphs. A pressing problem in the social sciences is the study of graphs representing a friendship network on social networks such as Facebook or Twitter. These networks are often too large and volatile for any practical direct analysis, but there are certain characteristics that are typically present in such a network. For example, we can assume that if Emily and Alice are friends and Alice and Danielle are friends, then it is more likely that Emily and Danielle are friends since they have a mutual friend. This is a network feature that we will refer to as *transitivity*; the fact that a friend of a friend is likely to be a friend. Another example is the phenomenon that people tend to have distinct friend groups. A person may have a friend group among colleagues and a friend group among family, each of which is highly connected within themselves, but it is less likely that there are many connections between a person's family and their colleagues. Random graphs are important tools in the investigation of modern networks. Randomness is a powerful assumption when addressing questions about what a graph may look like when only indirect or partial analysis is possible.

Complex networks that exhibit these properties have become omnipresent structures, especially with the popularity of technological and social networks, but also in economics

and healthcare [39, 63, 51]. The ubiquity of networks has hastened the need to develop new models that aid in the study of these elaborate structures and techniques to study both their local and global properties as well as how they change over time. Chapters 1 to 3 analyze probabilistic models of networks; specifically, exponential random graph models, which can be tuned to possess desirable properties by encoding certain local structures directly into the model.

Interval parking functions. Parking functions comprise a central object of study in combinatorics. Since their introduction in the 1960s, many connections to other combinatorial objects such as trees, lattice paths, Prüfer codes, etc. have been uncovered and several generalizations have been introduced. Classic parking functions consider n cars entering a one-way street with n labelled, open parking spaces. Each car has a preferred place that they wish to park given by a, a preference sequence of length n. If that spot is open, the car parks there, if not, they park in the next open spot. If no spot is open after their preferred spot, they drive away. If all cars park successfully, then a is a parking function.

In Chapters 4 and 5, we study another generalization of the classic object that we call *interval parking functions*. These differ from classic parking functions by not only specifying a spot where cars are willing to start parking along a one-way street, but also by providing a spot after which cars are no longer willing to park. This definition is motivated by the following scenario. Consider a group of cars travelling down 5th Ave in New York City between the 200th and 100th blocks, a one-way street flowing in the direction of smaller block numbers, in the heart of the Flatiron District. Suppose that one of the cars wishes to visit a storefront on the 160th block of 5th Ave. It is a reasonable assumption that these individuals do not want to walk 6 blocks from their parking spot to the store they intend to visit, if it can be helped. Instead, they would prefer to park in a smaller radius around the 160th block, say anywhere from the 180th block to the 140th block. Thus we would think of the 180th block as the beginning of the area in which they are willing to

park and the 140th block as the end. This extension of ordinary parking functions adds additional complexity while also being a quite intuitive generalizaiton based on the classic parking algorithm and many natural questions are investigated. How many interval parking functions of length n are there? When is a pair (a, b) an interval parking function? How do a and b affect where the cars end up parking?

Overview of chapters

This dissertation is divided into two parts. The first is intended to expand upon the knowledge of exponential random graph models by adding various forms of complexity to the model. Chapter 1 introduces the background and notation necessary for our study in Chapters 1 to 3. We will then investigate the phase transition that appears in certain two-parameter, edge-weighted exponential random graph models in Chapter 2. Specifically, we will rigorously define what a phase transition is, then we will investigate other asymptotic behavior of this edge-weighted model. In Chapter 3, we will explore the Glauber dynamics on vertex-weighted exponential random graphs. In particular, we determine the mixing time associated with the Glauber dynamics and identify where in the parameter space the mixing is (quasilinearly) fast and (exponentially) slow.

Chapters 4 and 5 studies several aspects of interval parking functions. Chapter 4 introduces parking functions and interval parking functions as well as basic aspects of the symmetric group as a Coxeter system and various partial orders that arise in our study of interval parking functions. The content of Chapter 5 gives an enumeration of the number of interval parking functions of length n, their relation to the left weak order and strong Bruhat order on Coxeter groups, and reveals a close connection between the outcomes of interval parking functions and the bubble sort order on the symmetric group, an instance of Armstrong's sorting order [9]. Several open questions are also presented and discussed.

Chapter 1: Introduction to exponential random graphs

1.1 Background

The first random graph models were proposed by Paul Erdős and Alfréd Rényi [33], and, contemporaneously and independently, by Edgar Gilbert [37]. This model, $G(n,p)$, considers n vertices with edges placed between distinct pairs of vertices independently with identical probability p. This description leads to a probability measure $\mathbf{P}_n^p$ on the set of simple, labelled graphs on n vertices, denoted by $\mathcal{G}_n$, where

$$\mathbf{P}_n^p(G) = p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}, \tag{1.1}$$

and $e(G) = |E(G)|$ is the number of edges in G with $E(G)$ being the set of edges in the graph $G \in \mathcal{G}_n$. Note that this construction can be generalized to considering configurations on underlying graph structures other than the complete graph, another common underlying structure is the d-dimensional integer lattice $\mathbb{Z}^d$.

Several natural questions about this model can then be investigated. What is the probability that for a given n and p a graph drawn from this particular distribution is connected? For a fixed k, what is the probability that for a given n and p a sampled graph has minimum degree $\geq k$? Questions of this form lead to intriguing results on the asymptotic behavior of graphs generated in this way. For values of p close to 0, the graph almost surely contains isolated vertices, while as p increases to $\log n/n$, the graph almost surely becomes connected. This abrupt structural change in the graph behavior is referred to as a *phase transition* and resembles the liquid to gas transition in physics when the temperature increases.

Much of the random graph literature has evolved from the famous Erdős-Rényi graph. While its simple formation has attracted significant mathematical interest, this construction lacks the ability to model real world networks, which exhibit many noticeable attributes such as clustering and transitivity. Since these are known to be present and defining features in large real-world networks, their omission in $G(n,p)$ makes it a less realistic model for modern networks. One of the first attempts to complicate random graph models was the effort of Fortuin and Kasteleyn. They created the *random cluster model* to encourage or discourage the formation of many connected components [34]. This is done by taking the model $G(n,p)$ and adding another parameter that gives weight to the number of connected components in the graph G,

$$\mathbf{P}_{n,c}^{p,q}(G) = Z^{-1} q^{k(G)} p^{e(G)} (1-p)^{\binom{n}{2}-e(G)}, \tag{1.2}$$

where $k(G)$ is the number of connected components in the graph G, $q > 0$, and Z is the *partition function*

$$Z = \sum_{G \in \mathcal{G}_n} q^{k(G)} p^{e(G)} (1-p)^{\binom{n}{2}-e(G)}. \tag{1.3}$$

Values of $q < 1$ favors graphs with fewer connected components while $q > 1$ favors those with more. When $q = 1$, we recover the classical $G(n,p)$. This model is of significant historical importance as it unifies the percolation, Ising, and Potts models in a single framework [34].

Another generalization of the Erdős-Rényi model on the path toward exponential random graphs is Jonasson's *random triangle model* [40]. The random triangle model aims to capture the importance of clustering and transitivity by introducing another parameter into the measure that counts the number of triangles in the graph G,

$$\mathbf{P}_{n,t}^{p,q}(G) = Z^{-1} q^{t(G)} p^{e(G)} (1-p)^{\binom{n}{2}-e(G)}, \tag{1.4}$$

where $q \geq 1$, $t(G)$ counts the number of triangles in G, and Z is similarly defined as in Equation (1.3) where $q^{k(G)}$ is replaced with $q^{t(G)}$.

Many more random graph models have been proposed and studied since the introduction of the Erdős-Rényi model. These models include Barabási-Albert [3], Watts-Strogatz [68], random geometric graphs [54], stochastic block model [1], etc. Exponential random graph models seek to unify many concepts within these models into a single framework by considering more generalized combinations of functions from the graph space to the reals.

Our central object of study in Chapters 1 to 3, the *exponential random graph model (ERGM)*, seeks to incorporate the known properties displayed by modern networks. The general form of measure for exponential random graphs is

$$\mathbf{P}_n^T(G) = Z^{-1} e^{T(G)}, \tag{1.5}$$

where Z is the normalization constant and T is a suitably chosen function from the graph space to the reals. The ERGM captures global network tendencies through local attributes and, by building the local attributes into the probability measure, yields graphs demonstrating the desired properties with a higher probability than those that do not. This is achieved by the following principle: Instead of using a single parameter p to indicate edge presence probability, exponential random graphs include in the exponent a linear combination of finite subgraphs coupled with positive or negative parameters to encourage or discourage the formation of these subgraph structures.

The concept of Markov graphs were first studied by Frank and Strauss [35]. Markov graphs are a special case of exponential random graphs where the only finite subgraphs considered are stars and triangles. Wasserman and Pattison [70] extended this framework by considering general subgraph counts. Inquiries into exponential random graphs have been made on the variational principle of the limiting normalization constant, concentration

of the limiting probability distribution, phase transitions, and asymptotic structures. See for example Chatterjee and Varadhan [22], Chatterjee and Diaconis [21], Radin and Yin [60], Lubetzky and Zhao [48] [49], Radin and Sadun [58] [59], Radin et al. [57], Kenyon et al. [41], Yin [72], Kenyon and Yin [42], Aristoff and Zhu [8], and Chatterjee and Dembo [20]. Many of these papers utilize the elegant theory of graph limits as developed by Lovász and coauthors (V.T. Sós, B. Szegedy, C. Borgs, J. Chayes, K. Vesztergombi, ...) [16] [17] [18] [46] [47]. Building on earlier work of Aldous [4] and Hoover [38], the graph limit theory creates a new set of tools for representing and studying the asymptotic behavior of graphs by connecting sequences of graphs G_n, which are discrete objects that lie in different probability spaces, to a unified graphon space $\mathcal{W}$, which is an abstract functional space equipped with a cut metric. Though the theory itself is tailored to dense graphs, parallel theories for sparse graphs are likewise emerging. See Benjamini and Schramm [11], Aldous and Steele [7], Aldous and Lyons [6], and Lyons [50] where the notion of local weak convergence is discussed and the works of Borgs et al. [14] [15] that are making progress towards enriching the existing L^∞ theory of dense graph limits by developing a limiting object for sparse graph sequences based on L^p graphons.

1.2 Another random graph model

As mentioned in the previous section, there are many different models for generating random graphs. Exponential random graphs are widely used in the social science literature as they seem to be appropriate for studying relationships among individuals in a social setting, while other random graph models may find widespread use in other disciplines. It should be noted that, when comparing different random graph models, context can be very important – it is strange to say that a hammer is always better than a screwdriver, but it is rather a matter of having the right tool for the job. We now briefly describe another random graph model and how it differs from ERGMs.

1.2.1 Stochastic block model. Many random graph models have been introduced as generalizations of $G(n,p)$. One such random graph model that builds upon $G(n,p)$ is the *stochastic block model* (SBM). Letting n be a predetermined number of vertices, one chooses a partition $\{\mathcal{C}_i\}_{i=1,\dots,m}$ of the vertex set where each $\mathcal{C}_i$ is referred to as a *community*, and a symmetric $m \times m$ matrix A with entries from $[0,1]$ representing edge presence probabilities. A graph is then generated as follows: any two vertices $u \in \mathcal{C}_i$ and $v \in \mathcal{C}_j$ are connected with probability A_{ij}. It is rather quick to see that the constant matrix $A_{ij} = p$ recovers the classic model $G(n,p)$, thereby rendering the vertex set partition meaningless. More interesting cases occur when the main diagonal of A is taken to be constant and off-diagonal entries differ. This simulates the situation where connections within communities are more (or less) likely than connections between distinct communities.

Stochastic block models have enjoyed wide use within the machine learning community [1]. A typical problem concerning these models is the community detection problem, that is, given a graph on n vertices, recover the communities $\mathcal{C}_1, \dots, \mathcal{C}_m$. One advantage that this model has over ERGMs is that it is a *generative model*, that is, the model gives a procedure for generating a random instance of the model. ERGMs are *static models* and do not admit a simple description for generating a random instance instead requiring sampling techniques such as Markov chain Monte Carlo. Eldan and Gross first studied the connection between exponential random graphs and the stochastic block model [32]. They concluded that for dense graphs, that is, when the number of edges is on the order of n^2, ERGMs behave like a mixture of stochastic block models. This work was then expanded by the author and their advisor to the case of vertex-weighted graphs [27].

The added complexity of SBM makes it a better candidate over Erdős-Rényi for modeling modern networks, but it still suffers from the same edge independence assumption $G(n,p)$. Since ERGMs do not make an assumption about edge independence, they are able to model more complex interconnected relationships between individuals in a network than

SBMs; however, this same assumption, which makes them a good candidate for modeling relationships in a network, is exactly what makes ERGMs difficult to analyze.

1.3 Graph limits

We now move on to introducing the necessary tools for the construction of ERGM probability measures. In order to construct these measures on $\mathcal{G}_n$ (and, similarly, on the spaces of edge-weighted and vertex-weighted graphs), we first illustrate how to record features of various finite subgraphs. Recall that a *graph* is a pair $G = (V, E)$ of vertices V and (undirected) edges $E \subset V \times V$ are (unordered) pairs of vertices that represent a connection between them. We write $V(G)$ and $E(G)$ to denote the vertex and edge sets of the graph G. We will supress the dependence on G when it is clear from the context. For a graph with n vertices ($|V(G)| = n$), we will take $V(G) = \{1, \ldots, n\}$ unless otherwise noted. By a *simple* graph H we mean that H has at most one edge between any two vertices and there are no edges from any vertex v back to itself. All graphs considered herein will be simple unless ortherwise noted. With this in mind, let H and G be finite simple graphs.

Definition 1.1. A *graph homomorphism* from H to G is a map $\phi : V(H) \to V(G)$ such that if $\{u, v\} \in E(H)$ then $\{\phi(u), \phi(v)\} \in E(G)$. Let $\mathrm{Hom}\,(H, G)$ be the set of all graph homomorphisms from H to G and $\hom(H, G) = |\mathrm{Hom}\,(H, G)|$. The *homomorphism density of H in G*, denoted by $t(H, G)$, is defined to be

$$t(H, G) = \frac{\hom(H, G)}{|V(G)|^{|V(H)|}}. \tag{1.6}$$

The homomorphism density $t(H, G)$ may be thought of as the probability that a random vertex map $V(H) \to V(G)$ is edge-preserving. Through documentation of the density value associated with different subgraphs H, local information of the graph G is encoded.

One of the questions that we consider in Chapter 2 is: how different are the exponential random graph model and $G(n, p)$ as the number of vertices increase? In order to address

this question, we require a notion that captures *similarity* in growing graphs. Suppose that $\{G_n\}$ is a sequence of graphs that become more and more similar as n tends to infinity, in the sense that $t(H, G_n)$ approaches some limit $t(H)$ for every finite simple graph H. Lovász and coauthors [16, 17, 18] identified a limiting object for such a sequence $\{G_n\}$ in rigorous mathematical terms. It is represented by a function $g \in \mathcal{W}$, where $\mathcal{W}$ is the space of all symmetric measurable functions $g : [0,1]^2 \to [0,1]$, i.e. $g(x,y) = g(y,x)$ for all $x, y \in [0,1]$. We refer to $\mathcal{W}$ as the *graphon space* and elements of $\mathcal{W}$ are called *graphons* or *graph limits*. Intuitively, we view $[0,1]$ as a continuum of vertices and $g(x,y)$ as the edge weight between graph vertices x and y (in the case of graphs without edge weights, we take $g(x,y) = 1$ to indicate the presence and $g(x,y) = 0$ to indicate the absence of an edge). The limiting density $t(H)$ may be read off from the limiting object g in the following manner. If H is a finite simple graph with $V(H) = [\![m]\!] = \{1, \ldots, m\}$, define

$$t(H, g) = \int_{[0,1]^m} \prod_{\{i,j\} \in E(H)} g(x_i, x_j)\, dx_1 \cdots dx_m. \tag{1.7}$$

Definition 1.2. A sequence of graphs $\{G_n\}$ converges to a graphon $g \in \mathcal{W}$ if for every finite simple graph H,

$$\lim_{n \to \infty} t(H, G_n) = t(H, g). \tag{1.8}$$

For any finite simple graph G on n vertices, a graphon representation of G may be constructed:

$$g^G(x,y) = \begin{cases} 1 & \{\lceil nx \rceil, \lceil ny \rceil\} \in E(G) \\ 0 & \text{otherwise.} \end{cases} \tag{1.9}$$

The definition is consistent because $t(H, G) = t(H, g^G)$ for any finite simple graph H where $t(H, G)$ is defined in Definition 1.1 and $t(H, g^G)$ in Equation (1.7).

Proposition 1.1. Let $H, G \in \mathcal{G}_n$. Then $t(H, G) = t(H, g^G)$.

Proof. Without loss of generality, let $V(H) = [\![m]\!]$. Then

$$\begin{aligned} t(H, g^G) &= \int_{[0,1]^m} \prod_{\{i,j\} \in E(H)} g(x_i, x_j)\, dx_1 \cdots dx_m \\ &= \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \prod_{\{a,b\} \in [m]} g(i_a/n, i_b/n) \\ &= \frac{\hom(H, G)}{n^m} = t(H, G). \end{aligned}$$

The second to last equality follows since

$$\begin{aligned} &\prod_{\{a,b\} \in [m]} g(i_a/n, i_b/n) = 1 \\ &\iff g(i_a/n, i_b/n) = 1 \text{ for all } \{a, b\} \in E(H) \\ &\iff \{i_a, i_b\} \in E(G) \text{ when } \{a, b\} \in E(H). \end{aligned}$$

■

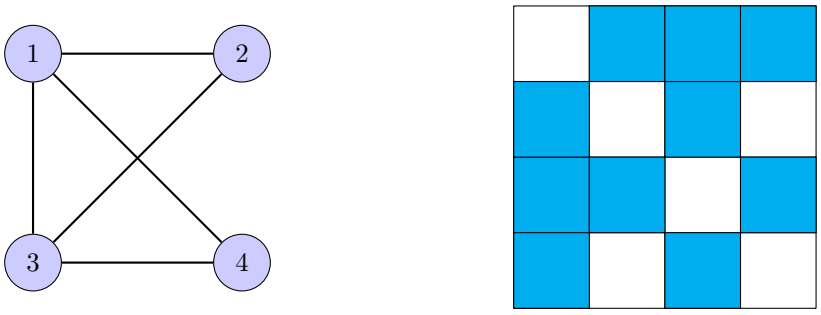

Figure 1.1: Simple graph G on 4 vertices and its corresponding graphon representation.

Example 1.1. As an illustration of the above graphon representation, Figure 1.1 shows a graph on 4 vertices, while the corresponding graphon g^G is depicted as a coloring of a $|V(G)| \times |V(G)|$ grid (in this case, 4×4), where a cyan square indicates presence of an edge and a white square indicates absence of an edge. We index starting in the top left corner

with coordinates $(0,0)$ down to the bottom right corner with coordinates $(1,1)$, similar to how matrices are indexed (this is the convention of Lovász in [46]). This construction can be extended to graphs with weighted edges by defining g^G to have the edge weight as its values.

There are clear advantages to working with the graphon space $\mathcal{W}$; the most immediate of which is that it allows one to consider all simple graphs, regardless of the number of vertices, as elements of the same space. This will prove useful in the establishment of large deviation principles for sequences of distributions on graphs after some technical concerns are addressed.

We now define a norm on the space $\mathcal{W}$. Let $f \in \mathcal{W}$ and define the *cut norm* as

$$\|f\|_{\square} = \sup_{S,T \subset [0,1]} \left| \int_{S \times T} f(x,y)\, dx\, dy \right|. \tag{1.10}$$

The cut norm was introduced on matrices by Frieze and Kannan [36] in the context of finding certain matrix approximations. The cut norm further induces a distance. For $f, h \in \mathcal{W}$ define the corresponding *cut distance* by

$$d_{\square}(f,h) = \|f - h\|_{\square}. \tag{1.11}$$

The cut distance as defined is a pseudometric. Note that if $d_{\square}(f,h) = 0$, then f and h either differ on a set of measure 0. To tackle this issue and obtain a metric, we introduce an equivalence relation $\sim$ on $\mathcal{W}$. For $f, h \in \mathcal{W}$ we say that $f \sim h$ if there exists a measure preserving bijection $\sigma : [0,1] \to [0,1]$ such that $f(x,y) = h_\sigma(x,y) := h(\sigma x, \sigma y)$. One can think of the measure preserving bijection σ as a relabeling of the vertices. This equivalence relation between graphons yields a quotient space $\widetilde{\mathcal{W}} = \mathcal{W}/\sim$, referred to as the *reduced graphon space*. We denote by $\widetilde{f}$ the equivalence class of f with respect to the relation $\sim$

on $\mathcal{W}$. In particular, if G is a finite simple graph, we associate to g^G its equivalence class $\widetilde{g}^G$ in $\widetilde{\mathcal{W}}$. Since the cut distance $d_\square$ is invariant under measure preserving bijections σ, we can define a cut distance $\delta_\square$ on $\widetilde{\mathcal{W}}$ as

$$\delta_\square(\widetilde{f}, \widetilde{h}) = \inf_{\sigma_1,\sigma_2} d_\square(f_{\sigma_1}, h_{\sigma_2}), \tag{1.12}$$

where $\widetilde{f}, \widetilde{h} \in \widetilde{\mathcal{W}}$ are representative elements of f and h respectively, and σ_1, σ_2 are measure preserving bijections of $[0, 1]$. The technical complication in $(\mathcal{W}, d_\square)$ is thus remedied and the quotient space $(\widetilde{\mathcal{W}}, \delta_\square)$ becomes a metric space.

One major advantage of the reduced graphon space $(\widetilde{\mathcal{W}}, \delta_\square)$ is that for any finite simple graph H, the homomorphism density functions $t(H, \cdot)$ in Equation (1.7) are continuous with respect to the cut metric $\delta_\square$. The following theorem from [17] relates the definition of convergence for a sequence of graphs to convergence of graphons under the $\delta_\square$ metric.

Theorem 1.2 (Borgs et. al., Theorem 3.8 in [17])**.** Let $\widetilde{f}_n$ be a sequence of graphons in $\widetilde{\mathcal{W}}$. Then the following are equivalent:

1. $t(H, \widetilde{f}_n)$ converges for all finite simple graphs H.
2. $\widetilde{f}_n$ is a cauchy sequence in the $\delta_\square$ metric.
3. There exists a graphon $\widetilde{f} \in \widetilde{\mathcal{W}}$ such that $t(H, \widetilde{f}_n) \to t(H, \widetilde{f})$ for all finite simple graphs H.

Furthermore, $t(H, \widetilde{f}_n) \to t(H, \widetilde{f})$ for all finite simple graphs H for some graphon $\widetilde{f} \in \widetilde{\mathcal{W}}$ if and only if $\delta_\square(\widetilde{f}_n, \widetilde{f}) \to 0$.

An important corollary of this theorem establishes a close connection between a convergent sequence of graphs and the limiting graphon.

Corollary 1.3 (Borgs et. al., Corollary 3.9 in [17])**.** For any convergent sequence $\{G_n\}$ of weighted graphs with uniformly bounded edge weights, there exists a graphon $\widetilde{f} \in \widetilde{\mathcal{W}}$ such that $\delta_\square(\widetilde{g}^{G_n}, \widetilde{f}) \to 0$. Conversely, any graphon $\widetilde{f} \in \widetilde{\mathcal{W}}$ can be obtained as the limit of a sequence of weighted graphs with uniformly bounded edge weights. The limit of a convergent graph sequence is unique in the following sense: If $\widetilde{g}^{G_n} \to \widetilde{f}$ and $\widetilde{g}^{G_n} \to \widetilde{h}$, then $\delta_\square(\widetilde{f}, \widetilde{h}) = 0$.

The following figure gives a visual realization of the convergence of a sequence of graphs and the limiting graphon. We can think of these pixel pictures as representations of graphs where each pixel indicates the presence (black) or absence (white) of an edge. Note that these pictures are symmetric about the main diagonal since we consider undirected edges. As the number of vertices increases without bound, the limiting pixel picture is a representation of the measurable function.

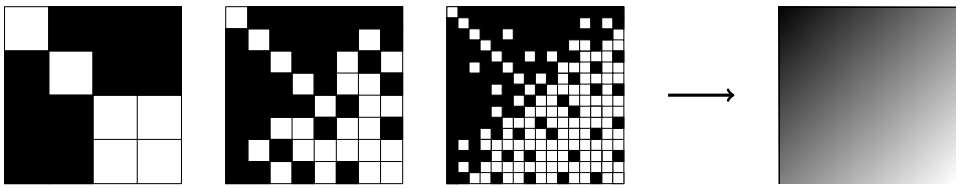

Figure 1.2: Graphical representation of a sequence of graphs on increasing number of vertices converging to a limiting graphon.

1.4 Gibbs measure

Generic exponential random graphs are constructed via a Gibbs measure on the set $\mathcal{G}_n$ of simple, labelled graphs G_n on n vertices. Let $T : \widetilde{\mathcal{W}} \to \mathbb{R}$ be a bounded continuous function on the reduced graphon space. Define a probability measure $\mathbf{P}_n^T : \mathcal{G}_n \to [0, 1]$ as

$$\mathbf{P}_n^T(G_n) = \exp\left(n^2\left(T(\widetilde{g^{G_n}}) - \psi_n^T\right)\right), \tag{1.13}$$

where ψ_n^T is the *partition function* or *normalization constant*

$$\psi_n^T = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n} \exp\left(n^2 T(\widetilde{g^{G_n}})\right). \tag{1.14}$$

More specifically, we are interested in k-parameter familys of exponential random graphs, where the function T in the definition of the Gibbs measure is a sum of k graph homomorphism densities,

$$T(\widetilde{g}^{G_n}) = \sum_{i=1}^{k} \beta_i t(H_i, \widetilde{g^{G_n}}), \tag{1.15}$$

where $\beta = (\beta_1, \ldots, \beta_k) \in \mathbb{R}^k$ and $H_1, \ldots, H_k$ are finite subgraphs of G_n (by convention we take H_1 to be a single edge). We will say that the k-parameter exponential random graph model is *attractive* if the parameters $\beta_2, \ldots, \beta_k$ are positive and *repulsive* if the parameters $\beta_2, \ldots, \beta_k$ are negative. We may alternatively write the Gibbs measure from Equation (1.13) on $\mathcal{G}_n$ as

$$\mathbf{P}_n^{\beta}(G_n) = \exp\left(n^2 \left(\sum_{i=1}^{k} \beta_i t(H_i, G_n)\right) - n^2 \psi_n^{\beta}\right), \tag{1.16}$$

with an associated normalization constant ψ_n^{β} (Equation (1.14)):

$$\psi_n^{\beta} = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n} \exp\left(n^2 \sum_{i=1}^{k} \beta_i t(H_i, G_n)\right). \tag{1.17}$$

In thermodynamics, the normalization constant is also referred to as the *Helmholtz free energy*. One can see in the following example that the $G(n,p)$ model introduced in Section 3.1 is indeed a special case of an exponential random graph model.

Example 1.2. Let $p \in (0,1)$ and H_1 a single edge. The Erdős-Rényi $G(n,p)$ model is a single parameter exponential random graph model, with a parameter β depending on p.

From Definition 1.1, $t(H_1, G_n) = 2\left|E(G_n)\right|/n^2$. Take

$$\beta = \frac{1}{2}\log\left(\frac{p}{1-p}\right) \quad \text{and} \quad \psi_n^\beta = -\frac{1}{n^2}\binom{n}{2}\log(1-p). \tag{1.18}$$

Simple calculation yields

$$\begin{aligned}\mathbf{P}_n^\beta(G_n) &= \exp\left(n^2\left(\beta t(H_1, G_n) - \psi_n^\beta\right)\right) \\ &= p^{|E(G_n)|}(1-p)^{\binom{n}{2}-|E(G_n)|} = \mathbf{P}_n^p(G_n).\end{aligned} \tag{1.19}$$

The random triangle model is also an exponential random graph model being a two parameter exponential random graph model with subgraphs H_1 a single edge and H_2 a triangle with an appropriate normalization constant.

One central difficulty in the study of exponential random graph models is the calculation of the normalization constant ψ_n^β for large values of n. Through the map that sends $G_n \in \mathcal{G}_n$ to $g^{G_n} \in \mathcal{W}$, the probability measure $\mathbf{P}_n^\beta$ on the space $\mathcal{G}_n$ induces a push-forward probability distribution on the space $\mathcal{W}$. We refer to this distribution on $\mathcal{W}$ as $\mathbf{Q}_n^\beta$. Through the quotient map $\mathcal{W} \to \widetilde{\mathcal{W}}$, we further obtain a push-forward probability distribution $\widetilde{\mathbf{Q}}_n^\beta$ on the reduced graphon space $\widetilde{\mathcal{W}}$. These probability distributions will be investigated in the next section and a characterization of $\lim_{n\to\infty}\psi_n^\beta$ is determined.

1.5 Large deviation principle

The goal of large deviation theory is to compute the asymptotic probabilities of rare events. In their wonderful paper [22], Chatterjee and Varadhan developed a large deviation principle for the Erdős-Rényi random graph model $G(n,p)$. This is a beautiful result in its own right and lays the groundwork for many subsequent results on the asymptotic behavior of graphs sampled from the exponential random graph model, as well as methods for determining the associated limiting normalization constant. Example 1.2 in the previous

Section 1.4 shows that the exponential random graph model is an extension of the classical Erdős-Rényi model. A natural question to ask is how significant is this extension? How does a typical exponential random graph behave as the number of vertices goes to infinity? For finite values of n, the exponential random graph model captures more diverse behavior than the classical Erdős-Rényi model [72, 52, 53] . Are the two models still appreciably different in the limit as n goes to infinity? The development of large deviation results help to answer these questions.

Recall that a function $f : \mathcal{X} \to \mathbb{R}$ is *lower semicontinuous* if the lower level sets $L_f(y) = \{x \in \mathcal{X} : f(x) \leq y\}$ are closed in $\mathcal{X}$ for all $y \in \mathbb{R}$.

Definition 1.3. Let $I : \mathcal{X} \to [0, \infty]$ be a lower semicontinuous function, where $\mathcal{X}$ is a Hausdorff space, and $\{r_n\} \nearrow \infty$ an increasing sequence of positive real numbers. A sequence of probability measures $\{\mu_n\} \subseteq \mathcal{M}_1(\mathcal{X})$, the space of probability measures on $\mathcal{X}$, is said to satisfy a *large deviation principle with rate function I and normalization r_n* if for all closed subsets F of $\mathcal{X}$,

$$\limsup_{n\to\infty} r_n^{-1} \log \mu_n(F) \leq - \inf_{x\in F} I(x), \tag{1.20}$$

and for all open subsets G of $\mathcal{X}$,

$$\liminf_{n\to\infty} r_n^{-1} \log \mu_n(G) \geq - \inf_{x\in G} I(x). \tag{1.21}$$

We will abbreviate this as LDP(μ_n, r_n, I).

Fix $p \in (0, 1)$. In foresight of the goal of formulating a large deviation principle for Erdős-Rényi random graphs, where edges can be thought of as Bernoulli random variables

with probability p, define $I_p : [0,1] \to \mathbb{R}$ as

$$\begin{aligned} I_p(u) &= \frac{1}{2} \sup_{a,b \in \mathbb{R}} \left\{ au + b(1-u) - \log(pe^a + (1-p)e^b) \right\} \\ &= \frac{1}{2} \sup_{a \in \mathbb{R}} \left\{ au - \log(pe^a + (1-p)) \right\}, \\ &= \frac{1}{2} u \log \frac{u}{p} + \frac{1}{2}(1-u) \log \frac{1-u}{1-p} \end{aligned} \tag{1.22}$$

and extend the domain of I_p to $\mathcal{W}$ such that for $f \in \mathcal{W}$,

$$\begin{aligned} I_p(f) &= \frac{1}{2} \sup_{a:[0,1]^2 \to \mathbb{R}} \left\{ \int a(x,y) f(x,y)\, dx\, dy - \int \log(pe^{a(x,y)} + (1-p))\, dx\, dy \right\} \\ &= \frac{1}{2} \int_0^1 \int_0^1 f(x,y) \log \frac{f(x,y)}{p} + (1 - f(x,y)) \log \frac{1 - f(x,y)}{1-p}\, dx\, dy \\ &= \int_0^1 \int_0^1 I_p(f(x,y))\, dx\, dy \end{aligned} \tag{1.23}$$

where the supremum in the first line is taken over bounded, measurable functions. It may seem as though the function $I_p(u)$ appears out of nowhere; however, there is good reason why one might expect it to appear in this context. We now provide some intuition as to why. Another argument using Stirling's approximation can be found in Chapter 1 of [61]. A nearly identical function to I_p occurs in large deviations of Bernoulli random variables. Let $p \in (0,1)$ as before and let $s \in (0,1)$. Consider a sequence of iid Bernoulli random variables X_i each with identical success probability p. Define $S_n = X_1 + \cdots + X_n$. We wish to analyze the behavior of the event $\mathbf{P}(S_n \geq ns)$. Using Markov's inequality, for any $t > 0$

$$\mathbf{P}(S_n \geq ns) = \mathbf{P}(S_n - ns \geq 0) = \mathbf{P}\left(e^{t(S_n - ns)} \geq 1\right) \leq e^{-nts} \mathbf{E}\left(e^{tS_n}\right). \tag{1.24}$$

The variables $X_1, \ldots, X_n$ are iid, so

$$\mathbf{E}\left(e^{tS_n}\right) = \left(\mathbf{E}\left(e^{tX_1}\right)\right)^n = \left(1 - p + pe^t\right)^n, \tag{1.25}$$

since $\mathbf{E}\left(e^{tX_1}\right)$ is the moment generating function of a Bernoulli random variable. Then for all $t > 0$,

$$\mathbf{P}(S_n \geq ns) \leq e^{-n\left(ts - \log\left(1-p+pe^t\right)\right)}. \tag{1.26}$$

Now we optimize the right hand side of the preceding inequality over all $t > 0$ to find that

$$\mathbf{P}(S_n \geq ns) \leq e^{-nh_p(s)} \tag{1.27}$$

where

$$\begin{aligned} h_p(s) &= \sup_t \left\{ts - \log\left(pe^t + (1-p)\right)\right\} \\ &= s\log\left(\frac{s}{p}\right) + (1-s)\log\left(\frac{1-s}{1-p}\right) \end{aligned} \tag{1.28}$$

With a little more work, one can show that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{S_n}{n} \geq s\right) = -h_p(s). \tag{1.29}$$

In other words, the sequence of measures $\mu_n(A) = \mathbf{P}(S_n/n \in A)$ for an iid sequence of Bernoulli random variables with success probability p satisfies a large deviation principle with rate function h_p and normalization n. Note that the law of large numbers implies that $S_n/n \to p$ almost surely. For $s > p$, one would expect then that $\mathbf{P}(S_n/n \geq s)$ goes to zero as $n \to \infty$. The large deviation principle gives the precise rate at which $\mathbf{P}(S_n/n \geq s)$ goes

to zero, that is, for $s > p$ the probability $\mathbf{P}(S_n/n \geq s)$ converges to 0 at the same rate as $e^{-nh_p(s)}$.

We now return to the setting as before and turn to the establishment of a large deviation principle for Erdős-Rényi random graphs. The function I_p can be further extended to $\widetilde{\mathcal{W}}$ by setting $I_p(\widetilde{f}) = I_p(f)$ where $f \in \mathcal{W}$ is any representative element of the equivalence class $\widetilde{f} \in \widetilde{\mathcal{W}}$. This raises the question as to whether I_p is well defined on $\widetilde{\mathcal{W}}$ and lower semicontinuous. The following lemma from [22] gives a positive answer. Along with the intuition of how I_p arises in the context of Bernoulli random variables, this makes I_p a good contender for a large deviation rate function.

Lemma 1.4 (Chatterjee and Varadhan, Lemma 2.1 in [22])**.** The function I_p is well defined on $\widetilde{\mathcal{W}}$ and is lower semicontinuous under the cut metric $\delta_\square$ on $\widetilde{\mathcal{W}}$.

Recall the measure $\mathbf{P}_n^p$ from Example 1.2. Let $\mathbf{Q}_n^p$ be the pushforward measure on the space $\mathcal{W}$ induced by $\mathbf{P}_n^p$. The following two theorems establish large deviation principles on the graphon spaces $\mathcal{W}$ and $\widetilde{\mathcal{W}}$ respectively.

Theorem 1.5 (Chatterjee and Varadhan, Theorem 2.2 in [22])**.** For any fixed $p \in (0, 1)$ the large deviation principle LDP$(\mathbf{Q}_n^p, n^2, I_p)$ holds in the weak topology on $\mathcal{W}$. That is, for every weakly closed set $F \subset \mathcal{W}$,

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \mathbf{Q}_n^p(F) \leq -\inf_{f\in F} I_p(f), \tag{1.30}$$

and for every weakly open set $G \subset \mathcal{W}$,

$$\liminf_{n\to\infty} \frac{1}{n^2} \log \mathbf{Q}_n^p(G) \geq -\inf_{f\in G} I_p(f). \tag{1.31}$$

We can see that this looks similar to the LDP(μ_n, n, h_p) from before where the n is replaced by n^2 for normalization and $1/2$ appears in the rate function due to there being

$\binom{n}{2}$ possible edges in a graph on n vertices. The large deviation principle on the reduced graphon space $\widetilde{\mathcal{W}}$ will prove to be useful as well.

Theorem 1.6 (Chatterjee and Varadhan, Theorem 2.3 in [22])**.** For any fixed $p \in (0,1)$ the large deviation principle LDP$(\widetilde{\mathbf{Q}}_n^p, n^2, I_p)$ holds in the metric topology induced by $\delta_\square$ on $\widetilde{\mathcal{W}}$. That is, for every closed set $\widetilde{F} \subset \widetilde{\mathcal{W}}$,

$$\limsup_{n\to\infty} \frac{1}{n^2} \log \widetilde{\mathbf{Q}}_n^p(\widetilde{F}) \leq - \inf_{\widetilde{f}\in\widetilde{F}} I_p(\widetilde{f}), \tag{1.32}$$

and for every open set $\widetilde{G} \subset \widetilde{\mathcal{W}}$,

$$\liminf_{n\to\infty} \frac{1}{n^2} \log \widetilde{\mathbf{Q}}_n^p(\widetilde{G}) \geq - \inf_{\widetilde{f}\in\widetilde{G}} I_p(\widetilde{f}). \tag{1.33}$$

Recall that one of the central difficulties in the study of exponential random graphs is the calculation of the normalization constant. To this end, define the *limiting normalization constant* $\psi_\infty^T = \lim_{n\to\infty} \psi_n^T$, where ψ_n^T is defined as in Equation (1.14). Chatterjee and Diaconis presented an alternative formulation of the limiting normalization constant ψ_∞^T in their seminal paper [24]. Let $T : \widetilde{\mathcal{W}} \to \mathbb{R}$ be a bounded continuous function on $(\widetilde{\mathcal{W}}, \delta_\square)$ and consider the probability measure $\mathbf{P}_n^T$ on $\mathcal{G}_n$ defined in Equation (1.13). Define $I : [0,1] \to \mathbb{R}$ as

$$I(u) = \frac{1}{2} u \log u + \frac{1}{2}(1-u)\log(1-u), \tag{1.34}$$

and extend I to $\widetilde{\mathcal{W}}$, in a similar way as for I_p (1.22) (1.23), by setting

$$I(\widetilde{f}) = \int_0^1 \int_0^1 I(f(x,y))\, dx\, dy, \tag{1.35}$$

where f is any representative element of the equivalence class $\widetilde{f} \in \widetilde{\mathcal{W}}$. Since I can be written as $I(u) = I_{1/2}(u) - (\log 2)/2$, Lemma 1.4 says that I is well defined and lower semicontinuous on $\widetilde{\mathcal{W}}$.

Theorem 1.7 (Chatterjee and Diaconis, Theorem 3.1 in [24])**.** If $T : \widetilde{\mathcal{W}} \to \mathbb{R}$ is a bounded continuous function and ψ_n^T and I are defined as in Equations (1.14), (1.34) and (1.35), then the limiting normalization constant

$$\psi_\infty^T = \lim_{n\to\infty} \psi_n^T = \sup_{\widetilde{f}\in\widetilde{\mathcal{W}}} \left\{ T(\widetilde{f}) - I(\widetilde{f}) \right\}. \tag{1.36}$$

The proof of Theorem 1.7 mainly utilizes the large deviation principle developed by Chatterjee and Varadhan in [22] for the $G(n, 1/2)$ model as well as the boundedness and continuity of the function T. By the continuity of T and the lower semicontinuity of I, the supremum in Equation (3.49) is achieved. Let $\widetilde{M}^* = (T - I)^{-1}\left(\left\{\psi_\infty^T\right\}\right)$. It follows that $\widetilde{M}^*$ is closed. Further, by the compactness of $\widetilde{\mathcal{W}}$, $\widetilde{M}^*$ is also compact. The follow-up Theorem 1.8 shows that for sufficiently large values of n, the quotient image $\widetilde{g}^{G_n}$ of a graph G_n drawn from the exponential random graph model defined by T (Equation (1.13)) lies close to the set of maximizers for the limiting normalization constant with exponentially high probability. Especially, if $\widetilde{M}^*$ is a singleton set, the theorem gives a law of large numbers for G_n; while if $\widetilde{M}^*$ is not a singleton set, the theorem points to the existence of a first order phase transition (a concept that will be expanded upon later in Section 2.7).

Theorem 1.8 (Chatterjee and Diaconis, Theorem 3.2 in [24])**.** Let G_n be drawn from $\mathbf{P}_n^T$ (Equation (1.13)) and $\widetilde{M}^*$ be defined as in the above paragraph. Then for any $\eta > 0$ there exists a $C > 0$ and a $\delta > 0$ such that for any $n \in \mathbb{N}$,

$$\mathbf{P}_n^T\left(\delta_\square\left(\widetilde{g}^{G_n}, \widetilde{M}^*\right) > \eta\right) \leq Ce^{-n^2\delta}. \tag{1.37}$$

In the case of attractive k-parameter exponential random graphs (Equation (1.16)), Theorems 1.7 and 1.8 may be reformulated and simplified. As was explained in the introduction, the goal of exponential random graph models is to capture the behavior of modern networks, such as clustering and transitivity. When T is a linear combination of homomorphism densities, this may be achieved through the adjustment of the parameters coupled with different subgraph statistics in the exponential probability measure. Desirable graph features are given higher probability (corresponding to a larger or more positive β) while undesirable graph features are given lower probability (corresponding to a smaller or more negative β).

The following theorems tell us that a graph drawn from the exponential random graph model in the attractive region of the parameter space is *weakly pseudorandom* [12], that means that it satisfies a number of equivalent properties that are shared by Erdős-Rényi random graphs.

Theorem 1.9 (Chatterjee and Diaconis, Theorem 4.1 in [24])**.** Consider the k-parameter exponential random graph model (Equation (1.16)). Suppose $\beta_2, \ldots, \beta_k$ are non-negative. Then the limiting normalization constant $\psi_\infty^\beta := \lim_{n\to\infty} \psi_n^\beta$ exists and is given by

$$\psi_\infty^\beta = \sup_{u\in[0,1]} \left\{ \sum_{i=1}^{k} \beta_i u^{e(H_i)} - I(u) \right\}, \tag{1.38}$$

where $e(H_i)$ is the number of edges in H_i and I is defined earlier in (1.34).

Theorem 1.10 (Chatterjee and Diaconis, Theorem 4.2 in [24])**.** Let G_n be an exponential random graph drawn from (Equation (1.16)). Suppose $\beta_2, \ldots, \beta_k$ are non-negative. Then G_n behaves like an Erdős-Rényi graph $G(n, u^*)$ in the large n limit in the sense of Equation (1.37), where u^* is picked randomly from the set U of maximizers of Equation (1.38).

Chapter 2: Phase transitions in edge-weighted exponential random graphs

2.1 Introduction

Despite their flexibility, conventionally used exponential random graphs admittedly have some shortcomings. The primary one that we will be concerned with in this section is the fact that they cannot directly model weighted networks as the underlying probability space consists of simple graphs only. Since many substantively important networks are weighted, this limitation is especially problematic. An alternative interpretation for simple graphs is such that the edge weights are iid and satisfy a Bernoulli distribution. Following this perspective, Yin [73] extended the exponential framework by putting a generic common distribution on the iid edge weights. After deriving a variational principle for the limiting normalization constant and an associated concentration of measure, an explicit characterization of the asymptotic phase transition was obtained for exponential models with uniformly distributed edge weights. This work expands upon the setting in [73] and places minimal assumptions on the edge-weights distribution, that is, it is non-degenerate and supported on the unit interval. By doing so, we recognize the essential properties associated with near-degeneracy and universality in edge-weighted exponential random graphs.

This section is organized as follows. In Section 2.2 we recall some basics of graph limit theory first introduced in Section 1.3, define the model of interest, and introduce key features of edge-weighted exponential random graphs. Then, in Section 2.3, we summarize important properties of Legendre duality between the cumulant generating function and the Cramér rate function for the edge-weights distribution. Section 2.4 demonstrates the

existence of a first order phase transition curve ending in a second order critical point in general edge-weighted exponential random graph models through a detailed analysis of the maximization problem for the normalization constant. Lastly, Section 2.5 explores the universal and non-universal asymptotics concerning the phase transition.

2.2 Background

Consider the set $\overline{\mathcal{G}}_n$ of all simple edge-weighted complete labelled graphs G_n on n vertices ("simple" means undirected, with no loops or multiple edges), where the edge weights x_{ij} between vertex i and vertex j are iid real random variables satisfying a non-degenerate common distribution μ that is supported on $[0, 1]$. As was shown in Section 1.3, any such graph G_n, irrespective of the number of vertices, may be represented as an element h^{G_n} of the graphon space by setting $h^{G_n}(x, y)$ as the edge weight between vertices $\lceil nx \rceil$ and $\lceil ny \rceil$ of G_n where $V(G_n) = \{1, \ldots, n\}$. Note that in this setting $\mathcal{G}_n \subset \overline{\mathcal{G}}_n$ for every n if we view an element $G \in \mathcal{G}_n$ as a complete graph where the 'absent' edges have weight 0 and the 'present' edges have weight 1. The common distribution μ for the edge weights yields probability measure $\mathbf{P}_n$ and the associated expectation $\mathbf{E}_n$ on $\overline{\mathcal{G}}_n$, and further induces probability measure $\mathbf{Q}_n$ on the space $\mathcal{W}$ under the graphon representation.

By a 2-parameter family of edge-weighted exponential random graphs we mean a family of probability measures $\mathbf{P}_n^\beta$ on $\overline{\mathcal{G}}_n$ such that, for any $G_n \in \overline{\mathcal{G}}_n$,

$$\mathbf{P}_n^\beta(G_n) = \exp\left(n^2\left(\beta_1 t(H_1, G_n) + \beta_2 t(H_2, G_n) - \psi_n^\beta\right)\right)\mathbf{P}_n(G_n), \tag{2.1}$$

where $\beta = (\beta_1, \beta_2)$ are 2 real parameters, H_1 is a single edge, H_2 is a finite simple graph with $p \geq 2$ edges, $t(H_i, G_n)$ is the density of graph homomorphisms, $\mathbf{P}_n$ is the probability measure induced by the common distribution μ for the edge weights, and ψ_n^β is the

normalization constant,

$$\psi_n^\beta = \frac{1}{n^2} \log \mathbf{E}_n \left(\exp \left(n^2 \left(\beta_1 t(H_1, G_n) + \beta_2 t(H_2, G_n) \right) \right) \right). \tag{2.2}$$

Since homomorphism densities $t(H_i, G_n)$ are preserved under vertex relabeling, the probability measure $\tilde{\mathbf{P}}_n^\beta$ and the associated expectation $\tilde{\mathbf{E}}_n^\beta$ (which coincides with $\mathbf{E}_n^\beta$) may likewise be defined.

Being exponential families with bounded support, one might expect exponential random graph models to enjoy a rather basic asymptotic form, though in fact, virtually all these models are highly nonstandard as n increases. The 2-parameter edge-weighted exponential random graph models are simpler than their k-parameter extensions but nevertheless exhibit a wealth of non-trivial characteristics and capture a variety of interesting features displayed by large networks. Furthermore, the relative simplicity provides insight into the expressive power of the exponential construction. In statistical physics, we refer to β_1 as the particle parameter and β_2 as the energy parameter. Accordingly, recall that the exponential model defined in Equation (2.1) is said to be "attractive" if β_2 is positive and "repulsive" if β_2 is negative. In this section we will concentrate on "attractive" 2-parameter models. The interest in these models is well justified. Consider the the graph of a social network, where the edge weights between different vertex pairs measure the strength of mutual friendship (perhaps represented by the frequency of interactions between them). Take H_1 an edge and H_2 a triangle. Since a friend of a friend is likely also a friend, the influence of a triangle that assesses the bond of a 3-way friendship should be emphasized, and this corresponds to taking $\beta_2 \geq 0$.

In this section we aim to study the global structure of edge-weighted exponential random graph models. Namely, in Section 2.4, we show that, in a particular region of the parameter space, a phase transition is present and the model is not appreciably different

from $G(n,p)$ using the Legendre transform. In Section 2.5, we identify this phase transition curve explicitly and determine that, when the model is similar to $G(n,p)$, the dual, θ, of p with respect to the Legendre transform and the limiting normalization constant ψ_∞^β exhibit universal asymptotic behavior in terms of the ERGM parameters β_1 and β_2.

2.3 Legendre transform and duality

We now present properties of the cumulant generating function $K(\theta)$ and the Cramér rate function $I(u)$ for the edge-weights distribution μ relevant to our investigation. We will see that $K(\theta)$ is convex on $\mathbb{R}$, which allows the application of the Legendre transform. Let $I : A \to \mathbb{R}$ be the *Legendre-Fenchel transform* of K given by

$$I(u) = \sup_{\theta \in \mathbb{R}} \{\theta u - K(\theta)\}, \tag{2.3}$$

where A, the domain of I, consists of all u so that $I(u) < \infty$. Note that in large deviation theory, I is commonly referred to as the Cramér conjugate rate function for the distribution μ. It follows from Lemma 2.1 that the Legendre transform connecting K and I is an involution, I is smooth and strictly convex everywhere it is defined, and there is a 1-1 relationship between K and I. Lemma 2.2 and Proposition 2.4 then discuss properties of $K(\theta)$ and $I(u)$ under the additional assumption that μ is symmetric. These properties will be useful in Section 2.5 when we explore universality in edge-weighted exponential random graphs. We say that a probability measure $\mathbf{P}$ on $[0,1]$ is *degenerate* if there exists a $c \in [0,1]$ such that $\mathbf{P}(X = c) = 1$.

Lemma 2.1. Consider a non-degenerate probability measure μ supported on $[0,1]$ (i.e., μ is not supported at only one point). Let $M(\theta) = \int e^{\theta x} \mu(dx)$ be the associated moment generating function and $K(\theta) = \log M(\theta)$ be the associated cumulant generating function. Then $K(\theta)$ is everywhere defined on $\mathbb{R}$, infinitely differentiable, and strictly convex.

Proof. The fact that K is well-defined and smooth follows from standard analytical arguments. Since μ is non-degenerate, $M(\theta) > 0$ and, by Cauchy-Schwarz,

$$\left(\int xe^{\theta x}\mu(dx)\right)^2 < \int x^2 e^{\theta x}\mu(dx) \int e^{\theta x}\mu(dx), \tag{2.4}$$

thus $M(\theta)M''(\theta) - [M'(\theta)]^2 > 0$. This implies that $K''(\theta) > 0$ for all $\theta \in \mathbb{R}$ and so K is strictly convex. ■

Lemma 2.2. Consider a non-degenerate probability measure μ supported on $[0, 1]$. Let $K(\theta)$ be the associated cumulant generating function. If μ is symmetric about the line $u = 1/2$, then $K'''(0)K'(0) + (p-2)\left(K''(0)\right)^2 \geq 0$, and equality is obtained only when $p = 2$.

Proof. Let X be a random variable distributed according to μ. By symmetry, $\mathbf{E}(X) = 1/2$ and $\mathbf{E}(X^3) = 3\mathbf{E}(X^2)/2 - 1/4$. This implies that $K'(0) = \mathbf{E}(X) = 1/2$ and $K''(0) = \mathbf{E}(X^2) - \left(\mathbf{E}(X)\right)^2 = \mathbf{E}(X^2) - 1/4$. Also,

$$K'''(0) = \mathbf{E}(X^3) - 3\mathbf{E}(X^2)\mathbf{E}(X) + 2\left(\mathbf{E}(X)\right)^3 = 0. \tag{2.5}$$

The claim thus follows. ■

Lemma 2.3. Consider a non-degenerate probability measure μ supported on $[0, 1]$. Let $I(u)$ be the associated Cramér rate function in Equation (2.3). Then the domain of I is a subset of $[0, 1]$.

Proof. Since μ is supported on $[0, 1]$, we have $0 \leq K(\theta) \leq \theta$ if $\theta \geq 0$, and $\theta \leq K(\theta) \leq 0$ if $\theta \leq 0$. This gives

$$I(u) = \sup\left\{\sup_{\theta\geq 0}\left\{\theta u - K(\theta)\right\}, \sup_{\theta\leq 0}\left\{\theta u - K(\theta)\right\}\right\} \tag{2.6}$$

Limiting Properties of $K(\theta)$	θ limit
$K(\theta) \to -\infty$ or $l < 0$	$\theta \to -\infty$
$K'(\theta) \to 0$	$\theta \to -\infty$
$K''(\theta) \to 0$	$\theta \to -\infty$
$K(\theta) \to \infty$	$\theta \to \infty$
$K'(\theta) \to 1$	$\theta \to \infty$
$K''(\theta) \to 0$	$\theta \to \infty$

Table 2.1: Limiting properties of $K(\theta)$ as $\theta \to \pm\infty$.

$$\geq \sup\left\{\sup_{\theta\geq 0}\left\{\theta\left(u-1\right)\right\}, \sup_{\theta\leq 0}\left\{\theta u\right\}\right\}.$$

If $u > 1$ then $\sup_{\theta\geq 0}\left\{\theta\left(u-1\right)\right\} = \infty$ and thus $I(u)$ is not finite. Similarly, if $u < 0$ then $\sup_{\theta\leq 0}\left\{\theta u\right\} = \infty$ and thus $I(u)$ is not finite. ■

Analyzing properties of $K(\theta)$ and $I(u)$ in detail will give a stronger conclusion than Lemma 2.3. We recognize that the cumulant generating function $K(\theta)$ satisfies $K(0) = 0$, $K'(0) = \mathbf{E}(X)$, and $K''(0) = \mathbf{Var}(X)$, where X is a random variable distributed according to μ. See Table 2.1 for important limiting properties of $K(\theta)$ as $\theta \to \pm\infty$. By Legendre duality, every $u \in (0,1)$ uniquely corresponds to a $\theta \in (-\infty, \infty)$, with $K'(\theta) = u$ and $I'(u) = \theta$. This implies that $I(\mathbf{E}(X)) = I'(\mathbf{E}(X)) = 0$, and $I(u)$ is decreasing on $(0, \mathbf{E}(X))$ and increasing on $(\mathbf{E}(X), 1)$. We also note that $I(0)$ and $I(1)$, depending on the probability distribution μ, may be either finite or grow unbounded. In the former case, the domain of I is $[0,1]$ (as for Bernoulli$(.5)$). In the latter case, the domain of I is $(0,1)$ (as for Uniform$(0,1)$).

Proposition 2.4. Consider a non-degenerate probability measure μ supported on $[0,1]$. Let $I(u)$ be the associated Cramér rate function defined in Equation (2.3). If μ is symmetric about the line $u = 1/2$, then $I(u)$ is also symmetric about the line $u = 1/2$.

Proof. Let $\theta \in \mathbb{R}$. Under the symmetry assumption, we will show, by a simple change of variable $x = 1 - y$, that $K(-\theta) = -\theta + K(\theta)$.

$$\begin{aligned} K(-\theta) &= \log \int e^{-\theta x} \mu(dx) = \log \int e^{-\theta(1-y)} \mu(dy) \\ &= \log \int e^{-\theta} e^{\theta y} \mu(dy) = -\theta + K(\theta). \end{aligned} \tag{2.7}$$

Let $u \in (0, 1)$. Following Legendre duality, $u = K'(\theta)$ for a unique θ. By Equation (2.7), this implies that $1 - u = 1 - K'(\theta) = K'(-\theta)$, i.e., $1 - u$ and $-\theta$ are unique duals of each other. We compute

$$\begin{aligned} I(u) &= \theta K'(\theta) - K(\theta) \\ &= \theta \left(1 - K'(-\theta)\right) - \left(K(-\theta) + \theta\right) \\ &= (-\theta) K'(-\theta) - K(-\theta) = I(1 - u). \end{aligned} \tag{2.8}$$

This verifies our claim. ■

2.4 Maximization analysis

In this section we demonstrate the existence of first order phase transitions in general edge-weighted exponential random graphs. Our main results are Theorem 2.7 and the consequent Corollary 2.8. In the standard statistical physics literature, phase transition is often associated with loss of analyticity in the normalization constant, which gives rise to discontinuities in the observed graph statistics. In the vicinity of a phase transition, even a tiny change in some local feature can result in a dramatic change of the entire system.

Definition 2.1. A *phase* is a connected region of the parameter space $\{\beta\}$, maximal for the condition that the limiting normalization constant $\psi_\infty^\beta := \lim_{n\to\infty} \psi_n^\beta$ is analytic. There is a *jth-order phase transition* at a boundary point of a phase if at least one jth-order partial derivative of ψ_∞^β is discontinuous there, while all lower order derivatives are continuous.

Following this philosophy, we will make use of two theorems from [73], which connect the occurrence of an asymptotic phase transition in our model with the solution of a certain maximization problem for the limiting normalization constant.

Theorem 2.5 (Yin, Theorem 3.4 in [73])**.** Consider a general 2-parameter exponential random graph model from Equation (2.1). Suppose β_2 is non-negative. Then the limiting normalization constant ψ_∞^β exists, and is given by

$$\psi_\infty^\beta = \sup_u \left(\beta_1 u + \beta_2 u^p - \frac{1}{2} I(u) \right), \tag{2.9}$$

where H_2 is a simple graph with $p \geq 2$ edges, I is the Cramér rate function from Equation (2.3), and the supremum is taken over all u in the domain of I, i.e., where $I < \infty$.

Theorem 2.6 (Yin, Theorem 3.5 in [73])**.** Let G_n be an exponential random graph drawn from Equation (2.1). Suppose β_2 is non-negative. Then G_n behaves like an Erdős-Rényi graph $G(n, u)$ in the large n limit,

$$\lim_{n\to\infty} \delta_\square(\tilde{h}^{G_n}, \tilde{u}) = 0 \text{ almost surely}, \tag{2.10}$$

where u is picked randomly from the set U of maximizers of Equation (2.9).

A significant part of computing phase boundaries for the 2-parameter exponential model is then a detailed analysis of a calculus problem coupled with probability estimates. However, as straightforward as it sounds, since the exact form of the Cramér rate function I is not readily obtainable for a generic edge-weights distribution μ, getting a clear picture of the asymptotic phase structure is not so simple and various tricks, especially the duality principle for the Legendre transform, need to be employed [75]. We note that our mechanism for 2-parameter models may be further generalized to a k-parameter setting, and the crucial idea is to minimize the effect of the ordered parameters on the limiting normal-

ization constant one by one. See [72] for an illustration of this procedure in the standard exponential random graph model (where μ is Bernoulli$(.5)$).

Assumption 1. Let p be the number of edges in H_2. Denote by $K(\theta)$ the cumulant generating function associated with the probability measure μ. We place a technical assumption that

$$K'''(\theta)K'(\theta) = -(p-2)\left(K''(\theta)\right)^2 \tag{2.11}$$

admits only one zero on $\mathbb{R}$.

Remark 1. This requirement on μ, which is satisfied by many common distributions including Bernoulli(0.5) and Uniform$(0,1)$ etc., is just a technicality to help explicitly identify the phase transition curve. For example, this unique zero occurs at $\theta = \log(p-1)$ for $\mu \sim$ Bernoulli(0.5). It is expected that the parameter space would still consist of a single phase with first order phase transition(s) across one (or more) curves and second order phase transition(s) along the boundaries should such Assumption 1 fail.

Theorem 2.7. Suppose the common distribution μ for the edge weights is supported on $[0,1]$ and non-degenerate. For any allowed H_2, the limiting normalization constant ψ_∞^β of Equation (2.1) is analytic at all (β_1, β_2) in the upper half-plane $(\beta_2 \geq 0)$ except on a certain decreasing curve $\beta_2 = r(\beta_1)$ which includes the endpoint (β_1^c, β_2^c). The derivatives $\frac{\partial}{\partial\beta_1}\psi_\infty^\beta$ and $\frac{\partial}{\partial\beta_2}\psi_\infty^\beta$ have (jump) discontinuities across the curve, except at the end point where all the second derivatives $\frac{\partial^2}{\partial\beta_1^2}\psi_\infty^\beta$, $\frac{\partial^2}{\partial\beta_1\partial\beta_2}\psi_\infty^\beta$ and $\frac{\partial^2}{\partial\beta_2^2}\psi_\infty^\beta$ diverge.

Corollary 2.8. For any allowed H_2, the parameter space $\{(\beta_1, \beta_2) : \beta_2 \geq 0\}$ consists of a single phase with a first order phase transition across the indicated curve $\beta_2 = r(\beta_1)$ and a second order phase transition at the critical point (β_1^c, β_2^c).

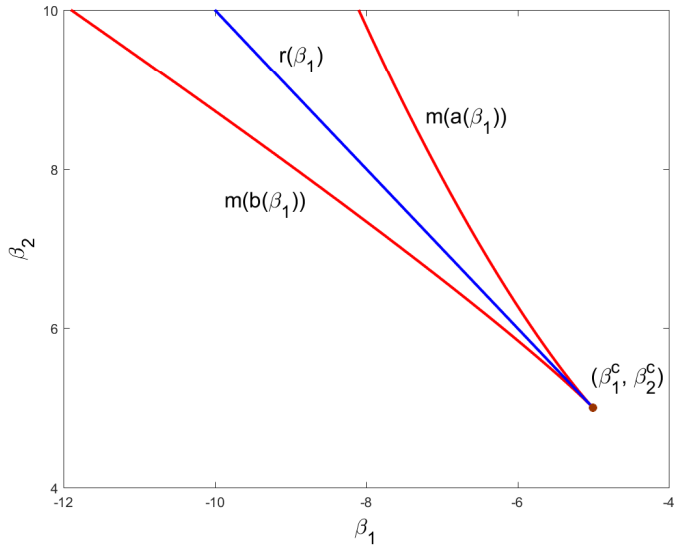

Figure 2.1: The V-shaped region (with phase transition curve $r(\beta_1)$ inside) for the Beta$(2, 2)$ distribution in the (β_1, β_2) plane. Graph drawn for $p = 2$.

Proof of Theorem 2.7. Let p be the number of edges in H_2. Denote by $I(u)$ the Cramér rate function associated with the probability measure μ. Define

$$L(u; \beta_1, \beta_2) = \beta_1 u + \beta_2 u^p - \frac{1}{2} I(u) \tag{2.12}$$

for $u \in [0, 1]$. We consider the maximization problem for $L(u; \beta_1, \beta_2)$ on the interval $[0, 1]$, where $-\infty < \beta_1 < \infty$ and $0 \leq \beta_2 < \infty$ are parameters. We note that by Theorem 2.5, the supremum should actually be taken over the domain of I, which might differ from $[0, 1]$ at the endpoints from the discussion following Lemma 2.3. However, when the domain of I does not include 0 (or 1), $L(0)$ (or $L(1)$) is negative infinity and so can not be the maximum. To locate the maximizers of $L(u)$, we examine the properties of $L'(u)$ and $L''(u)$,

$$L'(u) = \beta_1 + p\beta_2 u^{p-1} - \frac{1}{2} I'(u), \tag{2.13}$$

$$L''(u) = p(p-1)\beta_2 u^{p-2} - \frac{1}{2} I''(u).$$

Utilizing the duality principle for the Legendre transform between $I(u)$ and $K(\theta)$, we first analyze properties of $L''(u)$ on the interval $(0, 1)$. As a consequence of the Legendre transform,

$$I(u) + K(\theta) = \theta u, \tag{2.14}$$

where θ and u are unique duals of each other. Taking derivatives, we find that

$$u = K'(\theta) \quad \text{and} \quad I''(u)K''(\theta) = 1. \tag{2.15}$$

Consider the function

$$m(u) = \frac{I''(u)}{2p(p-1)u^{p-2}} \tag{2.16}$$

on $(0, 1)$. By Equation (2.15), we may analyze the properties of $m(u)$ through the function

$$n(\theta) = 2p(p-1)K''(\theta)\left(K'(\theta)\right)^{p-2}, \tag{2.17}$$

where $\theta \in \mathbb{R}$ and $m(u)n(\theta) = 1$. From the discussion following Lemma 2.3, we recognize that

$$\lim_{\theta \to -\infty} n(\theta) = 0, \tag{2.18}$$

$$\lim_{\theta \to 0} n(\theta) = 2p(p-1)\,\mathbf{Var}(X)\left(\mathbf{E}(X)\right)^{p-2},$$

$$\lim_{\theta \to \infty} n(\theta) = 0,$$

where X is a random variable distributed according to μ. Since

$$n'(\theta) = 2p(p-1)\left(K'(\theta)\right)^{p-3}\left(K'''(\theta)K'(\theta) + (p-2)\left(K''(\theta)\right)^2\right) \tag{2.19}$$

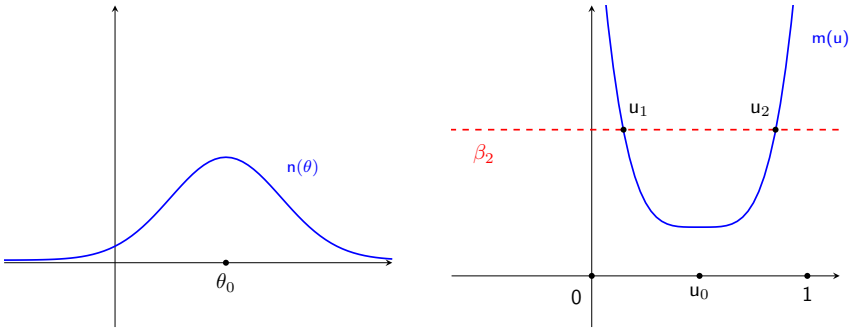

Figure 2.2: An illustrative plot of $n(\theta)$ and $m(u)$.

and $K'(\theta) > 0$ always, under Assumption 1 there exists a unique θ_0 such that $n'(\theta_0) = 0$. This unique global maximizer θ_0 for $n(\theta)$ corresponds to a unique global minimizer for $m(u)$, which we denote by u_0. Using duality, $m(u) > 0$ for all $u \in (0, 1)$ and grows unbounded on both ends. For $\beta_2 \leq m(u_0)$, $L''(u) \leq 0$ on $(0, 1)$. For $\beta_2 > m(u_0)$, $L''(u) < 0$ for $0 < u < u_1$ and $u_2 < u < 1$ and $L''(u) > 0$ for $u_1 < u < u_2$, where the transition points u_1 and u_2 satisfy $L''(u_1) = L''(u_2) = 0$. Sign properties of $L''(u)$ translate to monotonicity properties of $L'(u)$ over $(0, 1)$. For $\beta_2 \leq m(u_0)$, $L'(u)$ is decreasing over $(0, 1)$. For $\beta_2 > m(u_0)$, $L'(u)$ is decreasing from 0 to u_1, increasing from u_1 to u_2, and decreasing from u_2 to 1. See Figure 2.2 for an illustrative plot of $n(\theta)$ and $m(u)$.

The analytic properties of $L'(u)$ and $L''(u)$ entail analytic properties of $L(u)$ on the interval $[0, 1]$. Utilizing the properties pertaining to the duality of the Legendre transform in Equation (2.14) and Equation (2.15), $I(u)$ is a smooth convex function, with $I'(0) = -\infty$ and $I'(1) = \infty$. Therefore $L'(0) = \infty$ and $L'(1) = -\infty$, so $L(u)$ cannot be maximized at $u = 0$ or $u = 1$. For $\beta_2 \leq m(u_0)$, $L'(u)$ is decreasing from ∞ at 0 to $-\infty$ at 1 passing the u-axis only once. This intercept, which we denote by u^*, is the unique global maximizer for $L(u)$. Now consider $\beta_2 > m(u_0)$. If $L'(u_1) \geq 0$, then $L'(u)$ has a unique zero greater than u_2 and so $L(u)$ has a unique global maximizer at $u^* > u_2$. If $L'(u_2) \leq 0$, then $L'(u)$ has a unique zero less than u_1 and so $L(u)$ has a unique global maximizer at $u^* < u_1$.

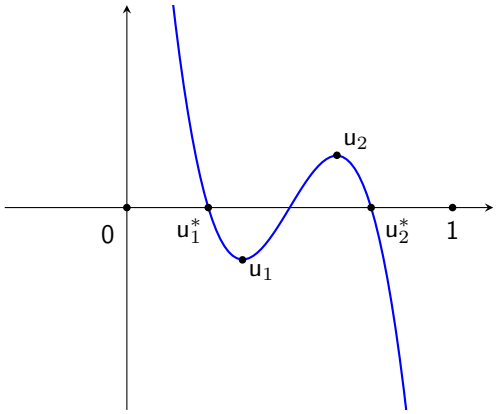

Figure 2.3: An illustrative plot of $L'(u)$ for $\beta_2 > m(u_0)$.

Lastly, suppose that $L'(u_1) < 0 < L'(u_2)$. Then $L(u)$ has two local maximizers. Denote them by u_1^* and u_2^*, with $0 < u_1^* < u_1 < u_0 < u_2 < u_2^* < 1$. See Figure 2.3 for an illustrative plot of $L'(u)$ in this case.

Define

$$f(u) = \frac{uI''(u)}{2(p-1)} - \frac{1}{2}I'(u). \tag{2.20}$$

Using $m(u_1) = m(u_2) = \beta_2$ (Equation (2.16)), $L'(u_1) = \beta_1 + f(u_1)$ and $L'(u_2) = \beta_1 + f(u_2)$. We compute

$$f'(u) = \frac{uI'''(u) - I''(u)(p-2)}{2(p-1)} = pu^{p-1}m'(u). \tag{2.21}$$

As a consequence of the relation between f' and m', following the previous analysis for m, f is decreasing on $(0, u_0)$ and increasing on $(u_0, 1)$. We check that similarly as m, f grows unbounded on both ends. Taking $u \to 0$ corresponds to taking $\theta \to -\infty$ in the dual space from Equation (2.14) and Equation (2.15), and the divergence is clear from the discussion following Lemma 2.3. To see that $f(u)$ diverges as $u \to 1$, we utilize Equation (2.21). By

the fundamental theorem of calculus,

$$f(u) - f(u_0) = \int_{u_0}^{u} f'(t)dt \geq pu_0^{p-1} \int_{u_0}^{u} m'(t)dt = pu_0^{p-1}(m(u) - m(u_0)), \qquad (2.22)$$

and grows to infinity as u approaches 1. Let X be a random variable distributed according to μ, we note some nice formulas for f and m for future reference:

$$f(\mathbf{E}(X)) = \frac{\mathbf{E}(X)}{2(p-1)\,\mathbf{Var}(X)}, \qquad (2.23)$$

$$m(\mathbf{E}(X)) = \frac{1}{2p(p-1)\,(\mathbf{E}(X))^{p-2}\,\mathbf{Var}(X)}.$$

In order for $L'(u_1) < 0$, we must have $\beta_1 < -f(u_1)$. Since f attains an absolute minimum at u_0, $f(u_1) > f(u_0)$, and then $\beta_1 < -f(u_0)$. The only possible region in the (β_1, β_2) plane where $L'(u_1) < 0 < L'(u_2)$ is thus bounded by $\beta_1 < -f(u_0)$ and $\beta_2 > m(u_0)$. Denote these two critical values for β_1 and β_2 by $\beta_1^c := -f(u_0)$ and $\beta_2^c := m(u_0)$.

Recall that $u_1 < u_0 < u_2$. By monotonicity of $f(u)$ on the intervals $(0, u_0)$ and $(u_0, 1)$, there exist continuous functions $a(\beta_1)$ and $b(\beta_1)$ of β_1, such that $L'(u_1) < 0$ for $u_1 > a(\beta_1)$ and $L'(u_2) > 0$ for $u_2 > b(\beta_1)$. As $\beta_1 \to -\infty$, $a(\beta_1) \to 0$ and $b(\beta_1) \to 1$. $a(\beta_1)$ is an increasing function of β_1, whereas $b(\beta_1)$ is a decreasing function, and they satisfy $f(a(\beta_1)) = f(b(\beta_1)) = -\beta_1$. The restrictions on u_1 and u_2 yield restrictions on β_2, and we have $L'(u_1) < 0$ for $\beta_2 < m(a(\beta_1))$ and $L'(u_2) > 0$ for $\beta_2 > m(b(\beta_1))$. As $\beta_1 \to -\infty$, $m(a(\beta_1)) \to \infty$ and $m(b(\beta_1)) \to \infty$. $m(a(\beta_1))$ and $m(b(\beta_1))$ are both decreasing functions of β_1, and they satisfy $L'(u_1) = 0$ when $\beta_2 = m(a(\beta_1))$ and $L'(u_2) = 0$ when $\beta_2 = m(b(\beta_1))$. As $L'(u_2) > L'(u_1)$ for every (β_1, β_2), the curve $m(b(\beta_1))$ lies below the curve $m(a(\beta_1))$, and together they generate the bounding curves of the V-shaped region in the (β_1, β_2) plane with corner point (β_1^c, β_2^c) where two local maximizers exist for $L(u)$.

By (Equation (2.21)), for sufficiently negative values of β_1, $f(a(\beta_1)) < m(a(\beta_1))$ and $f(b(\beta_1)) > m(b(\beta_1))$, so the straight line $\beta_1 = -\beta_2$ lies within this region.

Fix an arbitrary $\beta_1 < \beta_1^c$. Then $L'(u)$ shifts upward as β_2 increases and downward as β_2 decreases. As a result, as β_2 gets large, the positive area bounded by the curve $L'(u)$ increases, whereas the negative area decreases. By the fundamental theorem of calculus, the difference between the positive and negative areas is the difference between $L(u_2^*)$ and $L(u_1^*)$, which goes from negative ($L'(u_2) = 0$, u_1^* is the global maximizer) to positive ($L'(u_1) = 0$, u_2^* is the global maximizer) as β_2 goes from $m(b(\beta_1))$ to $m(a(\beta_1))$. Thus there must be a unique β_2: $m(b(\beta_1)) < \beta_2 < m(a(\beta_1))$ such that u_1^* and u_2^* are both global maximizers, and we denote this β_2 by $r(\beta_1)$. The parameter values of $(\beta_1, r(\beta_1))$ are exactly the ones for which positive and negative areas bounded by $L'(u)$ equal each other. An increase in β_1 induces an upward shift of $L'(u)$, and may be balanced by a decrease in β_2. Similarly, a decrease in β_1 induces a downward shift of $L'(u)$, and may be balanced by an increase in β_2. This justifies that $r(\beta_1)$ is monotonically decreasing in β_1. See Figure 2.1. Here we let X be a random variable distributed according to Beta$(2, 2)$, then $\mathbf{E}(X) = 1/2$ and $\mathbf{Var}(X) = 1/20$. By Lemma 2.2, $\theta_0 = 0$ and $u_0 = \mathbf{E}(X) = 1/2$, which by Equation (2.23) gives $(\beta_1^c, \beta_2^c) = (-5, 5)$. Also see Figure 1 in [60] and Figure 1 in [73] for related phase transition plots when the edge-weights distribution μ is respectively Bernoulli$(.5)$ and Uniform$(0, 1)$.

The rest of the proof follows as in the proof of the corresponding result (Theorem 2.1) in Radin and Yin [60], where some probability estimates were used. A (jump) discontinuity in the first derivatives of ψ_∞^β across the curve $\beta_2 = r(\beta_1)$ indicates a discontinuity in the expected local densities, while the divergence of the second derivatives of ψ_∞^β at the critical point (β_1^c, β_2^c) implies that the covariances of the local densities go to zero more slowly than $1/n^2$. We omit the proof details.

■

Remark 2. The maximization problem in Equation (2.9) is solved at a unique value u^* off the phase transition curve $\beta_2 = r(\beta_1)$, and at two values u_1^* and u_2^* along the curve. As $\beta_1 \to -\infty$ (resp. $\beta_2 \to \infty$), $u_1^* \to 0$ and $u_2^* \to 1$. The jump from u_1^* to u_2^* is quite noticeable even for small parameter values of β. For example, taking $p = 2$, $\beta_1 = -8$, and $\beta_2 = 8$ in Beta$(2, 2)$, numerical computations yield that $u_1^* \approx 0.165$ and $u_2^* \approx 0.835$.

2.5 Universal asymptotics

We now examine near degeneracy and universality in general edge-weighted exponential random graphs. All our findings in this section are derived based on the assumption that the non-degenerate probability measure μ for the edge weights is symmetric about the line $u = 1/2$. We remark that near degeneracy and universality are expected even when the edge weights are not symmetrically distributed, except that the universal straight line gets shifted vertically from $\beta_2 = -\beta_1$.

Proposition 2.9. Consider a non-degenerate probability measure μ supported on $[0, 1]$ and symmetric about the line $u = 1/2$. Take H_1 a single edge and H_2 a finite simple graph with $p \geq 2$ edges. The phase transition curve $\beta_2 = r(\beta_1)$ lies above the straight line $\beta_2 = -\beta_1$ when $p \geq 3$, and is exactly the portion of the straight line $\beta_2 = -\beta_1$ ($\beta_1 \leq -1/(4\,\mathbf{Var}(X))$ when $p = 2$. Here X is a random variable distributed according to μ.

Proof. From the proof of Theorem 2.7, there are two global maximizers u_1^* and u_2^* for $L(u)$ along the phase transition curve $\beta_2 = r(\beta_1)$, $0 < u_1^* < u_0 < u_2^* < 1$, where u_0 is the unique global minimizer for $m(u)$ from Equation (2.16). By Lemma 2.2, $u_0 = 1/2$ when $p = 2$ and $u_0 > 1/2$ when $p > 2$. Furthermore, the y-coordinate β_2^c of the critical point $(\beta_1^c, \beta_2^c) = (-f(u_0), m(u_0))$ is always positive. On the straight line $\beta_1 + \beta_2 = 0$, we rewrite $L(u) = \beta_1(u - u^p) - I(u)/2$. By Proposition 2.4, $I(u)$ is symmetric about the line $u = 1/2$. First suppose $p = 2$. Since $I(u)$ and $u - u^2$ are both symmetric, two global maximizers u_1^* and u_2^* exist for $L(u)$ and $(-f(u_0), m(u_0)) = (-1/(4\,\mathbf{Var}(X)), 1/(4\,\mathbf{Var}(X)))$ by

Equation (2.23). Next consider the generic case $p \geq 3$. Analytical calculations give that $u - u^p < (1-u) - (1-u)^p$ for $0 < u < 1/2$. Since $I(u)$ is symmetric, this says that for $\beta_1 < 0$ (resp. $\beta_2 > 0$), the global maximizer u^* of $L(u)$ satisfies $u^* \leq 1/2$ and so must be u_1^*. The conclusion readily follows. ■

Proposition 2.10. Consider a non-degenerate probability measure μ supported on $[0,1]$ and symmetric about the line $u = 1/2$. Assume the associated Cramér rate function in Equation (2.3) is bounded on $[0,1]$ (i.e. $I(0) = I(1)$ is finite). Take H_1 a single edge and H_2 a finite simple graph with $p \geq 2$ edges. The phase transition curve $\beta_2 = r(\beta_1)$ displays a universal asymptotic behavior as $\beta_1 \to -\infty$, specifically,

$$\lim_{\beta_1 \to -\infty} |r(\beta_1) + \beta_1| = 0. \tag{2.24}$$

Proof. Let $\beta_2 = -\beta_1 + \delta$ with $\delta > 0$ fixed. Define $F(u;\beta_1) = \beta_1(u - u^p)$ and $G(u;\delta) = \delta u^p - I(u)/2$ so that $L(u;\beta_1,\beta_2) = F(u;\beta_1) + G(u;\delta)$ by Equation (2.12). We will show, for sufficiently negative β_1, that the global maximizer u^* of $L(u)$ equals u_2^*. Together with Proposition 2.9, this implies that for these β_1, $-\beta_1 \leq r(\beta_1) \leq -\beta_1 + \delta$, which will prove the desired limit.

Under our assumption, $-I(u)$ is a continuous symmetric function that increases on $(0,1/2)$ and decreases on $(1/2,1)$, with a maximum attained at $u = 1/2$ and $-I(1/2) = 0$. Denote by $C := -I(0)/2 = -I(1)/2$ so that C is finite and negative and $G(0) = C$. Recall that $0 < u_1^* < u_0 < u_2^* < 1$, where u_1^* and u_2^* are two local maximizers for $L(u)$ and $u_0 \geq 1/2$ is the unique global minimizer for $m(u)$ (Equation (2.16)) that does not depend on β_1 and β_2. Rigorously, it may be that only one local maximizer u_1^* or u_2^* exist for $L(u)$, but this does not affect our argument below. From the continuity and boundedness of G on $[0,1]$, there exists $\eta \in (0, 1-u_0)$ such that if $0 \leq u < \eta$ then $G(u) - C < \delta/2$. Since $u - u^p = u(1-u^{p-1}) > 0$ on $(0,1)$ and vanishes at the endpoints 0 and 1, there

exists $\beta < 0$ such that for all $\beta_1 < \beta$ and $u \in [\eta, 1-\eta]$, $F(u) < C - \delta$ and therefore $L(u) < C - \delta + G(u) < C = L(0)$, so $u^* \in [0, \eta) \cup (1-\eta, 1]$. Similarly, using that $F(u) \leq 0$ for all $\beta_1 < 0$ and all $u \in [0, \eta)$, we have $L(u) \leq G(u) < C + \delta/2 < C + \delta = L(1)$ so $u^* \in (1-\eta, 1]$. Since $u_1^* < u_0 < 1 - \eta$, this says that $u^* = u_2^*$. ■

Proposition 2.9 and Proposition 2.10 have advanced our understanding of phase transitions in edge-weighted exponential random graphs, yet some fundamental questions remain unanswered. As explained in Section 2.4, a typical graph sampled from the exponential model looks like an Erdős-Rényi graph $G(n, u)$ in the large n limit, where the asymptotic edge presence probability $u(\beta_1, \beta_2) \to 0$ or 1 is prescribed according to the maximization problem (Equation (2.9)). However, the speed of u towards these two degenerate states is not at all clear. When a typical graph is sparse ($u \to 0$), how sparse is it? When a typical graph is nearly complete ($u \to 1$), how dense is it? Can we give an explicit characterization of the near degenerate graph structure as a function of the parameters? Theorems 2.11 and 2.12 are dedicated toward these goals.

Theorem 2.11. Consider a non-degenerate probability measure μ supported on $[0, 1]$ and symmetric about the line $u = 1/2$. Take H_1 a single edge and H_2 a finite simple graph with $p \geq 2$ edges. Let $\beta_1 < -\beta_2$ and $\beta_2 \geq 0$. For large n and (β_1, β_2) sufficiently far away from the origin, a typical graph drawn from the model looks like an Erdős-Rényi graph $G(n, u)$, where the edge presence probability u depends on the distribution μ, but its dual θ universally satisfies $\theta \asymp 2\beta_1$.

Proof. Let $\beta_1 = a\beta_2$ with $a < -1$. Resorting to Legendre duality, Equation (2.9) gives a condition on θ, the dual of u:

$$\beta_1 + p\beta_2 (K'(\theta))^{p-1} = \frac{1}{2}\theta. \tag{2.25}$$

By Proposition 2.9, $u \to 0$ for (β_1, β_2) sufficiently far away from the origin, which corresponds to $\theta \to -\infty$ in the dual space. From Table 2.1, $K'(\theta) \to 0$ as $\theta \to -\infty$, we have

$$\frac{\theta}{2\beta_2} = a + p(K'(\theta))^{p-1} \to a. \tag{2.26}$$

The universal asymptotics of $\theta \asymp 2\beta_1$ is verified.

We claim that u on the other hand depends on the specific distribution μ. We will derive the asymptotics of u in two special cases, Bernoulli$(.5)$ and Uniform$(0, 1)$. In both cases, $u = K'(\theta)$ by Legendre duality. For Bernoulli$(.5)$,

$$u = K'(\theta) = \frac{e^\theta}{1 + e^\theta} \asymp e^\theta \asymp e^{2\beta_1}. \tag{2.27}$$

While for Uniform$(0, 1)$,

$$u = K'(\theta) = \frac{e^\theta}{e^\theta - 1} - \frac{1}{\theta} \asymp -\frac{1}{\theta} \asymp -\frac{1}{2\beta_1}. \tag{2.28}$$

■

Theorem 2.12. Consider a non-degenerate probability measure μ supported on $[0, 1]$ and symmetric about the line $u = 1/2$. Assume the associated Cramér rate function (Equation (2.3)) is bounded on $[0, 1]$ (i.e. $I(0) = I(1)$ is finite). Take H_1 a single edge and H_2 a finite simple graph with $p \geq 2$ edges. Let $\beta_1 > -\beta_2$ and $\beta_2 \geq 0$. For large n and (β_1, β_2) sufficiently far away from the origin, a typical graph drawn from the model looks like an Erdős-Rényi graph $G(n, u)$, where the edge presence probability u depends on the distribution μ, but its dual θ universally satisfies $\theta \asymp 2(\beta_1 + p\beta_2)$.

Proof. Let $\beta_1 = a\beta_2$ with $a > -1$. Resorting to Legendre duality, Equation (2.9) gives condition Equation (2.25) on θ, the dual of u. By Proposition 2.10, $u \to 1$ for (β_1, β_2) sufficiently far away from the origin, which corresponds to $\theta \to \infty$ in the dual space. From

Table 2.1, $K'(\theta) \to 1$ as $\theta \to \infty$, we have

$$\frac{\theta}{2\beta_2} = a + p(K'(\theta))^{p-1} \to a + p. \tag{2.29}$$

The universal asymptotics of $\theta \asymp 2(\beta_1 + p\beta_2)$ is verified.

We claim that u on the other hand depends on the specific distribution μ. We will derive the asymptotics of u in two special cases, Bernoulli$(.5)$ and Uniform$(0, 1)$. In both cases, $u = K'(\theta)$ by Legendre duality. For Bernoulli$(.5)$,

$$K'(\theta) = \frac{e^\theta}{1 + e^\theta} \asymp 1 - e^{-\theta} \asymp 1 - e^{-2(\beta_1 + p\beta_2)}. \tag{2.30}$$

While for Uniform$(0, 1)$,

$$K'(\theta) = \frac{e^\theta}{e^\theta - 1} - \frac{1}{\theta} \asymp 1 - \frac{1}{\theta} \asymp 1 - \frac{1}{2(\beta_1 + p\beta_2)}. \tag{2.31}$$

■

See Table 2.2 and Table 2.3. Even for β with small magnitude, the asymptotic tendency of the optimal θ (hence the optimal u) is quite evident. Here we take $p = 2$. The asymptotic characterizations of u obtained in Theorem 2.11 and Theorem 2.12 make possible a deeper analysis of the asymptotics of the limiting normalization constant ψ_∞^β of the exponential model in the following Theorem 2.13 and Theorem 2.14. Interestingly, universality is observed only in the nearly complete region but not the sparse region of the parameter space.

Theorem 2.13. Consider a non-degenerate probability measure μ supported on $[0, 1]$ and symmetric about the line $u = 1/2$. Take H_1 a single edge and H_2 a finite simple graph with

β_1	β_2	θ_{opt}	u_{opt}	$\exp(2\beta_1)$	$1-\exp(-2(\beta_1+p\beta_2))$
-2	-4	-4.23	0.014	0.018	
1	1	5.99	0.998		0.998

Table 2.2: Asymptotic comparison for Bernoulli$(.5)$ near degeneracy.

β_1	β_2	θ_{opt}	u_{opt}	$-1/(2\beta_1)$	$1-1/(2(\beta_1+p\beta_2))$
-4	-6	-10.32	0.097	0.125	
3	2	13.40	0.925		0.929

Table 2.3: Asymptotic comparison for Uniform$(0,1)$ near degeneracy.

$p \geq 2$ edges. Let $\beta_1 < -\beta_2$ and $\beta_2 \geq 0$. For (β_1, β_2) sufficiently far away from the origin, the limiting normalization constant ψ_∞^β depends on the distribution μ.

Proof. Let $\beta_1 = a\beta_2$ with $a < -1$. By Theorem 2.5,

$$\psi_\infty^\beta = \beta_1 u + \beta_2 u^p - \frac{1}{2} I(u), \tag{2.32}$$

where u is chosen so that the above equation is maximized and $u \to 0$ for (β_1, β_2) sufficiently far away from the origin. Resorting to Legendre duality, this gives

$$\psi_\infty^\beta = \beta_1 K'(\theta) + \beta_2 (K'(\theta))^p - \frac{1}{2}\left(\theta K'(\theta) - K(\theta)\right), \tag{2.33}$$

where θ is the dual of u and approaches $-\infty$ when (β_1, β_2) diverge. By Equation (2.25),

$$\psi_\infty^\beta = (1-p)\beta_2 (K'(\theta))^p + \frac{1}{2} K(\theta). \tag{2.34}$$

Since $\beta_2 \asymp \theta/(2a)$ as $\theta \to -\infty$ from Theorem 2.11, asymptotically we have

$$\begin{aligned} \psi_\infty^\beta &\asymp \frac{1-p}{2a}\theta(K'(\theta))^p + \frac{1}{2}K(\theta) \\ &\asymp (1-p)\beta_2\left(K'(2\beta_1)\right)^p + \frac{1}{2}K(2\beta_1). \end{aligned} \tag{2.35}$$

■

Remark 3. Many common distributions including Bernoulli$(.5)$ and Uniform$(0,1)$ satisfy $\theta K'(\theta)/K(\theta) \to 0$ as $\theta \to -\infty$, in which case the asymptotics in Theorem 2.13 may be further reduced to $\psi_\infty^\beta \asymp K(\theta)/2 \asymp K(2\beta_1)/2$.

Theorem 2.14. Consider a non-degenerate probability measure μ supported on $[0,1]$ and symmetric about the line $u = 1/2$. Assume the associated Cramér rate function (Equation (2.3)) is bounded on $[0,1]$ (i.e. $I(0) = I(1)$ is finite). Take H_1 a single edge and H_2 a finite simple graph with $p \geq 2$ edges. Let $\beta_1 > -\beta_2$ and $\beta_2 \geq 0$. For (β_1, β_2) sufficiently far away from the origin, the limiting normalization constant ψ_∞^β universally satisfies $\psi_\infty^\beta \asymp \beta_1 + \beta_2$.

Proof. Let $\beta_1 = a\beta_2$ with $a > -1$. Similarly as in the proof of Theorem 2.13, Theorem 2.5 gives Equation (2.32), where u is chosen so that the equation is maximized and $u \to 1$ for (β_1, β_2) sufficiently far away from the origin. Since the first two terms diverge to $\beta_1 + \beta_2$ while the last term is bounded by our assumption, the claim easily follows. ■

Remark 4. The boundedness assumption on I in Theorem 2.14 is only used as a sufficient condition to ensure that $u \to 1$ for $\beta_1 > -\beta_2$ in the upper half-plane and far away from the origin and is not necessary for the derivation of the universal asymptotics for ψ_∞^β. Indeed, since $\theta \asymp 2(\beta_1 + p\beta_2)$ by Theorem 2.12, using $K(\theta)/\theta \asymp K'(\theta) \asymp 1$ in Equation (2.34), we have

$$\psi_\infty^\beta \asymp (1-p)\beta_2 + (\beta_1 + p\beta_2) \asymp \beta_1 + \beta_2. \tag{2.36}$$

This universal asymptotic phenomenon is observed for example in Uniform$(0, 1)$, whose associated Cramér rate function I is not bounded.

Chapter 3: Mixing time of vertex-weighted exponential random graphs

3.1 Introduction

We now turn our attention to the evolution of vertex-weighted exponential random graphs. As motivation, consider the dynamics of spreading events in a complex network. There are sensitive control points collectively known as "influential spreaders", whose infection maximizes the overall fraction of infected vertices. For information diffusion over Twitter for example, the influential spreaders may be thought of as a celebrity or a news source. Vertices in the network thus carry with themselves some distinguishing features, a phenomenon that could not be directly modeled by standard exponential random graphs since their underlying probability space consists of simple graphs only. By placing weights on the vertices, this section addresses this limitation of the exponential model.

Before proceeding further, we provide another reason why the vertex-weighted model may be of interest [10]. Let $\dot{\mathcal{G}}_n$ be the set of all vertex-weighted labeled graphs G_n on n vertices. Assume that the vertex weights are iid real random variables subject to a common distribution ν supported on $[0, 1]$, the edge weight between two vertices is a product of the vertex weights, and the triangle weight among three vertices is a product of the edge weights. Let U be a random variable distributed according to ν and denote its expectation with respect to ν by $\mathbf{E}$. Further denote the expected edge weight of $G_n \in \dot{\mathcal{G}}_n$ by e and the expected triangle weight by t. Then we have $e = \mathbf{E}(U)^2$ and $t = (\mathbf{E}(U^2))^3$. Note that by suitably choosing ν, the entire region between the upper boundary of the realizable edge-triangle densities and the Erdős-Rényi curve ($e^3 \leq t \leq e^{3/2}$) may be attained for

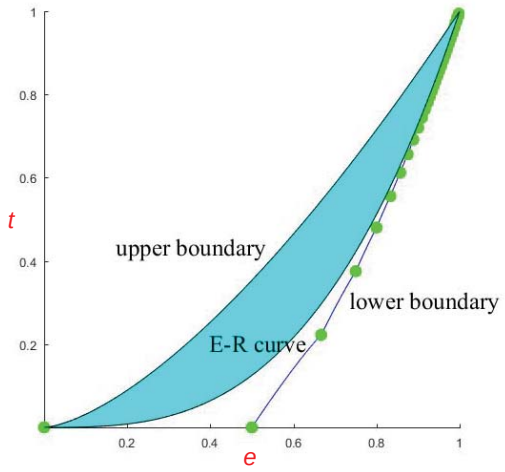

Figure 3.1: The cyan region shows where the expected edge and triangle densities lie for a vertex-weighted random graph model.

vertex-weighted random graphs. See Figure 3.1. If we take U to be Bernoulli, the upper boundary is reproduced:

$$\mathbf{E}(U^2) = \mathbf{E}(U) \implies t = e^{3/2}. \tag{3.1}$$

If we take U to be a constant a.s., the Erdős-Rényi curve is recovered:

$$\mathbf{E}(U^2) = \mathbf{E}(U)^2 \implies t = e^3. \tag{3.2}$$

By contrast, since simple graphs may be interpreted as having iid Bernoulli$(.5)$ weights on the edges, the underlying graph space of standard exponential random graphs lies at a single point $(1/2, 1/8)$ on the Erdős-Rényi curve. Even without incorporating the exponential construction, assigning vertex weights alone adds intriguing characteristics to the model.

3.1.1 The model. In this section we will restrict our attention to vertex-weighted exponential random graph models where the vertex weights take values in 0 and 1 only. We include in the exponent a combination of edge and triangle densities, both with non-

negative parameters. Even under this simplification, the vertex-weighted model depicts captivating behaviors in large-scale networks. Recall that by Theorem 1.10 in Section 1.3, exponential random graph models with non-negative parameters behave like Erdős-Rényi random graphs in the large n limit; however, this is not the case in this section as we are considering the model only for finite n. Instead of Erdős-Rényi, this model emphasizes the formation of cliques, and is particularly suited for the modeling of a broad range of social networks. Consider the Facebook friend graph for example, where we make the simplifying assumption that a person is either interested (vertex value 1) or not (vertex value 0) in building a friendship. Then having value 1 at three distinct vertices will force the formation of a triangle rather than a two-star, which is rooted in the notion that a friend of a friend is more likely to be a friend. An added benefit of this setting is that the model may be considered as an extension of the lattice gas (Ising) model on a graph, and the techniques of spin models may be employed in our investigation [69].

A graph $G_n \in \dot{\mathcal{G}}_n$ may be viewed as an element $X \in \mathcal{X} := \{0,1\}^n$, referred to as *configurations*, that attributes weights 0 or 1 to the ordered vertices of G_n. Denote by $X(i)$ the weight of vertex i. Borrowing terminology from spin models, the vertex weight $X(i)$ will be called the *spin* at i. By iid-ness, the spins at different vertices are independent, and subject to a common distribution ν that assumes value 0 with probability $1-p$ and 1 with probability p for some $p \in (0,1)$. Let H_1 be the number of edges for the configuration X and H_2 be the number of triangles. They may be formulated explicitly in terms of lattice gas (Ising) spins

$$H_1(X) = \sum_{i \neq j} X(i)X(j) \text{ and } H_2(X) = \sum_{i \neq j \neq k} X(i)X(j)X(k), \tag{3.3}$$

where the inequality $n_1 \neq n_2 \neq \cdots \neq n_k$ means that $n_i \neq n_j$ for any $i \neq j$. We rescale the edge and triangle parameters in the exponent, $H = (\alpha_1/n)H_1 + (\alpha_2/n^2)H_2$, so that the

total contribution of a single vertex to the weights is $O(1)$. We are now ready to introduce a Gibbs distribution on the set of spin configurations $\mathcal{X}$. To avoid cumbersome notation, we suppress the n-dependence in many of the quantities under discussion.

Definition 3.1. Take $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, and $p \in (0,1)$. Let $X \in \mathcal{X}$ be a spin configuration. Denote by $\omega(X)$ the number of vertices with spin 1 in X. Assign a Gibbs probability measure on $\mathcal{X}$ as

$$\pi(X) = Z^{-1} \exp(H(X)) p^{\omega(X)} (1-p)^{n-\omega(X)}, \tag{3.4}$$

where $H = (\alpha_1/n)H_1 + (\alpha_2/n^2)H_2$ is the combination of edge and triangle weights and $Z = Z(n, p, \alpha_1, \alpha_2)$ is the normalizing constant, also referred to as the partition function.

The configuration space $\mathcal{X}$ can be partially ordered in the sense that for $X, Y \in \mathcal{X}$, we say that $X \leq Y$ if and only if $X(i) \leq Y(i)$ for every $i \in \{1, \ldots, n\}$. To model the evolution of the network towards equilibrium, we will adopt (single-site) Glauber dynamics, which is a discrete-time irreducible and aperiodic Markov chain $(X_t)_{t=0}^{\infty}$ on $\mathcal{X}$. Under the Glauber dynamics, the random graph evolves by selecting a vertex i and updating the spin $X(i)$ according to π conditioned to agree with the spins at all vertices not equal to i. By sampling from the exponential distribution using Glauber dynamics, we learn the global structure of the network as well as parameters describing the interactions. Explicitly, let $X \in \mathcal{X}$ be a configuration and set the initial state $X_0 = X$. The next step of the Markov chain, X_1, is obtained as follows. Choose a vertex i uniformly from $\{1, \ldots, n\}$. Let $X_1(j) = X(j)$ for all $j \neq i$, $X_1(i) = 1$ with probability P_+ and $X_1(i) = 0$ with probability P_-, where the update probabilities P_+ and P_- are given by

$$P_+(X, i) = \frac{p \exp(H'(X, i))}{p \exp(H'(X, i)) + (1-p)} \tag{3.5}$$

and

$$P_{-}(X,i) = \frac{1-p}{p\exp(H'(X,i)) + (1-p)}. \tag{3.6}$$

Here $H' = (\alpha_1/n)S + (\alpha_2/n^2)T$ depends only on the spins at vertices other than i, with

$$S(X,i) = \sum_{j\neq i} X(j) \text{ and } T(X,i) = \sum_{j\neq k\neq i} X(j)X(k) = \frac{S(X,i)(S(X,i)-1)}{2}. \tag{3.7}$$

For $X, Y \in \mathcal{X}$, the transition matrix for the Glauber dynamics is then

$$P(X,Y) = \frac{1}{n}\sum_{i} \frac{f(Y(i))e^{Y(i)H'(X,i)}}{f(Y(i))e^{Y(i)H'(X,i)} + f(1-Y(i))e^{(1-Y(i))H'(X,i)}} \mathbf{1}_{\left\{ \substack{Y(j)=X(j) \\ j\neq i} \right\}}, \tag{3.8}$$

where $Y(i) \in \{0,1\}$ and we define f such that $f(0) = 1-p$ and $f(1) = p$ to lighten the notation.

3.1.2 Mixing time. The Gibbs distribution π is stationary and reversible for the Glauber dynamics chain. By the convergence theorem for ergodic Markov chains, the Glauber dynamics will converge to the stationary distribution and our goal is to obtain some estimates on the mixing time, since it greatly affects the efficiency of simulation studies and sampling algorithms [12] [21]. Given $\varepsilon > 0$, the mixing time for this Markov chain is defined as

$$t_{\text{mix}}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}, \tag{3.9}$$

where

$$d(t) = \max_{X\in\mathcal{X}} \left\| P^t(X,\cdot) - \pi \right\|_{\text{TV}} \tag{3.10}$$

measures the total variation distance to stationarity of the Glauber dynamics chain after t steps. As is standard, we take $t_{\text{mix}} = t_{\text{mix}}(1/4)$. The mixing time is thus defined to be the minimum number of discrete time steps such that, starting from an arbitrary configuration X, the chain is within total variation distance $1/4$ from the stationary distribution π. For

background on mixing times, see Aldous and Fill [5] and Levin et al. [45]. Our results will indicate that the mixing time can vary enormously depending on the choice of parameters.

3.1.3 Normalized magnetization. Given a spin configuration $X \in \mathcal{X}$, the normalized magnetization c of X is defined as

$$c(X) = \frac{1}{n}\sum_{i=1}^{n} X(i). \tag{3.11}$$

The image of a Markov chain under a map is not usually itself a Markov chain; however, in this case, it turns out that the normalized magnetization is a Markov chain. Adopting (single-site) Glauber dynamics on $\mathcal{X}$, $(c_t)_{t=0}^{\infty}$ is a projection of the chain $(X_t)_{t=0}^{\infty}$ and so is also aperiodic and irreducible. Set the initial state $c_0 = c$. From the mechanism described in Section 3.1.1, after one Glauber update, c_1 will take on one of three values: $c - 1/n$, c, or $c+1/n$. If a spin 0 vertex is chosen and updated to spin 1, $c_1 = c+1/n$. Alternatively, if a spin 1 vertex is chosen and updated to spin 0, $c_1 = c - 1/n$. When no spins are updated, $c_1 = c$. By Equation (3.5), the probability that we select a spin 0 vertex and update it to spin 1 is

$$P_u = \frac{n - cn}{n}\frac{p\exp\left(\alpha_1 c + \frac{\alpha_2}{2}c(c-\frac{1}{n})\right)}{p\exp\left(\alpha_1 c + \frac{\alpha_2}{2}c(c-\frac{1}{n})\right) + (1-p)}. \tag{3.12}$$

Similarly, by Equation (3.6), the probability that we select a spin 1 vertex and update it to spin 0 is

$$P_d = \frac{cn}{n}\frac{1-p}{p\exp\left(\alpha_1(c-\frac{1}{n}) + \frac{\alpha_2}{2}(c-\frac{2}{n})(c-\frac{1}{n})\right) + (1-p)}. \tag{3.13}$$

Combining Equation (3.12) and Equation (3.13), the magnetization c_t moves up with probability P_u, down with probability P_d, and remains unchanged with probability $1 - P_u - P_d$. For n large enough, $P_u \asymp (1-c)\lambda(c)$ and $P_d \asymp c(1-\lambda(c))$, where

$$\lambda(c) = \frac{p\exp\left(\alpha_1 c + \frac{\alpha_2}{2}c^2\right)}{p\exp\left(\alpha_1 c + \frac{\alpha_2}{2}c^2\right) + (1-p)} \tag{3.14}$$

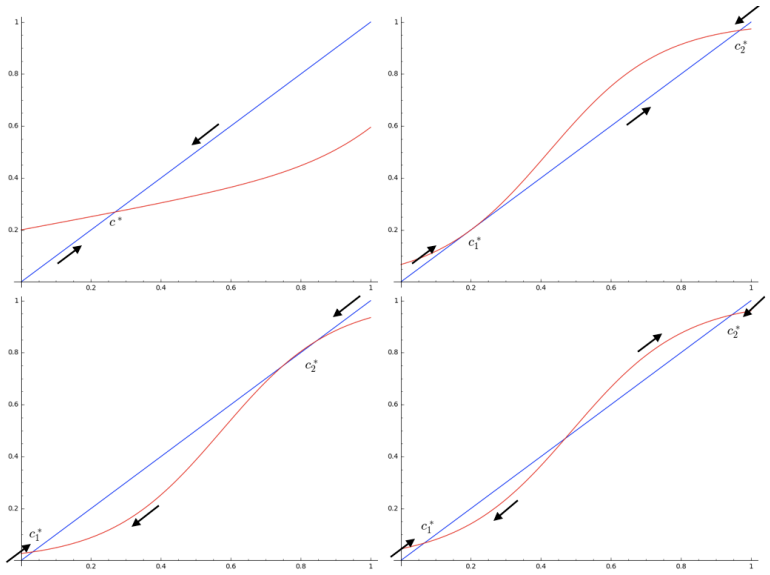

Figure 3.2: Behavior of the λ function in different regions of the parameter space with arrows indicating whether the fixed point is an attractor or a repellor.

represents the asymptotic probability that a chosen vertex is updated to spin 1. This implies that the expected magnetization drift is asymptotically $(\lambda(c) - c)/n$, and a rigorous justification may be found in Lemma 3.1.

3.1.4 Phase classification. The magnetization chain (c_t) is a deciding factor in the convergence of the Glauber dynamics chain (X_t). Note that $0 \leq c_t \leq 1$ and λ is a smooth and increasing function on $[0, 1]$. Since $\lambda(0) > 0$ and $\lambda(1) < 1$, $\lambda(c) = c$ admits at least one solution in $(0, 1)$. If the solution c is unique and not an inflection point, i.e. $\lambda'(c) < 1$ (referred to as the *high-temperature phase*), then independent of the initial position all configurations will be driven towards it, and the burn-in stage will cost $O(n)$ steps. The burn-in stage is the time required for the magnetization chain to be close to an attractor λ with high probability for any suitably chosen initial configuration. See the upper left plot of Figure 3.2. Conversely, if there exist at least two solutions c such that $\lambda'(c) < 1$ (referred to as the *low-temperature phase*), then the burn-in procedure will take the configurations to different attractor states depending on their initial positions. See the lower right plot of

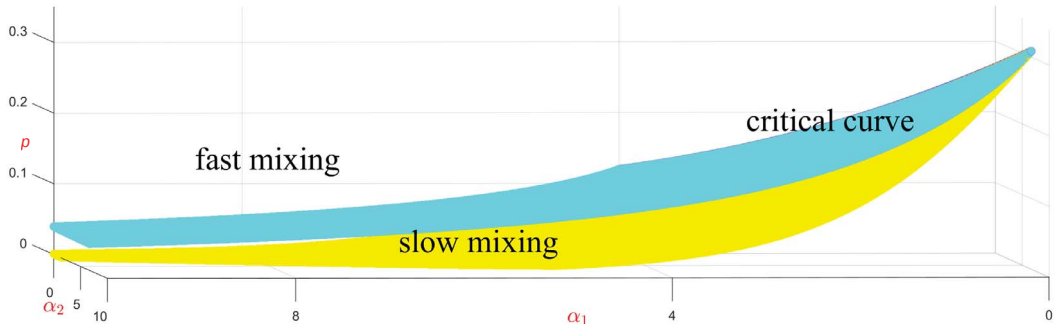

Figure 3.3: Surfaces in the parameter space illustrating the region with fast vs. slow mixing and identifying the critical curve.

Figure 3.2. Once the configuration is close to an attractor, the Glauber dynamics allows an exponentially small flow of probability for it to leave. A detailed examination of the burn-in period will be provided in Section 3.2.

In Section 3.3, through estimating the average distance after one update between two coupled configurations that agree everywhere except at a single vertex, we show that the Glauber dynamics X_t mixes in $O(n \log n)$ steps in the high-temperature phase. Relating to coupon collecting and employing spectral methods, the same asymptotic lower bound $\Omega(n \log n)$ is validated. While in Section 3.4, by a conductance argument using the Cheeger inequality, we establish exponentially slow mixing of the Glauber dynamics X_t in the low-temperature phase. Finally, in Section 3.5, we give evidence that the burn-in will cost $O(n^{3/2})$ steps along the "critical curve", and the Glauber dynamics X_t is thus expected to mix in $O(n^{3/2})$ steps. See Figure 3.3. The cyan and canary yellow surfaces separate the high and low-temperature phases, with their intersection marked by the critical curve. Convergence of the Glauber dynamics X_t elsewhere on the two surfaces corresponds to the situation where $\lambda(c) = c$ has at least two solutions and one solution c satisfies $\lambda'(c) = 1$. See the upper and right and lower left plots of Figure 3.2. The mixing time largely depends on the movement of the chain around the inflection point and is not addressed.

3.2 Burn-in period

We start by running the Glauber dynamics for an initial burn-in period. This will ensure that the associated magnetization chain is with high probability close to an attractor of λ. Let $X \in \mathcal{X}$ be any spin configuration. Set the initial state $X_0 = X$ and let c_t be the normalized magnetization of X_t at time t. We use $\mathbf{P}_X$ and $\mathbf{E}_X$ respectively to denote the underlying probability measure and associated expectation. To keep the notation light, we omit the explicit dependence on X when it is clear from the context.

Lemma 3.1. The expected drift in c_t after one step of the Glauber dynamics, starting from the configuration X, is given by

$$\mathbf{E}(c_{t+1} - c_t \mid c_t) = \frac{1}{n}\left(\lambda(c_t) - c_t\right) + O\left(\frac{1}{n^2}\right), \tag{3.15}$$

where λ is defined as in Equation (3.14).

Proof. From our discussion in Section 3.1.3, we compute

$$\begin{aligned}\mathbf{E}(c_{t+1} - c_t \mid c_t) = \frac{1}{n}(1 - c_t)\frac{p\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}c_t\left(c_t - \frac{1}{n}\right)\right)}{p\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}c_t\left(c_t - \frac{1}{n}\right)\right) + (1-p)} \\ - \frac{1}{n}c_t\frac{1-p}{p\exp\left(\alpha_1\left(c_t - \frac{1}{n}\right) + \frac{\alpha_2}{2}\left(c_t - \frac{1}{n}\right)\left(c_t - \frac{2}{n}\right)\right) + (1-p)}.\end{aligned} \tag{3.16}$$

Note that lower order fluctuations may be extracted from the exponents since

$$\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}c_t\left(c_t - \frac{1}{n}\right)\right) = \exp\left(-\frac{\alpha_2}{2n}c_t\right)\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}\left(c_t\right)^2\right), \tag{3.17}$$

and

$$\begin{aligned}\exp\left(\alpha_1\left(c_t - \frac{1}{n}\right) + \frac{\alpha_2}{2}\left(c_t - \frac{1}{n}\right)\left(c_t - \frac{2}{n}\right)\right) \\ = \exp\left(-\frac{\alpha_1}{n} - \frac{3\alpha_2}{2n}c_t + \frac{\alpha_2}{n^2}\right)\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}\left(c_t\right)^2\right),\end{aligned} \tag{3.18}$$

which gives

$$\begin{aligned}\mathbf{E}(c_{t+1} - c_t \mid c_t) = \frac{1}{n}\left(\lambda(c_t) - c_t\right) \\ + \frac{\lambda(c_t)}{n}(1 - c_t)\frac{\exp\left(-\frac{\alpha_2}{2n}c_t\right) - 1}{1 + \left(\frac{p}{1-p}\right)\exp\left(-\frac{\alpha_2}{2n}c_t\right)\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}\left(c_t\right)^2\right)} \\ + \frac{\lambda(c_t)}{n}c_t\frac{\exp\left(-\frac{\alpha_1}{n} - \frac{3\alpha_2}{2n}c_t + \frac{\alpha_2}{n^2}\right) - 1}{1 + \left(\frac{p}{1-p}\right)\exp\left(-\frac{\alpha_1}{n} - \frac{3\alpha_2}{2n}c_t + \frac{\alpha_2}{n^2}\right)\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}\left(c_t\right)^2\right)}.\end{aligned} \tag{3.19}$$

Note that for the first fraction,

$$-\frac{\alpha_2}{2n}c_t \leq \exp\left(-\frac{\alpha_2}{2n}c_t\right) - 1 \leq -\frac{\alpha_2}{2n}c_t + \frac{\alpha_2^2}{8n^2}c_t^2$$

implying that

$$\frac{\exp\left(-\frac{\alpha_2}{2n}c_t\right) - 1}{1 + \left(\frac{p}{1-p}\right)\exp\left(-\frac{\alpha_2}{2n}c_t\right)\exp\left(\alpha_1 c_t + \frac{\alpha_2}{2}\left(c_t\right)^2\right)} = O\left(\frac{1}{n}\right).$$

Following a similar argument, the other fraction is $O(1/n)$ as well, and the conclusion readily follows. ■

Applying Lemma 3.1, the following Theorem 3.2 and Corollary 3.3 and Corollary 3.4 show that if the associated magnetization c_0 of the initial configuration is significantly different from an attractor c^* but bounded away from any other solution of the fixed point equation $\lambda(c) = c$, then there is a drift of the Glauber dynamics towards a configuration whose normalized magnetization is closer to c^* than the starting state, i.e. $c_t \to c^*$. Theorem 3.2 is proved when $c_0 > c^*$, and an analogous result holds for $c_0 < c^*$ using a similar line of reasoning. See Figure 3.2 for an illustration of this burn-in procedure.

Theorem 3.2. Suppose $\lambda(c^*) = c^*$ and $\lambda'(c^*) < 1$. Let $\overline{c} > c^*$ be the smallest value satisfying $\lambda(\overline{c}) = \overline{c}$. (If no such $\overline{c}$ exists, take $\overline{c} = 1$.) Let the initial magnetization be c_0 with $c^* + \mu < c_0 < \overline{c} - \mu$ for some $\mu > 0$. Then there exist $\eta, d > 0$ depending only on $\mu, p, \alpha_1, \alpha_2$ such that $T = dn$ and

$$\mathbf{P}\left(c_T \leq c_0 - \eta\right) \geq 1 - e^{-\Omega(\sqrt{n})}. \tag{3.20}$$

Proof. Since $\lambda'(c^*) < 1$, $\lambda(c) - c < 0$ for $c \in [c^* + \mu, \overline{c} - \mu]$. By the extreme value theorem, the maximum of the smooth function $\lambda(c) - c$ is attained for some value $\underline{c}$ in the compact interval $[c^* + \mu, \overline{c} - \mu]$. Define $\gamma > 0$ as $\gamma = -(\lambda(\underline{c}) - \underline{c})/2$. Choose $\eta > 0$ so that $[c_0 - 2\eta, c_0 + \eta] \subseteq [c^* + \mu, \overline{c} - \mu]$. Let

$$D_t(\eta) = \left\{c_t : c_0 - 2\eta \leq c_t \leq c_0 + \eta\right\}. \tag{3.21}$$

By Lemma 3.1, for $c_t \in D_t(\eta)$ and n sufficiently large, $\mathbf{E}(c_{t+1} - c_t \mid c_t) \leq -\gamma/n$. Utilizing the negative drift $-\gamma/n$ of the random walk c_t and employing a moment generating function method, we first show that "bad" magnetization, i.e. $c_t > c_0 + \eta$ for some $0 \leq t \leq T$, occurs with exponentially small probability.

Define S_{t_1,t_2} as

$$S_{t_1,t_2} = \sum_{t=t_1+1}^{t_2} \left(c_t - c_{t-1} + \frac{\gamma}{2n}\right) \mathbf{1}_{D_{t-1}}. \tag{3.22}$$

The random variable S_{t_1,t_2} records the change in "good" magnetization c_t from t_1 to t_2, shifted by $\gamma/(2n)$ per time step. Let $\mathcal{F}_t$ be the natural filtration. Note that $e^{\theta S_{t_1,t_2-1}}$ is constant given $\mathcal{F}_{t_2-1}$, so, by the tower property of conditional expectation,

$$\mathbf{E}\left(e^{\theta S_{t_1,t_2}}\right) = \mathbf{E}\left(e^{\theta S_{t_1,t_2-1}} \mathbf{E}\left(e^{\theta(c_{t_2} - c_{t_2-1} + \frac{\gamma}{2n})\mathbf{1}_{D_{t_2-1}}} \mid \mathcal{F}_{t_2-1}\right)\right). \tag{3.23}$$

Using linearity of conditional expectation, we write

$$\begin{aligned}\mathbf{E}\left(e^{\theta(c_{t_2}-c_{t_2-1}+\frac{\gamma}{2n})\mathbf{1}_{D_{t_2-1}}} \mid \mathcal{F}_{t_2-1}\right) &= \sum_{k=0}^{\infty}\mathbf{E}\left(\frac{\theta^k(c_{t_2}-c_{t_2-1}+\frac{\gamma}{2n})^k}{k!}\mathbf{1}^k_{D_{t_2-1}} \mid \mathcal{F}_{t_2-1}\right)\\ &= 1+\mathbf{E}\left(\theta(c_{t_2}-c_{t_2-1}+\frac{\gamma}{2n})\mathbf{1}_{D_{t_2-1}} \mid \mathcal{F}_{t_2-1}\right)\\ &\quad+\sum_{k=2}^{\infty}\mathbf{E}\left(\frac{\theta^k(c_{t_2}-c_{t_2-1}+\frac{\gamma}{2n})^k}{k!}\mathbf{1}^k_{D_{t_2-1}} \mid \mathcal{F}_{t_2-1}\right).\end{aligned} \tag{3.24}$$

Recall that $\mathbf{E}(c_{t+1}-c_t \mid c_t) \le -\gamma/n$ for $c_t \in D_t$, and so

$$\mathbf{E}\left(e^{\theta(c_{t_2}-c_{t_2-1}+\frac{\gamma}{2n})\mathbf{1}_{D_{t_2-1}}} \mid \mathcal{F}_{t_2-1}\right) \le 1-\frac{\gamma\theta}{2n}\mathbf{1}_{D_{t_2-1}}+O\left(\frac{\theta^2}{n^2}\right). \tag{3.25}$$

Taking $\theta = c\sqrt{n}$ for a sufficiently small constant c, the above conditional expectation is less than 1. Iterating this procedure gives

$$\mathbf{E}\left(e^{\theta S_{t_1,t_2}}\right) \le \mathbf{E}\left(e^{\theta S_{t_1,t_2-1}}\right) \le \cdots \le \mathbf{E}\left(e^{\theta S_{t_1,t_1}}\right) = 1. \tag{3.26}$$

By the Chernoff bound,

$$\mathbf{P}\left(S_{t_1,t_2} \ge \frac{\eta}{2}\right) \le \frac{\mathbf{E}\left(e^{\theta S_{t_1,t_2}}\right)}{e^{\theta\frac{\eta}{2}}} = e^{-\Omega(\sqrt{n})}. \tag{3.27}$$

Consider the set

$$B_{t_1,t_2}(\eta) = \left(\bigcap_{t_1\le t<t_2} D_t(\eta)\right)\bigcap\left\{c_{t_2}-c_{t_1} > \frac{\eta}{2}\right\}, \tag{3.28}$$

consisting of all "good" magnetizations at time t_1 up to time t_2, with an increase of at least $\eta/2$ from t_1 to t_2. Subject to $c_t \in D_t$ for $t_1 \leq t < t_2$ and $c_{t_2} - c_{t_1} > \eta/2$,

$$\begin{aligned} S_{t_1,t_2} &= \sum_{t=t_1+1}^{t_2} \left(c_t - c_{t-1} + \frac{\gamma}{2n} \right) \\ &= c_{t_2} - c_{t_1} + \frac{\gamma}{2n}(t_2 - t_1) > \frac{\eta}{2}, \end{aligned} \tag{3.29}$$

from which the containment $B_{t_1,t_2}(\eta) \subseteq \{S_{t_1,t_2} \geq \eta/2\}$ follows. Hence

$$\mathbf{P}\left(\bigcup_{0 \leq t_1 < t_2 \leq T} B_{t_1,t_2} \right) \leq n^2 e^{-\Omega(\sqrt{n})} = e^{-\Omega(\sqrt{n})}. \tag{3.30}$$

Take n large enough. Suppose $c_t > c_0 + \eta$ for some $0 \leq t \leq T$. Then there exists a t_1 such that $c_0 - 2\eta \leq c_s$ for all $t_1 \leq s \leq t$. Define t_2 to be the least time greater than t_1 with $c_{t_2} > c_0 + \eta$. Then $c_t \in D_t$ for all $t_1 \leq t < t_2$ and $c_{t_2} - c_{t_1} > \eta/2$. This implies that

$$\{c_t : c_t > c_0 + \eta \text{ for some } 0 \leq t \leq T\} \subseteq \bigcup_{0 \leq t_1 < t_2 \leq T} B_{t_1,t_2}, \tag{3.31}$$

and further implies that

$$\mathbf{P}\left(c_t > c_0 + \eta \text{ for some } 0 \leq t \leq T\right) \leq e^{-\Omega(\sqrt{n})}. \tag{3.32}$$

We have thus shown that the normalized magnetization c_t remains below $c_0 + \eta$ for all $0 \leq t \leq T$ with exponentially high probability, provided that c_0 is suitably bounded away from any other fixed point of λ (Note: we say an event has exponentially high probability if its complement has exponentially small probability). Next we show that c_T ends below $c_0 - \eta$ with exponentially high probability. We prove this by showing that c_t actually reaches $c_0 - 2\eta$ with exponentially high probability, and then by the preceding argument will have

exponentially small probability of increasing to $c_0 - \eta$. We have

$$\mathbf{P}\left(c_t \geq c_0 - 2\eta \text{ for all } 0 \leq t \leq T\right) \leq \mathbf{P}\left(\bigcap_{0 \leq t \leq T} D_t(\eta)\right) + \mathbf{P}\left(c_t > c_0 + \eta \text{ for some } 0 \leq t \leq T\right). \tag{3.33}$$

Subject to $c_t \in D_t(\eta)$ for $0 \leq t \leq T$ and noticing that at worst $c_0 = 1$ and $c_T = 0$,

$$\begin{aligned} S_{0,T} &= \sum_{t=1}^{T} \left(c_t - c_{t-1} + \frac{\gamma}{2n}\right) \mathbf{1}_{D_{t-1}} \\ &= c_T - c_0 + \frac{\gamma}{2n} T \geq -1 + \frac{\gamma}{2n} T. \end{aligned} \tag{3.34}$$

Using the Chernoff bound on $S_{0,T}$ and assume that $d > 2/\gamma$,

$$\begin{aligned} \mathbf{P}\left(c_t \geq c_0 - 2\eta \text{ for all } 0 \leq t \leq T\right) &\leq \mathbf{P}\left(S_{0,T} \geq -1 + \frac{\gamma}{2n} T\right) + e^{-\Omega(\sqrt{n})} \\ &\leq \frac{\mathbf{E}\left(e^{\theta S_{0,T}}\right)}{e^{\theta\left(-1 + \frac{\gamma}{2n} T\right)}} + e^{-\Omega(\sqrt{n})} \\ &= e^{-\Omega(\sqrt{n})}. \end{aligned} \tag{3.35}$$

Finally,

$$\begin{aligned} \mathbf{P}\left(c_T \geq c_0 - \eta\right) \leq \mathbf{P}\left(c_T \geq c_0 - \eta \text{ and } c_t < c_0 - 2\eta \text{ for some } 0 \leq t \leq T\right) \\ + \mathbf{P}\left(\{c_t \geq c_0 - 2\eta \text{ for all } 0 \leq t \leq T\}\right) \leq e^{-\Omega(\sqrt{n})}, \end{aligned} \tag{3.36}$$

where to bound the first probability on the right, we apply the bound on "bad" magnetization of Equation (3.32) with minor adaptation: initial magnetization c_t (for some $0 \leq t \leq T$) in place of c_0, time interval under consideration $[t, T]$ in place of $[0, T]$, and the increase in magnetization $c_0 - 2\eta \to c_0 - \eta$ in place of $c_0 \to c_0 + \eta$. ■

Repeated application of Theorem 3.2 shows that after a burn-in period on the order of $O(n)$, any suitably chosen configuration ends up close to an attractor c^* with exponentially high probability. Recall the definition of high-temperature phase and low-temperature phase from Section 3.1.4. The following corollaries are immediate.

Corollary 3.3. In the high-temperature phase, suppose that c^* is the unique solution to $\lambda(c) = c$ and $\lambda'(c^*) < 1$. For any $\varepsilon > 0$, there exists $\alpha > 0$ such that for any initial configuration with associated magnetization c_0, when $t \geq \alpha n$ we have

$$\mathbf{P}\left(c_t \geq c^* + \varepsilon\right) \leq e^{-\Omega(\sqrt{n})} \tag{3.37}$$

and

$$\mathbf{P}\left(c_t \leq c^* - \varepsilon\right) \leq e^{-\Omega(\sqrt{n})}. \tag{3.38}$$

Corollary 3.4. In the low-temperature phase, suppose that c^* is a solution to $\lambda(c) = c$ and $\lambda'(c^*) < 1$. Take $\varepsilon > 0$. If the associated magnetization c_0 for some initial configuration satisfies $c^* - \varepsilon \leq c_0 \leq c^* + \varepsilon$, then there exists $\beta > 0$,

$$\mathbf{P}\left(\sup_{0<t<e^{\beta\sqrt{n}}} c_t \geq c^* + 2\varepsilon\right) \leq e^{-\Omega(\sqrt{n})} \tag{3.39}$$

and

$$\mathbf{P}\left(\inf_{0<t<e^{\beta\sqrt{n}}} c_t \leq c^* - 2\varepsilon\right) \leq e^{-\Omega(\sqrt{n})}. \tag{3.40}$$

3.3 Fast mixing at high-temperature

In this section we study the mixing time of Glauber dynamics in the high-temperature phase. We first establish an upper bound $O(n \log n)$ using path coupling techniques of Bubley and Dyer [19]. Consider two arbitrary spin configurations $X, Y \in \mathcal{X}$. Taking "attractive" parameters $\alpha_i \geq 0$ ensures that we may apply a monotone coupling on the

chain (X_t, Y_t): X_t is a version of the Glauber dynamics with starting state X and Y_t is a version of the Glauber dynamics with starting state Y, if $X_0 \leq Y_0$ then $X_t \leq Y_t$ for all t. We write $\mathbf{P}_{X,Y}$ and $\mathbf{E}_{X,Y}$ for the underlying probability measure and associated expectation. To keep the notation light, we omit the explicit dependence on X and Y when it is clear from the context. To understand how far apart X_t and Y_t are, we introduce the Hamming distance, which records the number of vertices where the two configurations disagree. Define $\rho : \mathcal{X} \times \mathcal{X} \to \{0, \ldots, n\}$ by

$$\rho(X, Y) = \sum_{i=1}^{n} |X(i) - Y(i)| . \tag{3.41}$$

Following a standard contraction argument, it suffices to estimate the average distance after one Glauber update between two coupled configurations X and Y with Hamming distance $\rho(X, Y) = 1$.

Lemma 3.5. Assume that $\sup_{0 \leq c \leq 1} \lambda'(c) < 1$. Let $X, Y \in \mathcal{X}$ be two spin configurations satisfying $X \leq Y$ and $\rho(X, Y) = 1$. Set the initial state $X_0 = X$ and $Y_0 = Y$. Then there exists $\delta > 0$ depending only on p, α_1, α_2 such that a single step of the Glauber dynamics can be coupled in such a way that when n is sufficiently large:

$$\mathbf{E}\left(\rho(X_1, Y_1)\right) \leq e^{-\delta/n}. \tag{3.42}$$

Proof. Let $X, Y \in \mathcal{X}$ be two configurations such that $X \leq Y$ and there exists a single vertex i such that $X(i) = 0$ and $Y(i) = 1$. Let U be a uniform random variable on $[0, 1]$. We apply the standard monotone coupling, where U is used as the common source of noise to update both chains so that they agree as often as possible. From the mechanism described in Section 3.1.1, the chain evolves by selecting a vertex j uniformly at random and updating

the spin at j. Set

$$X_1(j) = \begin{cases} 1 & U \leq p_+(X,j), \\ 0 & U > p_+(X,j), \end{cases} \qquad Y_1(j) = \begin{cases} 1 & U \leq p_+(Y,j), \\ 0 & U > p_+(Y,j), \end{cases} \tag{3.43}$$

and $X_1(k) = X(k)$ and $Y_1(k) = Y(k)$ for all $k \neq j$. Define the function $f(S)$ as

$$f(S) = \frac{p \exp\left(\frac{\alpha_1}{n} S + \frac{\alpha_2}{2n^2} S(S-1)\right)}{p \exp\left(\frac{\alpha_1}{n} S + \frac{\alpha_2}{2n^2} S(S-1)\right) + (1-p)}. \tag{3.44}$$

If $j = i$, then $p_+(X,j) = f\left(S(X,j)\right) = p_+(Y,j)$ and so $\rho(X_1, Y_1) = 0$. For $j \neq i$, we have $p_+(X,j) = f\left(S(X,j)\right)$ while $p_+(Y,j) = f\left(S(X,j)+1\right)$, where $0 \leq S(X,j) \leq n-2$. Since f is a smooth and increasing function, this shows that $p_+(X,j) \leq p_+(Y,j)$. Hence $\rho(X_1, Y_1) = 2$ if $p_+(X,j) < U \leq p_+(Y,j)$ and $\rho(X_1, Y_1) = 1$ otherwise.

We wish to find an upper bound for

$$\mathbf{E}\left(\rho(X_1, Y_1)\right) = 1 - \frac{1}{n} + \frac{1}{n} \sum_{j \neq i} \left(p_+(Y,j) - p_+(X,j)\right). \tag{3.45}$$

To that end, we compute, by the mean value theorem

$$p_+(Y,j) - p_+(X,j) = f\left(S(X,j)+1\right) - f\left(S(X,j)\right) = \frac{g'(\overline{c})}{n}, \tag{3.46}$$

where $0 \leq \overline{c} \leq 1$ and $g'(c)$ is defined as

$$g'(c) = \frac{p(1-p) \exp\left(\alpha_1 c + \frac{\alpha_2}{2} c\left(c - \frac{1}{n}\right)\right)}{p \exp\left(\alpha_1 c + \frac{\alpha_2}{2} c\left(c - \frac{1}{n}\right)\right) + (1-p)} \left(\alpha_1 + \alpha_2 c - \frac{\alpha_2}{2n}\right). \tag{3.47}$$

Compare $g'(c)$ against $\lambda'(c)$, where $\lambda(c)$ is defined as in Equation (3.14),

$$\lambda'(c) = \frac{p(1-p) \exp\left(\alpha_1 c + \frac{\alpha_2}{2} c^2\right)}{p \exp\left(\alpha_1 c + \frac{\alpha_2}{2} c^2\right) + (1-p)} \left(\alpha_1 + \alpha_2 c\right). \tag{3.48}$$

Via standard analytical arguments as in the proof of Lemma 3.1, the difference $g'(c) - \lambda'(c) = O(1/n)$. Therefore

$$\begin{aligned}
\mathbf{E}\left(\rho(X_1, Y_1)\right) &\leq 1 - \frac{1}{n} + \frac{1}{n} \sum_{j \neq i} \left(\frac{1}{n} \sup_{0 \leq c \leq 1} \lambda'(c) + O\left(\frac{1}{n^2}\right) \right) \\
&\leq 1 - \frac{1}{n} + \frac{n-1}{n^2} \sup_{0 \leq c \leq 1} \lambda'(c) + O\left(\frac{1}{n^2}\right) \\
&\leq 1 - \frac{1 - \sup_{0 \leq c \leq 1} \lambda'(c)}{n} + O\left(\frac{1}{n^2}\right). \qquad (3.49)
\end{aligned}$$

Define $\delta > 0$ as $\delta = \left(1 - \sup_{0 \leq c \leq 1} \lambda'(c)\right)/2$. Then for n sufficiently large, we obtain

$$\mathbf{E}\left(\rho(X_1, Y_1)\right) \leq 1 - \frac{\delta}{n} \leq e^{-\delta/n}. \qquad (3.50)$$

■

The requirement $\sup_{0 \leq c \leq 1} \lambda'(c) < 1$ in Lemma 3.5 may be weakened. By Corollary 3.3, in the high-temperature phase, with exponentially high probability, after $O(n)$ time steps the associated magnetization of all configurations are within an ε-neighborhood of the unique solution c^* of λ with $\lambda'(c^*) < 1$. The supremum referenced in Equation (3.49) thus need not be taken over the entire interval $[0, 1]$ but just $[c^* - \varepsilon, c^* + \varepsilon]$, and is guaranteed to be less than 1 using smoothness of λ. Now take any two configurations $X, Y \in \mathcal{X}$ with $\rho(X, Y) = k$, where $1 \leq k \leq n$. (At worst $X(i) = 0$ and $Y(i) = 1$ for all $i \in \{1, \ldots, n\}$.) There is a sequence of states $X_0, \ldots, X_k$ such that $X_0 = X$, $X_k = Y$, and each neighboring pair X_i, X_{i+1} are unit Hamming distance apart. Applying Lemma 3.5 for configurations at unit distance, we have $\mathbf{E}\left(\rho(X_1, Y_1)\right) \leq ne^{-\delta/n}$. Iterating gives

$$\mathbf{E}\left(\rho(X_t, Y_t)\right) \leq ne^{-\delta t/n}. \qquad (3.51)$$

Theorem 3.6. In the high-temperature phase, the mixing time for the Glauber dynamics is $O(n \log n)$.

Proof. By Theorem 14.6 and Corollary 14.7 of [45], Equation (3.51) implies

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{n\left(\log n - \log \varepsilon\right)}{\delta} \right\rceil. \tag{3.52}$$

Setting $\varepsilon = 1/4$,

$$t_{\text{mix}} \leq \left\lceil \frac{n\left(\log n + \log 4\right)}{\delta} \right\rceil. \tag{3.53}$$

■

Next in Theorem 3.7, by checking the total variance distance from the stationary distribution at time $t^* := (n \log n)/4$, we establish a matching lower bound $\Omega(n \log n)$ for the Glauber dynamics in the high-temperature phase. Together with Theorem 3.6, the correct order for the mixing time, $\Theta(n \log n)$, is validated. To this end, let γ be the spectral gap associated with the Glauber dynamics transition matrix. By Lemma 13.12 and Remark 13.13 of [45],

$$\gamma = \inf_{\substack{f:\mathcal{X}\to\mathbb{R} \\ \mathbf{Var}_\pi(f)\neq 0}} \frac{\varepsilon(f)}{\mathbf{Var}_\pi(f)}, \tag{3.54}$$

where the Dirichlet form

$$\varepsilon(f) = \frac{1}{2} \sum_{X,Y\in\mathcal{X}} \left(f(X) - f(Y)\right)^2 \pi(X) P(X,Y), \tag{3.55}$$

and the variance under stationary distribution

$$\mathbf{Var}_\pi(f) = \sum_{X\in\mathcal{X}} \left(f(X)\right)^2 \pi(X) - \left(\sum_{X\in\mathcal{X}} f(X)\pi(X)\right)^2. \tag{3.56}$$

Theorem 3.7. In the high-temperature phase, the mixing time for the Glauber dynamics is $\Omega(n \log n)$.

Proof. Let $f(X)$ count the number of vertices with spin 1 in a configuration $X \in \mathcal{X}$. Then, for this particular f, we have that

$$\gamma \leq \frac{\varepsilon(f)}{\mathbf{Var}_\pi(f)}, \tag{3.57}$$

From the mechanism described in Section 3.1.1, for configurations $X, Y \in \mathcal{X}$, $P(X, Y)$ defined in Equation (3.8) is zero unless X and Y differ at at most one vertex. This implies that

$$\varepsilon(f) \leq \frac{1}{2} \sum_{X,Y \in \mathcal{X}} \pi(X) P(X, Y) = \frac{1}{2} \sum_{X \in \mathcal{X}} \pi(X) = \frac{1}{2}, \tag{3.58}$$

and when applied to Equation (3.57), further implies that $2\,\mathbf{Var}_\pi(f) \leq 1/\gamma$. Hence

$$\log 2\,(2\,\mathbf{Var}_\pi(f) - 1) \leq \log 2 \left(\frac{1}{\gamma} - 1 \right) \leq t_{\text{mix}}, \tag{3.59}$$

where the second inequality uses spectral representation techniques (for details, see for example Theorem 12.4 of [45]). By Theorem 3.6, $t_{\text{mix}} = O(n \log n)$, which then gives $\mathbf{Var}_\pi(f) = O(n \log n)$. Let configuration X be chosen according to the stationary distribution π. By Chebyshev's inequality,

$$\pi\left(|f(X) - \mathbf{E}_\pi(f(X))| > n^{2/3} \right) \leq \frac{\mathbf{Var}_\pi(f(X))}{n^{4/3}} = O(n^{-1/3} \log n). \tag{3.60}$$

Therefore asymptotically

$$\pi\left(|f(X) - \mathbf{E}_\pi(f(X))| \leq n^{2/3} \right) \to 1. \tag{3.61}$$

Let $X^+, X^- \in \mathcal{X}$ be configurations such that $X^+(i) = 1$ and $X^-(i) = 0$ for every $i \in \{1, \ldots, n\}$. Assume that X^+ and X^- are coupled using the standard monotone coupling as described in the proof of Lemma 3.5. So $X_t^+ \geq X_t^-$ for all t. Let R_{t^*} denote the number of vertices not yet selected by the Glauber Dynamics by time t^*. By a coupon collecting argument,

$$\mathbf{E}(R_{t^*}) \asymp n^{3/4} \text{ and } \mathbf{Var}(R_{t^*}) \leq \mathbf{E}(R_{t^*}). \tag{3.62}$$

(For details, see for example Lemma 7.12 of [45].) Let $\varepsilon > 0$. Again, by Chebyshev's inequality, it follows that, asymptotically,

$$\mathbf{P}\left(|R_{t^*} - \mathbf{E}\left(R_{t^*}\right)| > (1-\varepsilon)\mathbf{E}\left(R_{t^*}\right)\right) \leq \frac{\mathbf{Var}\left(R_{t^*}\right)}{(1-\varepsilon)^2\left(\mathbf{E}\left(R_{t^*}\right)\right)^2} \leq O(n^{-3/4}), \tag{3.63}$$

which, using set containment, implies that

$$\mathbf{P}\left(R_{t^*} < \varepsilon\mathbf{E}\left(R_{t^*}\right)\right) \leq \mathbf{P}\left(|R_{t^*} - \mathbf{E}\left(R_{t^*}\right)| > (1-\varepsilon)\mathbf{E}\left(R_{t^*}\right)\right) \leq O(n^{-3/4}). \tag{3.64}$$

Therefore $R_{t^*} = \Omega(n^{3/4})$ with probability tending to 1 asymptotically. Since $f(X_{t^*}^+) - f(X_{t^*}^-) \geq R_{t^*}$, we also conclude that $f(X_{t^*}^+) - f(X_{t^*}^-) = \Omega(n^{3/4})$ with probability tending to 1 asymptotically.

Define sets A and B respectively as

$$A := \left\{X \in \mathcal{X} \ : \ \left|f(X_{t^*}^+) - \mathbf{E}_\pi(f(X))\right| \leq n^{2/3}\right\} \tag{3.65}$$

and

$$B := \left\{X \in \mathcal{X} \ : \ \left|f(X_{t^*}^-) - \mathbf{E}_\pi(f(X))\right| \leq n^{2/3}\right\}. \tag{3.66}$$

Then by the triangle inequality, their intersection $A \cap B$, if nonempty, satisfies $f(X_{t^*}^+) - f(X_{t^*}^-) = O(n^{2/3})$. Since $n^{3/4} > n^{2/3}$, this contradicts what was established in the previous

paragraph. Therefore the sets A and B are asymptotically disjoint and so one of them has probability bounded above by $1/2 - o(1)$. Without loss of generality, suppose A satisfies

$$\mathbf{P}\left(\left|f(X^+_{t^*}) - \mathbf{E}_\pi(f(X))\right| \leq n^{2/3}\right) \leq \frac{1}{2} - o(1). \tag{3.67}$$

By definition of the total variation distance and the bounds in Equation (3.61) and Equation (3.67),

$$\left\|P^{t^*}(X^+, \cdot) - \pi\right\|_{\text{TV}} \geq \left|P^{t^*}(X^+, A) - \pi(A)\right| \geq 1 - \frac{1}{2} + o(1). \tag{3.68}$$

Since at t_{mix}, the default for $t_{\text{mix}}(\varepsilon)$, the distance must be less than or equal to $1/4$, this shows that the mixing time is asymptotically bigger than $t^* = (n \log n)/4$ proving the lower bound. ■

3.4 Slow mixing at low-temperature

In this section we study the mixing time of Glauber dynamics in the low-temperature phase. Rather than analyzing the spin update probability λ defined in Equation (3.14) directly, we find asymptotic expressions for components of the partition function Z (see Definition 3.1). For $k \in \{0, \ldots, n\}$, define $A_k = \{X : |\{i : X(i) = 1\}| = k\}$. That is, the set A_k consists of spin configurations $X \in \mathcal{X}$ whose number of vertices with spin 1 is k and number of vertices with spin 0 is $n - k$. Then

$$\pi(A_k) = \frac{1}{Z}\binom{n}{k} \exp\left(\frac{\alpha_1}{n}\binom{k}{2} + \frac{\alpha_2}{n^2}\binom{k}{3}\right) p^k (1-p)^{n-k}. \tag{3.69}$$

Let a_k be such that $a_k = Z\pi(A_k)$. Notice that $\sum_{k=0}^{n} a_k = Z$.

Lemma 3.8. Let $c \in [0, 1]$ and a_k be defined as above, we have

$$\log(a_{\lfloor cn \rfloor}) = n(\varphi_{p,\alpha_1,\alpha_2}(c) + o(1)), \tag{3.70}$$

where

$$\varphi(c) = \frac{\alpha_1}{2}c^2 + \frac{\alpha_2}{6}c^3 - c\log\frac{c}{p} - (1-c)\log\frac{1-c}{1-p}. \tag{3.71}$$

Proof. Stirling's formula states that

$$n! \asymp \sqrt{2\pi}e^{-n}n^n\sqrt{n}. \tag{3.72}$$

The binomial coefficient admits the following asymptotic formula

$$\binom{n}{\lfloor cn \rfloor} \asymp \frac{1}{\sqrt{2\pi}c^{cn}\sqrt{c}(1-c)^{(1-c)n}\sqrt{1-c}\sqrt{n}}. \tag{3.73}$$

Hence

$$\begin{aligned}\log(a_{\lfloor cn \rfloor}) &\asymp n\left(\frac{\alpha_1}{2}c^2 + \frac{\alpha_2}{6}c^3 - c\log\frac{c}{p} - (1-c)\log\frac{1-c}{1-p}\right) \\ &\quad - \log\left(\sqrt{2\pi c(1-c)n}\right) \\ &= n(\varphi(c) + o(1)).\end{aligned} \tag{3.74}$$

■

Next in Lemmas 3.9 and 3.10, we reveal a deep relationship between φ (commonly referred to as the "free energy density") and the spin update probability λ. As we will see, local maximizers for φ correspond to fixed points of λ, and concavity of φ at the local maximizer (indicating whether it is a local maximum or minimum) translates to the attractor/repellor characterization on the fixed point of λ previously described in Theorem 2.7.

Lemma 3.9. Let λ and φ be defined as in Equation (3.14) and Equation (3.71). Then $\lambda(c) = c \iff \varphi'(c) = 0$.

Proof. We have the following string of equivalences

$$\begin{aligned}\lambda(c) = c &\iff \left(\frac{p}{1-p}\right)\exp\left(\alpha_1 c + \frac{\alpha_2}{2}c^2\right) = \frac{c}{1-c} \\ &\iff \alpha_1 c + \frac{\alpha_2}{2}c^2 + \log\frac{p}{1-p} - \log\frac{c}{1-c} = 0 \\ &\iff \varphi'(c) = 0. \end{aligned} \tag{3.75}$$

■

Lemma 3.10. Let λ and φ be defined as in Equation (3.14) and Equation (3.71). Suppose $\varphi'(c) = 0$. Then $\varphi''(c) > 0 \iff \lambda'(c) > 1$, $\varphi''(c) < 0 \iff \lambda'(c) < 1$, and $\varphi''(c) = 0 \iff \lambda'(c) = 1$.

Proof. Suppose $\varphi'(c) = 0$. From Lemma 3.9, $\lambda(c) = c$. Therefore

$$\lambda'(c) = c(1-c)\left(\alpha_1 + \alpha_2 c\right). \tag{3.76}$$

The claim readily follows since

$$\varphi''(c) = \alpha_1 + \alpha_2 c - \frac{1}{c(1-c)}. \tag{3.77}$$

■

Using a conductance argument whose idea goes back at least to Griffiths et al. [37], we now show that in the region where $\varphi(c)$ has at least two local maximizers, the mixing time for the Glauber dynamics is at least exponential. In the language of λ, this establishes exponentially slow mixing of the Glauber dynamics when $\lambda(c) = c$ has at least two solutions c satisfying $\lambda'(c) < 1$. Recall the *Cheeger constant* or *bottleneck ratio* of a Markov

chain with stationary distribution π as

$$\Phi_* = \min_{S \subseteq \mathcal{X}:\, \pi(S) \leq 1/2} \frac{Q(S, S^c)}{\pi(S)}, \tag{3.78}$$

where S is a set in the configuration space and Q is the edge measure given by

$$Q(X, Y) = \pi(X)P(X, Y) \text{ and } Q(A, B) = \sum_{X \in A, Y \in B} Q(X, Y). \tag{3.79}$$

Recall that $P(X, Y)$ defined in Equation (3.8) is the transition probability from configuration X to configuration Y, and so $Q(A, B)$ is the probability of moving from set A to set B in one step of the chain when starting from the stationary distribution.

Theorem 3.11. In the low-temperature phase, the mixing time for the Glauber dynamics is $e^{\Omega(n)}$.

Proof. Notice that $\varphi'(c) \to \infty$ as $c \to 0$ and $\varphi'(c) \to -\infty$ as $c \to 1$, so the local maximizers of φ are contained in $(0, 1)$. Let c_1 be the smallest and c_2 be the largest local maximizer of φ where $c_1 \neq c_2$. There exists $\varepsilon > 0$ such that for all $c < c_1$ and $c_1 < c \leq c_1 + \varepsilon$, $\varphi(c) < \varphi(c_1)$, while for all $c_2 - \varepsilon \leq c < c_2$ and $c > c_2$, $\varphi(c) < \varphi(c_2)$, with $c_1 + \varepsilon < c_2 - \varepsilon$. Define the following two sets

$$S_1 = \{A_0, \ldots, A_{\lfloor (c_1+\varepsilon)n \rfloor}\} \tag{3.80}$$

and

$$S_2 = \{A_{\lfloor (c_2-\varepsilon)n \rfloor}, \ldots, A_n\}. \tag{3.81}$$

For n large enough, $S_1 \cap S_2 = \varnothing$, $A_{\lfloor (c_1+\varepsilon)n \rfloor} \neq A_{\lfloor c_1 n \rfloor}$, and $A_{\lfloor (c_2-\varepsilon)n \rfloor} \neq A_{\lfloor c_2 n \rfloor}$. Since S_1 and S_2 are disjoint, at least one of them has probability bounded above by $1/2$. Without

loss of generality, suppose S_1 is such that $\pi(S_1) \leq \frac{1}{2}$. Then

$$\Phi_* \leq \frac{Q(S_1, S_1^c)}{\pi(S_1)} = \frac{1}{\pi(S_1)} \sum_{i=0}^{\lfloor (c_1+\varepsilon)n \rfloor} \sum_{X \in A_i} \pi(X) \sum_{j=\lfloor (c_1+\varepsilon)n \rfloor + 1}^{n} \sum_{Y \in A_j} P(X,Y). \tag{3.82}$$

Since a single step of the Glauber dynamics only changes the value of the normalized magnetization by at most $1/n$, the only non-zero transition probability in Equation (3.82) is the transition from $A_{\lfloor (c_1+\varepsilon)n \rfloor}$ to $A_{\lfloor (c_1+\varepsilon)n \rfloor + 1}$. It follows that

$$\begin{aligned} \Phi_* &\leq \frac{1}{\pi(S_1)} \sum_{i=0}^{\lfloor (c_1+\varepsilon)n \rfloor} \sum_{X \in A_i} \pi(X) \leq \frac{\pi(A_{\lfloor (c_1+\varepsilon)n \rfloor})}{\pi(A_{\lfloor c_1 n \rfloor})} \\ &= \frac{a_{\lfloor (c_1+\varepsilon)n \rfloor}}{a_{\lfloor c_1 n \rfloor}} \asymp \frac{e^{n\varphi(c_1+\varepsilon)}}{e^{n\varphi(c_1)}} = e^{n(\varphi(c_1+\varepsilon)-\varphi(c_1))}, \end{aligned} \tag{3.83}$$

where the asymptotics are derived in Lemma 3.8. Define $\delta > 0$ as

$$\delta = \left(\varphi(c_1) - \varphi(c_1+\varepsilon)\right)/2.$$

Then the bottleneck ratio satisfies $\Phi(S_1) \leq e^{-\delta n}$. Using Theorem 7.3 of [45],

$$t_{\text{mix}} \geq \frac{1}{4\Phi_*} \geq \frac{e^{\delta n}}{4} = e^{\Omega(n)}. \tag{3.84}$$

■

Call a Markov chain local if at most $o(n)$ vertices are selected in each step. The argument used in the proof of Theorem 3.11 actually shows that in the low-temperature phase, the mixing is exponentially slow for any local Markov chain, with the (single-site) Glauber dynamics being one such instance. We remark that there is a difference in the qualitative nature of the phase transition investigated in this section as compared with that in the standard statistical physics literature. While the asymptotic phase transitions in the rigorous

statistical physics sense occur at parameter values giving non-unique global maximizers of the free energy density, the asymptotic transition from high-temperature phase to low-temperature phase arises as a consequence of the non-uniqueness of local maximizers for the free energy density. This discrepancy may not come as a surprise, since in simulations it is often hard to distinguish between a local maximizer and a global maximizer and the algorithm may become trapped at a local maximizer; one solution might be to add controlled moves based on network geometry.

3.5 Slower burn-in along critical curve

In this section we study the mixing time of the Glauber dynamics along the critical curve, corresponding to parameters for which $\lambda(c) = c$ admits a unique solution c with $\lambda'(c) = 1$. We first identify explicitly the high-temperature vs. low-temperature phase. As explained in Section 3.4, the two phases may be alternatively determined by whether there is a unique local maximizer for φ defined in Equation (3.71). The phase identification thus reduces to a 3-dimensional intricate calculus problem. Though it sounds straightforward, various tricks are needed to solve it analytically. The crucial idea is to minimize the effect of the parameters p, α_1, α_2 on the free energy density φ one by one. (See [72] for more details of the calculation in a related model.) Denote by $l(c) := \varphi(c) - \log(1-p)$.

Proposition 3.12. Fix α_2. Consider the maximization problem for

$$l_{\alpha_2}(c; p, \alpha_1) = \left(\log \frac{p}{1-p}\right) c + \frac{\alpha_1}{2} c^2 + \frac{\alpha_2}{6} c^3 - c \log c - (1-c)\log(1-c) \tag{3.85}$$

on the interval $[0, 1]$, where $0 < p < 1$ and $-\infty < \alpha_1 < \infty$ are parameters. Then there is a V-shaped region in the (p, α_1)-plane with corner point (p^c, α_1^c),

$$p^c = \frac{\bar{c} \exp\left(\frac{4\bar{c}-3}{2(1-\bar{c})^2}\right)}{\bar{c} \exp\left(\frac{4\bar{c}-3}{2(1-\bar{c})^2}\right) + (1-\bar{c})}, \tag{3.86}$$

$$\alpha_1^c = \frac{2 - 3\bar{c}}{\bar{c}(1-\bar{c})^2}, \tag{3.87}$$

where $\bar{c}$ is uniquely determined by

$$\alpha_2 = \frac{2\bar{c} - 1}{\bar{c}^2(1-\bar{c})^2}. \tag{3.88}$$

Outside this region, $l_{\alpha_2}(c)$ has only one local maximizer c^*. Inside this region, $l_{\alpha_2}(c)$ has exactly two local maximizers c_1^* and c_2^*.

Proof. The location of maximizers of $l_{\alpha_2}(c)$ on the interval $[0, 1]$ is closely related to the properties of its derivatives:

$$l'_{\alpha_2}(c) = \log\frac{p}{1-p} + \alpha_1 c + \frac{\alpha_2}{2}c^2 - \log\frac{c}{1-c},$$

$$l''_{\alpha_2}(c) = \alpha_1 + \alpha_2 c - \frac{1}{c(1-c)},$$

$$l'''_{\alpha_2}(c) = \alpha_2 + \frac{1-2c}{c^2(1-c)^2}. \tag{3.89}$$

We check that $l'''_{\alpha_2}(c)$ is monotonically decreasing on $[0, 1]$, $l'''_{\alpha_2}(0) = \infty$, and $l'''_{\alpha_2}(1) = -\infty$. Thus there is a unique $\bar{c}$ in $(0, 1)$ such that $l'''_{\alpha_2}(\bar{c}) = 0$, with $l'''_{\alpha_2}(c) > 0$ for $c < \bar{c}$ and $l'''_{\alpha_2}(c) < 0$ for $c > \bar{c}$. Since the correspondence between α_2 and $\bar{c}$ is one-to-one, we may describe α_2 by Equation (3.88).

This implies that $l''_{\alpha_2}(c)$ is increasing from 0 to $\bar{c}$, and decreasing from $\bar{c}$ to 1, with the global maximum achieved at $\bar{c}$, where

$$l''_{\alpha_2}(\bar{c}) = \alpha_1 + \frac{3\bar{c} - 2}{\bar{c}(1-\bar{c})^2}. \tag{3.90}$$

Let α_1^c be defined as in Equation (3.87) so that $l''_{\alpha_2}(\bar{c}; \alpha_1^c) = 0$. It follows that for $\alpha_1 \leq \alpha_1^c$, $l''_{\alpha_2}(c) \leq 0$ on the entire interval $[0, 1]$; whereas for $\alpha_1 > \alpha_1^c, l''_{\alpha_2}(c)$ takes on both positive and negative values, and we denote the transition points by c_1 and c_2 ($c_1 < \bar{c} < c_2$). For fixed α_2, c_1 and c_2 are solely determined by α_1, and vice versa. Let $m(c) = \alpha_1 - l''_{\alpha_2}(c)$ so that $\alpha_1 = m(c_1) = m(c_2)$. We have $m(0) = m(1) = \infty, m(c)$ is decreasing from 0 to $\bar{c}$, and increasing from $\bar{c}$ to 1.

We proceed to analyze properties of $l'_{\alpha_2}(c)$ and $l_{\alpha_2}(c)$ on the interval $[0, 1]$. For $\alpha_1 \leq \alpha_1^c, l'_{\alpha_2}(c)$ is monotonically decreasing. For $\alpha_1 > \alpha_1^c, l'_{\alpha_2}(c)$ is decreasing from 0 to c_1, increasing from c_1 to c_2, then decreasing again from c_2 to 1. We write down the explicit expressions of $l'_{\alpha_2}(c_1)$ and $l'_{\alpha_2}(c_2)$:

$$l'_{\alpha_2}(c_1) = \log \frac{p}{1-p} + \frac{1}{1-c_1} - \log \frac{c_1}{1-c_1} + \frac{1-2\bar{c}}{2\bar{c}^2(1-\bar{c})^2} c_1^2,$$

$$l'_{\alpha_2}(c_2) = \log \frac{p}{1-p} + \frac{1}{1-c_2} - \log \frac{c_2}{1-c_2} + \frac{1-2\bar{c}}{2\bar{c}^2(1-\bar{c})^2} c_2^2. \tag{3.91}$$

Notice that $l_{\alpha_2}(c)$ is a bounded continuous function, $l'_{\alpha_2}(0) = \infty$, and $l'_{\alpha_2}(1) = -\infty$, so $l_{\alpha_2}(c)$ cannot be maximized at 0 or 1. For $\alpha_1 \leq \alpha_1^c, l'_{\alpha_2}(c)$ crosses the c-axis only once, going from positive to negative. Thus $l_{\alpha_2}(c)$ has a unique local maximizer $c^* = \bar{c}$. For $\alpha_1 > \alpha_1^c$, the situation is more complicated. If $l'_{\alpha_2}(c_1) \geq 0$ (resp. $l'_{\alpha_2}(c_2) \leq 0$), $l_{\alpha_2}(c)$ has a unique local maximizer at a point $c^* > c_2$ (resp. $c^* < c_1$). If $l'_{\alpha_2}(c_1) < 0 < l'_{\alpha_1}(c_2)$, then $l_{\alpha_2}(c)$ has two local maximizers c_1^* and c_2^*, with $c_1^* < c_1 < \bar{c} < c_2 < c_2^*$.

Let

$$n(c) = \frac{1}{1-c} - \log \frac{c}{1-c} + \frac{1-2\bar{c}}{2\bar{c}^2(1-\bar{c})^2} c^2 \tag{3.92}$$

so that $l'_{\alpha_2}(c_1) = \log(p/(1-p)) + n(c_1)$ and $l'_{\alpha_2}(c_2) = \log(p/(1-p)) + n(c_2)$. We have $n(0) = \infty$, $n(1) = \infty$, with derivative $n'(c)$ given by

$$n'(c) = c\left(\frac{1-2\bar{c}}{\bar{c}^2(1-\bar{c})^2} - \frac{1-2c}{c^2(1-c)^2}\right) = c\left(l'''_{\alpha_2}(\bar{c}) - l'''_{\alpha_2}(c)\right). \tag{3.93}$$

As $l'''_{\alpha_2}(c)$ is monotonically decreasing on $[0,1]$, $n(c)$ is decreasing from 0 to $\bar{c}$, and increasing from $\bar{c}$ to 1, with the global minimum achieved at $\bar{c}$,

$$n(\bar{c}) = \frac{1}{1-\bar{c}} - \log\frac{\bar{c}}{1-\bar{c}} + \frac{1-2\bar{c}}{2(1-\bar{c})^2}. \tag{3.94}$$

This implies that $l'_{\alpha_2}(c_1; p, \alpha_1^c) \geq 0$ for $p \geq p^c$ where p^c is defined in Equation (3.86). The only possible region in the (p, α_1)-plane where $l'_{\alpha_2}(c_1) < 0 < l'_{\alpha_1}(c_2)$ is thus bounded by $p < p^c$ and $\alpha_1 > \alpha_1^c$.

Finally, we analyze the behavior of $l'_{\alpha_1}(c_1)$ and $l'_{\alpha_1}(c_2)$ more closely when p and α_1 are chosen from this region. Recall that $c_1 < \bar{c} < c_2$. By monotonicity of $n(c)$ on the intervals $(0, \bar{c})$ and $(\bar{c}, 1)$, there exist continuous functions $a(p)$ and $b(p)$, such that $l'_{\alpha_2}(c_1) < 0$ for $c_1 > a(p)$ and $l'_{\alpha_2}(c_2) > 0$ for $c_2 > b(p)$. As $p \to 0$, $a(p) \to 0$ and $b(p) \to 1$. $a(p)$ is an increasing function, whereas $b(p)$ is a decreasing function, and they satisfy $n(a(p)) = n(b(p)) = -p$. The restrictions on c_1 and c_2 yield restrictions on α_1, and we have $l'_{\alpha_2}(c_1) < 0$ for $\alpha_1 < m(a(p))$ and $l'_{\alpha_2}(c_2) > 0$ for $\alpha_1 > m(b(p))$. As $p \to 0, m(a(p)) \to \infty$ and $m(b(p)) \to \infty$. $m(a(p))$ and $m(b(p))$ are both decreasing functions of p and they satisfy $l'_{\alpha_2}(c_1; p, m(a(p)) = l'_{\alpha_2}(c_2; p, m(b(p)) = 0$. As $l'_{\alpha_2}(c_2; p, \alpha_1) > l'_{\alpha_2}(c_1; p, \alpha_1)$ for every (p, α_1), the curve $m(b(p))$ must lie below the curve $m(a(p))$, and together they generate the bounding curves for the V-shaped region in the (p, α_1)-plane with corner point (p^c, α_1^c) where two local maximizers exist for $l_{\alpha_2}(c)$. ■

From Proposition 3.12, the critical curve is traced out by $(p^c, \alpha_1^c, \alpha_2)$ defined in Equations (3.86) to (3.88), where we take $1/2 \leq \bar{c} = c^* \leq 2/3$ to meet the non-negativity constraints on α_1^c and α_2. See Figure 3.3. We delve deeper into the behavior of the function λ defined in Equation (3.14) along the critical curve. To lighten the notation, we denote the associated parameters by (p, α_1, α_2), and the unique local maximizer by c^* with $1/2 \leq c^* \leq 2/3$.

Lemma 3.13. Along the critical curve, we have

(1) $\lambda(c^*) = c^*$, $\lambda'(c^*) = 1$, $\lambda''(c^*) = 0$, and $\lambda'''(c^*) \leq -8$.

(2) $\lambda''(c) \leq 0$ for $c \geq c^*$, and $\lambda''(c) \geq 0$ for $c \leq c^*$.

Proof. The first claim follows from direct computation. We spell out some details. It is clear that $\lambda(c^*) = c^*$. Letting $A = p\exp(\alpha_1 c + \frac{\alpha_2}{2}c^2)/(1-p)$, we write λ's first few derivatives as

$$\lambda'(c) = \frac{\alpha_1 + \alpha_2 c}{1+A}\lambda(c),$$

$$\lambda''(c) = \lambda(c)\frac{\alpha_2}{1+A} + (1-A)\frac{\lambda'(c)^2}{\lambda(c)},$$

$$\begin{aligned}\lambda'''(c) = \lambda'(c)\frac{\alpha_2}{1+A} - \alpha_2\lambda(c)\lambda'(c) + 2(1-A)\frac{\lambda'(c)\lambda''(c)}{\lambda(c)} \\ - (1-A)\frac{\lambda'(c)^3}{\lambda(c)^2} - (\alpha_1 + \alpha_2 c)A\frac{\lambda'(c)^2}{\lambda(c)}.\end{aligned} \tag{3.95}$$

Substituting the parameter values in Equations (3.86) to (3.88) yields $\lambda'(c^*) = 1$, $\lambda''(c^*) = 0$, and

$$\lambda'''(c^*) = \frac{-6(c^*)^2 + 6(c^*) - 2}{(c^*)^2(1-c^*)^2} \leq -8. \tag{3.96}$$

For the second claim, we show that for $c \geq c^*$, $\lambda''(c) \leq 0$. The parallel claim may be verified using a similar line of reasoning. As derived previously,

$$\lambda''(c) = \frac{A}{(1+A)^3}\left((1+A)\alpha_2 + (1-A)(\alpha_1 + \alpha_2 c)^2\right). \tag{3.97}$$

Notice that $\lambda''(c) \leq 0$ precisely when

$$\frac{\alpha_2 + (\alpha_1 + \alpha_2 c)^2}{(\alpha_1 + \alpha_2 c)^2 - \alpha_2} \leq A, \tag{3.98}$$

where equality holds when $c = c^*$. For c increasing from c^*, A is increasing whereas the left hand of the above inequality is decreasing. Our claim thus follows. ■

By Lemma 3.13, the expected magnetization drift $(\lambda(c^*) - c^*)/n$ (Equation (3.15)) drops from first order to third order along the critical curve as compared with other parameter regions. As a result, we anticipate that the burn-in will be slower. The following Theorem 3.14 establishes an upper bound. Utilizing coupling techniques from Levin et al. [44] with minor adaptation, an $O(n^{2/3})$ mixing of the Glauber dynamics is further expected. For details, see Lemma 2.9 and Theorem 4.1 of [44].

Theorem 3.14. Along the critical curve, the burn-in time for the Glauber dynamics is $O(n^{3/2})$.

Proof. From Lemma 3.1,

$$\mathbf{E}(c_{t+1} - c^* \mid c_t - c^*) = \frac{1}{n}\left(\lambda(c_t) - \lambda(c^*)\right) + \left(1 - \frac{1}{n}\right)(c_t - c^*) + O\left(\frac{1}{n^2}\right). \tag{3.99}$$

Let $e_t = c_t - c^*$ and define $g(e_t) = \lambda(e_t + c^*) - \lambda(c^*)$. Then

$$\mathbf{E}(|e_{t+1}| \mid e_t) = \frac{1}{n} g(|e_t|) + \left(1 - \frac{1}{n}\right)|e_t| + O\left(\frac{1}{n^2}\right) \text{ for } e_t \geq 0,$$

$$\mathbf{E}(|e_{t+1}| \mid e_t) = -\frac{1}{n} g(-|e_t|) + \left(1 - \frac{1}{n}\right) |e_t| + O\left(\frac{1}{n^2}\right) \text{ for } e_t < 0. \tag{3.100}$$

Define $\tau_0 = \min\{t \geq 0 : |e_t| \leq 1/n\}$. Note that e_t does not change sign when $t < \tau_0$. Multiplying both sides of Equation (3.100) by the indicator function $\mathbf{1}_{\{\tau_0 > t\}}$ and using that $g(0) = 0$ and $\mathbf{1}_{\{\tau_0 > t+1\}} \leq \mathbf{1}_{\{\tau_0 > t\}}$, we have

$$\mathbf{E}\left(|e_{t+1}|\mathbf{1}_{\{\tau_0 > t+1\}} \mid e_t\right) \leq \frac{1}{n} g\left(|e_t|\mathbf{1}_{\{\tau_0 > t\}}\right) + \left(1 - \frac{1}{n}\right) |e_t|\mathbf{1}_{\{\tau_0 > t\}} + O\left(\frac{1}{n^2}\right),$$

$$\mathbf{E}(|e_{t+1}|\mathbf{1}_{\{\tau_0 > t+1\}} \mid e_t) \leq -\frac{1}{n} g(-|e_t|\mathbf{1}_{\{\tau_0 > t\}}) + \left(1 - \frac{1}{n}\right) |e_t|\mathbf{1}_{\{\tau_0 > t\}} + O\left(\frac{1}{n^2}\right). \tag{3.101}$$

Let $\theta_t^+ = \mathbf{E}\left(|e_t|\mathbf{1}_{\{\tau_0 > t\}}\right)$. By Lemma 3.13, $g(e) \leq 0$ for $e \geq 0$ and $g(e) \geq 0$ for $e \leq 0$, g is concave down on the non-negative axis and concave up on the negative axis. Taking expectation of both sides of Equation (3.101) and applying Jensen's inequality on g,

$$\theta_{t+1}^+ - \theta_t^+ \leq -\frac{1}{n}\left(\theta_t^+ - g(\theta_t^+)\right) + O\left(\frac{1}{n^2}\right),$$

$$\theta_{t+1}^+ - \theta_t^+ \leq -\frac{1}{n}\left(g(-\theta_t^+) - \left(-\theta_t^+\right)\right) + O\left(\frac{1}{n^2}\right). \tag{3.102}$$

Let $\mu > 0$ and suppose that $\theta_t^+ \geq \mu$. As in the proof of Theorem 3.2, by the extreme value theorem, there exists $\gamma(\mu) > 0$ such that

$$\theta_{t+1}^+ - \theta_t^+ \leq -\frac{\gamma(\mu)}{n}. \tag{3.103}$$

Utilizing the negative drift $-\gamma(\mu)/n$, there exists a time $t^* = O(n)$ so that $\theta_t^+ \leq 1/4$ for all $t \geq t^*$. Consider the Taylor series expansion of $g(\theta_t^+)$ and $g(-\theta_t^+)$ and using Lemma 3.13:

$$g(\theta_t^+) = \lambda'(c^*)\theta_t^+ + \frac{\lambda''(c^*)}{2}(\theta_t^+)^2 + \frac{\lambda'''(d_1)}{6}(\theta_t^+)^3 \leq \theta_t^+ - \frac{4}{3}(\theta_t^+)^3,$$

$$g(-\theta_t^+) = -\lambda'(c^*)\theta_t^+ + \frac{\lambda''(c^*)}{2}(\theta_t^+)^2 - \frac{\lambda'''(d_2)}{6}(\theta_t^+)^3 \geq -\theta_t^+ + \frac{4}{3}(\theta_t^+)^3, \tag{3.104}$$

where $d_1 \in [c^*, c^* + \theta_t^+]$ and $d_2 \in [c^* - \theta_t^+, c^*]$. Then it follows that

$$\theta_{t+1}^+ \leq \theta_t^+ - \frac{4}{3n}(\theta_t^+)^3 + O\left(\frac{1}{n^2}\right) \tag{3.105}$$

for $t \geq t^*$.

The remainder of the proof follows analogously as in the proof of Theorem 4.1 in [44]. For some c,

$$\lim_{c\to\infty} \mathbf{P}\left(\tau_0 > cn^{3/2}\right) = 0 \tag{3.106}$$

uniformly in n, which implies that the Glauber dynamics may be coupled so that the magnetizations agree in $O(n^{3/2})$ time steps. ■

3.6 Generalizations and future work

Numerous extensions can be made about these vertex-weighted exponential models. We started the discussion with the edge-triangle lattice gas (Ising) model, but clearly more complicated subgraph densities can be considered. Denote by K_k a complete graph on k vertices so that an edge is K_2 and a triangle is K_3. For consistency, denote by $C_0 = 1$, $C_1 = S(X, i)$, and $C_2 = T(X, i)$. As in Equation (3.7), the crucial quantity

$$C_m(X, i) = \sum_{i \neq i_1 \neq i_2 \neq \ldots \neq i_m} X_{i_1} X_{i_2} \cdots X_{i_m} \tag{3.107}$$

satisfies $C_m = \binom{C_1}{m}$, which may be justified by noting that a term in the sum defining C_m has value 1 if and only if every vertex in the product has spin 1. Moving on, we have employed a discrete-time update of the network, but the network may be updated on a continuous-time basis, and this may be realized by posing iid Poisson clocks and examining the corresponding heat kernel. More significantly, rather than the simplifying assumption that a person is either interested or not in building a friendship, in reality a person probably has different levels of interest in forming a connection, and an edge is placed between two people when the joint interest exceeds a certain threshold value. Also, in social networks people have diverse attributes; only people with the same attribute or those with more than a specified number of attributes will establish a tie, which will fall within the regime of the random cluster model and multilayer networks. All these extensions are quite challenging both theoretically and computationally, especially when network geometry comes into play, but we hope to address at least some of them in future work.

After we gain an understanding of the small-world observed structure of big network data using Markov chain dynamics, we may use this knowledge for the prediction and control of general spreading processes on large-scale networks. These processes include the social influence of opinions, users' decisions to adopt products, and epidemic intervention strategies, etc. To illustrate, we return to information diffusion over Twitter mentioned at the beginning of this section. Updating the weight of a vertex corresponding to a celebrity or a news source will definitely have more impact than that for ordinary people. So instead of running the Glauber dynamics that chooses a vertex at random, we choose the "hubs" of the network to update. This selective procedure decreases the mixing time and drives the spreading dynamics more efficiently towards equilibrium. Other properties of the chains may be studied simultaneously. For example, the cover time of the network may be interpreted as a realization of a "web crawl", and the hitting time may be interpreted as the necessary local queries to determine the global connectivity.

Chapter 4: Parking functions and orders on Coxeter systems

4.1 Introduction

We now turn our focus to the second topic of this dissertation and begin by briefly recalling the theory of parking functions, introduced in various contexts in [43, 56, 62]; see [71] for a comprehensive survey. Consider a parking lot with n parking spots placed sequentially along a one-way street. A line of n labelled cars enter the street, one by one, beginning with car 1 and ending with car n where each car has a spot in which they prefer to park, denoted $a(i)$. The i^{th} car drives to its preferred spot $a(i)$ and parks there if it is open; if the spot is already occupied then the car parks in the first available spot after $a(i)$. The list of preferences $a = (a(1), \ldots, a(n))$ is called a *parking function* if all cars successfully park; in this case the *outcome* is the permutation $\mathcal{O}(a) = x = (x(1), \ldots, x(n))$, where the i^{th} car parks in spot $x(i)$. It is well-known that there are $(n+1)^{n-1}$ parking functions of length n.

4.1.1 Overview. Parking functions were first defined by Konheim and Weiss [43] while researching computer storage as a way to study hash functions. Over the years, they have become an interesting and well established area of combinatorics research in their own right. In fact, parking functions have become a central object of study in combinatorics with connections to labelled trees, non-crossing partitions, lattice paths, the Shi arrangement, and other topics [43, 66, 71, 65]. There are also many extensions of classical parking functions that have been studied as well such as $\mathbf{u}$-parking functions, G-parking functions, parking with variable size cars, etc [71, 55, 31].

In this section, we study a generalization of parking functions in which the i^{th} car is willing to park only in the discrete interval

$$[\![a(i), b(i)]\!] = \{a(i), a(i)+1, \ldots, b(i)\} \subseteq \{1, \ldots, n\}.$$

If all cars can successfully park then we say that the pair

$$(a, b) = ((a(1), \ldots, a(n)), (b(1), \ldots, b(n)))$$

is an *interval parking function*, or IPF, of length n. Note that if $b(i) = n$ for all i, then we recover the classical case described above. It is easy to show that there are $n!(n+1)^{n-1}$ IPFs of length n, and that if (a, b) is an IPF then the sequences a and $b^* = (n+1-b(n), \ldots, n+1-b(1))$ must both be parking functions, raising the question of the relationship between the permutations $\mathcal{O}(a)$ and $\mathcal{O}(b^*)$. Both of these facts are stated and proven in Section 4.1.3.

We say that a pair of permutations $(x, y) \in \mathfrak{S}_n \times \mathfrak{S}_n$ is *reachable*, written $x \trianglerighteq_R y$, if there exists an IPF (a, b) such that $x = \mathcal{O}(a)$ and $y^* = \mathcal{O}(b^*)$. Reachability is *not* a partial order on $\mathfrak{S}_n$ because it is not transitive; however, its transitive closure is a partial order, which we call the *pseudoreachability order*. One of the main results of this section is that the pseudoreachability order on $\mathfrak{S}_n$ is precisely the *bubble-sort order* on $\mathfrak{S}_n$ (see [13, Example 3.4.3]), which in turn is an instance of the more general *sorting order* defined by Armstrong for Coxeter systems [9]. In particular, pseudoreachability lies between Bruhat and (left) weak order in $\mathfrak{S}_n$, and it is a self-dual distributive lattice, poset-isomorphic to the product $C_2 \times \cdots \times C_n$, where C_i denotes the chain with i elements.

We also investigate connections between reachability and pattern avoidance conditions on the permutations x and y. Section 5.5 contains partial results in this direction: Theorems 5.12 and 5.16 give sufficient conditions for a pair (x, y) to be reachable, provided that $x \geq_B y$ where $\geq_B$ is the Bruhat order on $\mathfrak{S}_n$. We will introduce and discuss some prelim-

inaries of this order in Section 4.4.3. Finally, we conjecture that the number of reachable pairs $(x, y) \in \mathfrak{S}_n \times \mathfrak{S}_n$ gives, what we believe to be, the first combinatorial interpretation of a certain analytically defined exponential generating function; the details are in Section 5.6.

4.1.2 Preliminaries. Double square brackets denote discrete integer intervals. Define $[\![m, n]\!] = \{m, \ldots, n\}$ for $m, n \in \mathbb{Z}$ and $[\![n]\!] = [\![1, n]\!]$.

Lists of positive integers (including permutations) will be regarded as functions, so we will write $a = (a(1), \ldots, a(n))$ rather than $a = (a_1, \ldots, a_n)$. Thus notation such as $x\,[\![a, b]\!]$ is the image of the discrete interval $[\![a, b]\!]$ under the function x; therefore, $x\,[\![a, b]\!] = \{x(a), x(a+1), \ldots, x(b)\}$. To simplify notation, we sometimes drop the parentheses and commas: e.g., $2431 = (2, 4, 3, 1)$.

Let $a = (a(1), \ldots, a(n))$ and $b = (b(1), \ldots, b(n)) \in \mathbb{Z}^n$. We write $a \leq_C b$ if $a(i) \leq b(i)$ for all $i \in [\![n]\!]$; this is the *componentwise partial order* on $\mathbb{Z}^n$. The *reverse complement* of $x \in [\![n]\!]^n$ is the vector $x^* = (n+1-x(n), \ldots, n+1-x(1))$. Reverse complementation is an involution that reverses componentwise order. Specifically, $(x^*)^* = x$ and $x \leq_C y \iff y^* \leq_C x^*$ for all $x, y \in \mathbb{Z}^n$.

4.1.3 Parking functions and interval parking functions. We now recall some facts about parking functions before proceeding to our central object of study. Let $a \in [\![n]\!]^n$ be a list of parking preferences. Algorithm 1 defines a $O(n^2)$ parking procedure to determine whether a list of preferrences a is a parking function. Note that if a is a parking function, the algorithm can easily be amended to return the outcome as well.

If Algorithm 1 succeeds in parking every car, then the preference vector a is called a *parking function*. The set of all parking functions $a = (a(1), \ldots, a(n))$ is denoted PF_n. It is well known that $|\,\mathsf{PF}_n\,| = (n+1)^{n-1}$. This was first proven by Konheim and Weiss [43] in their paper defining parking functions by analytic methods. A shorter, more elegant proof due to Pollack was presented by Riordan [62]. Roughly speaking, Pollack's proof proceeds as follows. In addition to the original n spaces, another space is added and the $n+1$

Algorithm 1 Parking Procedure A

```
1: procedure PARK(a)
2:     Initialize list of open spaces
3:     for i = 1, ..., n do
4:         Try to park car i in spot a(i)
5:         while car i has not parked do
6:             if car i tries spot j and spot j is open then
7:                 remove spot j from list of open spaces
8:             else
9:                 move on to spot j + 1
10:            end if
11:        end while
12:        if car i has reached the end of the street then
13:            return False
14:        end if
15:    end for
16:    return True
17: end procedure
```

spaces are arranged clockwise in a circle. The cars enter the street with preferred spaces $1, 2, \ldots, n, n+1$ and the $n+1$-st space is treated like any other preference. Cars begin at space 1 and move clockwise around the circle until they park in the first available space at or after their desired space. Since there are $n+1$ spaces and n cars, every preference sequence leaves 1 space unoccupied. Suppose that $(a_1, \ldots, a_n)$ is a preference sequence where preferences are chosen from $[\![1, n+1]\!]$. Then there exists a unqiue $j \in [\![1, n+1]\!]$ such that

$$((a_1 + j, \ldots, a_n + j) \pmod{n+1}) + 1$$

is a parking function. There are $(n+1)^n$ possible preference sequences and, due to symmetry, $1/(n+1)$ of them leave the $n+1$-st space unoccupied. Therefore

$$|\,\mathsf{PF}_n\,| = \frac{(n+1)^n}{(n+1)} = (n+1)^{n-1}.$$

In fact, an equivalent characterization to all cars with preference list a parking is that $\tilde{a}(i) \leq i$ for all i where $\tilde{a}$ is the non-decreasing rearrangement of a. We state and prove this well-known fact here for thoroughness of exposition.

Proposition 4.1. $a \in [\![n]\!]^n$ is a parking function if and only if $\tilde{a}(i) \leq i$ for all $i \in [\![n]\!]$.

Proof. Let $a \in [\![n]\!]^n$ and assume that there exists an $i \in [\![n]\!]$ such that $\tilde{a}(i) > i$. Then there are $n + 1 - i$ cars in a that will try to park in the last $n + 1 - \tilde{a}(i) < n + 1 - i$ spots, therefore a cannot be a parking function. Conversely, assume that the parking preference list $a \in [\![n]\!]^n$ does not permit all cars to park. Let j be the empty spot with the smallest index. We know that at least $n+1-j$ cars preferred parking in the last $n-j$ spots, therefore $\tilde{a}(j) \geq j + 1 > j$. ■

This leads to the characterization of the set of parking functions of length n as

$$\mathsf{PF}_n = \{a \in [\![n]\!]^n : \ \tilde{a}(i) \leq i \ \ \forall i\}.$$

In particular, every rearrangement of a parking function is a parking function. Thus we can use this characterization to write a shorter algorithm to determine whether a is a parking function. Since the keys in a are in the range $[\![1, n]\!]$, we can use counting sort to acheive $O(n)$ worst case run time.

Algorithm 2 Parking Procedure B

```
1: procedure PARK(a, n)
2:     for i = 1, ..., n do
3:         if sorted a(i) > i then
4:             return False
5:         end if
6:     end for
7:     return True
8: end procedure
```

In fact, we can take advantage of the proof that there are $(n+1)^{n-1}$ parking functions of length n and Proposition 4.1 to give a procedure that returns a uniformly random parking function. We note that this fact is known and appears in Diaconis and Hicks [28].

Algorithm 3 Generate Random Parking Function

```
1: procedure RandPark(n)
2:     Generate uniformly random a ∈ Z_{n+1}^n
3:     while not Park(a, n) do
4:         a ← a + (1, ..., 1) (mod n + 1)
5:     end while
6:     return a + (1, ..., 1)
7: end procedure
```

From our informal proof of the number of parking functions of length n, there exists a unique j such that

$$((a_1 + j, \ldots, a_n + j) \pmod{n+1}) + 1$$

is a parking function. Therefore the while loop executes at most $n+1$ times and we have a linear time algorithm to generate a random parking function.

We now modify Algorithm 1 to obtain our central object of study. Again, consider a parking lot with n parking spots placed sequentially along a one-way street. Now each car has space they prefer to park in, $a(i)$, as well as a spot after which they refuse to park, $b(i)$. A line of n labelled cars enter the street, one by one, beginning with car 1 and ending with car n where each car has a spot in which they prefer to park, denoted $a(i)$. The i^{th} car drives to its preferred spot $a(i)$ and parks there if it is open; if the spot is already occupied then the car parks in the first available spot after $a(i)$ before spot $b(i)+1$. The pair (a, b) is called an *interval parking function* if all cars successfully park

Definition 4.1. If Algorithm 4 succeeds in parking every car, then $\mathbf{c} = (a, b)$ is called an *interval parking function* of length n, or IPF. The set of all interval parking functions for n cars is denoted IPF_n. The *feasible interval* for the i^{th} car is $[\![a(i), b(i)]\!]$.

Algorithm 4 Interval Parking Procedure

procedure INTERVALPARK(a, b)
  Initialize list of open spaces
  for $i = 1, \ldots, n$ **do**
    Try to park car i in spot $a(i)$
    while car i has not parked **do**
      if car i tries spot j and spot j is open **then**
        remove spot j from list of open spaces
      else
        move on to spot $j + 1$
      end if
    end while
    if car i has reached spot $b(i) + 1$ or the end of the street **then**
      return False
    end if
  end for
  return True
end procedure

For example,

$$\mathsf{IPF}_2 = \{(11, 12),\ (11, 22),\ (12, 12),\ (12, 22),\ (21, 21),\ (21, 22)\}.$$

Unlike ordinary parking functions, IPFs are *not* invariant under the action of $\mathfrak{S}_2$ by permuting cars. For example, $(11, 12)$ is an IPF but $(11, 21)$ is not.

4.2 Properties of interval parking functions

Proposition 4.2. Let $a, b \in [n]^n$. Then:

1. $a \in \mathsf{PF}_n$ if and only if $(a, (n, \ldots, n)) \in \mathsf{IPF}_n$.

2. $(a, b) \in \mathsf{IPF}_n$ if and only if $a \in \mathsf{PF}_n$ and $\mathcal{O}(a) \leq_C b$.

Proof. For (1), if $b(i) = n$ for all i then Algorithm 4 is identical to Algorithm 1. For (2), if the given conditions hold, then the execution of Algorithm 4 mimics that of Algorithm 1. On the other hand, if a is not a parking function, then some car will not find a spot, while if $\mathcal{O}(a) \not\leq_C b$ then some car will not find a spot in its own feasible interval. ■

As a consequence of the proof of Proposition 4.2 2, the outcome $\mathcal{O}(\mathbf{c})$ of $\mathbf{c} = (a, b)$ is just $\mathcal{O}(a)$. Moreover, for every $a \in \mathsf{PF}_n$, there are precisely $n!$ choices for b such that $(a, b) \in \mathsf{IPF}_n$. In particular,

$$|\mathsf{IPF}_n| = n!(n+1)^{n-1}. \tag{4.1}$$

Proposition 4.3. Let $\mathbf{c} = (a, b) \in \mathsf{IPF}_n$. Then

1. $a \leq_C \mathcal{O}(\mathbf{c}) \leq_C b$,
2. $b^* \in \mathsf{PF}_n$, and
3. $\mathcal{O}(b^*)^* \leq_C b$.

Proof. 1. Evidently $a \leq_C \mathcal{O}(\mathbf{c}) \leq_C b$ based on Algorithm 4.

2. Since $\mathcal{O}(\mathbf{c}) \leq_C b$, $b^* \leq_C \mathcal{O}(\mathbf{c})^*$ and $\mathcal{O}(\mathbf{c})^*$ is a permutation, therefore b^* is a parking function.

3. By (1), b^* is a parking function. Thus $b^* \leq_C \mathcal{O}(b^*)$. Conjugation reverses the order $\leq_C$ and is an involution, so $\mathcal{O}(b^*)^* \leq_C (b^*)^* = b$. ■

4.3 Coxeter groups

Before moving on to the main results of interval parking functions, we recall some facts about Coxeter groups and various partial orders that are significant in the combinatorics of Coxeter groups. Coxeter groups are important mathematical objects that lie in the intersection of algebraic geometry, group theory, and combinatorics, as well as the theory of partially ordered sets and lattices. Informally speaking, Coxeter groups, named after the geometer H. S. M. Coxeter, are abstract groups where the relations between generators are written in terms of reflections. Formally, let S be a set of *generators* where

$$S = \{r_1, \ldots, r_n\}.$$

A *Coxeter group* is a group W with presentation

$$\langle r_1, \ldots, r_n \mid (r_i r_j)^{m_{ij}} = 1 \rangle$$

where $m_{ii} = 1$ and $m_{ij} \geq 2$ for all $i \neq j$. It is permissible that $m_{ij} = \infty$, this convention implies that there is no relation of the form $(r_i r_j)^m = 1$ for any m. The pair (W, S) is referred to as a *Coxeter system*.

We will be limiting ourselves to studying the symmetric group, $\mathfrak{S}_n$, as a Coxeter system of type A_{n-1}. In particular, we are concerned with the symmetric group with generators

$$S = \{s_1, \ldots, s_{n-1}\}$$

given by adjacent transpositions, $s_i = (i,\ i+1)$, and presentation

$$\left\langle s_1, \ldots, s_{n-1} \mid (s_i\, s_i)^1 = 1,\ (s_i\, s_{i+1})^3 = 1,\ (s_i\, s_j)^2 = 1 \ \text{ for all } \ |i-j| > 1 \right\rangle.$$

It will also be helpful to have a notation for any transposition of two elements i and j, we denote this t_{ij}. With this notation, we have $s_i = t_{i,i+1}$. We will write permutations as words in S. Note that there are infinitely many ways that one could write an element $x \in \mathfrak{S}_n$ as a word in S. For example, $2143 \in \mathfrak{S}_4$ can be written as

$$s_1 s_3 = s_1 s_2 s_2 s_3 = s_2 s_3 s_2 s_3 s_2 s_3 s_1 s_3,$$

and in many other, even more inefficient, ways. The *length* of the word w, denoted $\ell(w)$, is defined to be the length of the smallest word for w in S. A word of length $\ell(w)$ is a *reduced word* for w. For the example of $w = 2143$, the word $s_1 s_3$ is a reduced word for w, as is $s_3 s_1$, and $\ell(w) = 2$.

4.4 Partial orders on Coxeter systems

Our interest in the symmetric group as a Coxeter system lies mainly in the theory of partial orders on $\mathfrak{S}_n$. First we recall some facts about partial orders and $\mathfrak{S}_n$. If $>$ is a partial ordering on a set S, then $\gtrdot$ denotes a covering relation. Recall that $x \gtrdot y$ is a *covering relation* if $x > y$ and there exists no z such that $x > z > y$. Furthermore, if $>_1$ is a partial order at least as strong as $>_2$, that is, $x >_2 y$ implies $x >_1 y$, then $x >_2 y$ and $x \gtrdot_1 y$ together imply $x \gtrdot_2 y$.

The symmetric group of all permutations of $[\![n]\!]$ is denoted by $\mathfrak{S}_n$. We will as far as possible follow the notation and terminology for the symmetric group used in [13]. We set $e = (1, \ldots, n)$ (the identity permutation) and $w_0 = (n, n-1, \ldots, 1)$. Note that permutations in this section are written in *one-line* notation. The permutation transposing i and j and fixing all other values is denoted t_{ij}, and we set $s_i = t_{i,i+1}$. The elements $s_1, \ldots, s_{n-1}$ are the *standard generators* for $\mathfrak{S}_n$. Our convention for multiplication is right to left, which is consistent with treating permutations as bijective functions $[\![n]\!] \to [\![n]\!]$. Thus $t_{ij}x$ is obtained by transposing the *digits* i, j wherever they appear in x, while xt_{ij} is obtained by transposing the digits in the i^{th} and j^{th} *positions*.

Recall that a *graded poset* is a partially ordered set, $(P, >)$ equipped with a rank function $\rho : P \to \mathbb{N}$ that satisfies

- $\rho(x) < \rho(y)$ whenever $x < y$, and
- $\rho(y) = \rho(x) + 1$ whenever y covers x.

For our purposes, we will be concerned with $\mathfrak{S}_n$ and rank function $\ell : \mathfrak{S}_n \to \mathbb{N}$ where

$$\ell(x) = \left|\{(i,j) \ : \ x(i) > x(j) \text{ where } 1 \le i < j \le n\}\right|. \tag{4.2}$$

This notion of length coincides with the word length from above and counts the number of inversions in a permutation. In our study of IPF's, we will encounter three different partial orders on $\mathfrak{S}_n$: the weak order, sorting order, and Bruhat order.

4.4.1 Weak order.

Definition 4.2. The *left weak order*, denoted $\geq_W$, is the transitive closure of the relations $u > s_i u$ whenever $\ell(u) > \ell(s_i u)$.

There is also a right weak order that is analogously defined as $u > us_i$ whenever $\ell(u) > \ell(us_i)$. It turns out that the resulting posets are isomorphic and so we will restrict ourselves to considering the left weak order and may refer to it simply as the weak order. With ℓ as the rank function, $(\mathfrak{S}_n, \geq_W)$ is a graded poset with top element w_0 and bottom element e.

4.4.2 Sorting order. In 2009, Armstrong introduced a family of partial orders on an arbitrary Coxeter system collectively referred to as *sorting orders*. We briefly describe this family of partial orders, for more details see [9]. Let

$$S^* = S^0 \cup S^1 \cup S^2 \cup \cdots \cup S^\infty$$

be the collection of all finite and semi-infinite words in S, where S^n is the set of all words of length n in S and S^∞ is the set of all semi-infinite words in S. Given an arbitrary word $\omega \in S^*$, let $W_\omega \subseteq W$ be the set of group elements that appear as subwords of ω. The word ω is referred to as the *sorting word*. For every subword α of ω we identify α with its index set $I(\alpha) \subseteq I(\omega)$, where $I(\alpha)$ describes the positions of the letters of α in ω. For every element $w \in W_\omega$, let $\mathrm{sort}_\omega(w)$ denote the reduced word for w that is lexicographically first among subwords of ω. The word $\mathrm{sort}_\omega(w)$ is referred to as the *ω-sorted word* of w.

Definition 4.3. The *ω-sorting order*, denoted $\geq_\omega$, on W_ω is defined as $u \geq_\omega w$ if

$$I(\mathrm{sort}_\omega(w)) \subseteq I(\mathrm{sort}_\omega(u)).$$

One nice consequence of this definition is that for any sorting word $\omega \in S^*$, there exists a reduced subword ω' of ω such that the ω'-sorting order and the ω-sorting order coincide.

Example 4.1. Consider $\mathfrak{S}_4$ and let $\omega = s_1s_2s_3s_1s_2s_1$. Note that ω is a reduced word for the permutation 4321. Then $I(\omega) = \{1, 2, 3, 4, 5, 6\}$. For the permutation $\alpha = 3241$, the ω-sorting word for α is $\mathrm{sort}_\omega(\alpha) = s_1s_2s_3s_1$ and $I(\mathrm{sort}_\omega(\alpha)) = \{1, 2, 3, 4\}$.

An intuitive way to think about the sorting order is that we want to convert a group element $\alpha^{-1} \in W_\omega$ into the identity element and $\mathrm{sort}_\omega(\alpha)$ records the steps in the process. As previously stated, we will study how a specific sorting order on $\mathfrak{S}_n$ relates to outcomes interval parking functions. Let

$$\omega_n = (s_1s_2 \cdots s_{n-1})(s_1s_2 \cdots s_{n-2}) \cdots (s_1s_2)(s_1),$$

a reduced word for $w_0 \in \mathfrak{S}_n$, where certain consecutive subwords have been grouped for clarity. This grouping will be convenient later on as well. Note that in this case, $W_{\omega_n} = \mathfrak{S}_n$. The particular sorting order we are interested in is the ω_n-sorting order on $\mathfrak{S}_n$, this also referred to as the *bubble sorting order* because it coincides with the sequence of steps one takes to sort an array of numbers using the bubble sort algorithm, the popular, albeit inefficient, sorting algorithm one commonly learns in a first course in computer science. In particular, the ω_n-sorting order is a graded lattice that is strictly between the weak order and the Bruhat order. In particular, for $x, y \in \mathfrak{S}_n$, we have the following string of implications

$$x \geq_W y \implies x \geq_{\omega_n} y \implies x \geq_B y.$$

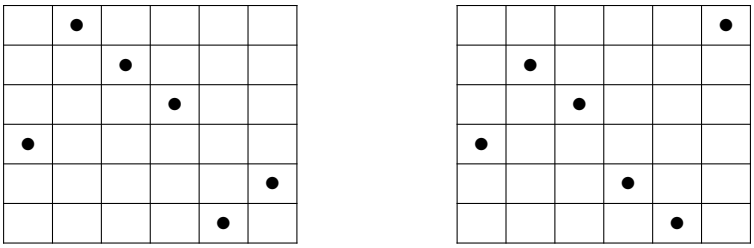

Figure 4.1: Diagram of the permuations $x = 365412$ and $y = 354216$.

Furthermore, the poset $(W, \geq_\omega)$ is a maximal lattice in the sense that if any Bruhat covers are added, the resulting poset is no longer a lattice.

4.4.3 Bruhat order. For a Coxeter system (W, S), the Bruhat order, also referred to as the strong order or the Bruhat-Chevalley order, is another partial order on W that is stronger than both the weak order and ω_n-sorting order. The Bruhat order can be defined in many equivalent ways and we will present a few different ways to think about it in this section.

Definition 4.4. The *Bruhat order*, denoted $\geq_B$, is the transitive closure of the relations $u > t_{ij}u$ whenever $\ell(u) > \ell(t_{ij}u)$.

This can be a very convenient description if we are only concerned with covering relations, but can be difficult to use when we are not dealing with a covering relation. For example, it is pretty clear to see that $x = 365412$ and $y = 354216$ do not differ by a single transposition. How can we determine whether x and y are comparable in Bruhat order, and, more precisely, how they compare to one another? We will present two different ways to do this. The first is a diagrammatic criterion and the second is a tableau criterion. For more information on these, see [13, 74]. The *diagram* of a permutation $z \in \mathfrak{S}_n$ is given by placing dots in the cells with coordinates $(i, z(i))$ in the plane for $1 \leq i \leq n$ where the lower left square has coordinates $(1, 1)$ and the upper right has coordinates (n, n). See Figure 4.1 for an example of the diagram for the permutations x and y. For $z \in \mathfrak{S}_n$, let

$$z\langle i, j\rangle = |\{k \in [\![i]\!] \ : \ \pi(k) \geq j\}| \tag{4.3}$$

for $i, j \in [\![n]\!]$. This quantity $z\langle i, j\rangle$ has a convenient interpretation in terms of the diagram of the permutation z, $z\langle i, j\rangle$ counts the number of dots in the northwest corner above and including the point with coordinates (i, j). For example, x from Figure 4.1 has $x\langle 2, 3\rangle = 2$, $x\langle 4, 3\rangle = 3$, and $x\langle 3, 1\rangle = 3$. In fact, for any $z \in \mathfrak{S}_n$,

$$z\langle n, i\rangle = n + 1 - i \text{ and } z\langle i, 1\rangle = i$$

for $i \in [\![n]\!]$.

Theorem 4.4 (Diagrammatic criterion for Bruhat order)**.** Let $x, y \in \mathfrak{S}_n$. Then $x \geq_B y$ if and only if $x\langle i, j\rangle \geq y\langle i, j\rangle$ for all $i, j \in [\![n]\!]$.

For proof, see [13]. We also give another equivalent formulation of the Bruhat order that we will not use in any proofs, but will be useful when thinking about how the situtation of interval parking functions relates to ordinary parking functions. This criterion, which is referred to as the *tableau criterion* for Bruhat order indicates how one could partially extend the non-decreasing rearrangement condition for determining whether a parking preference vector is indeed a parking function to interval parking functions.

Theorem 4.5 (Tableau criterion for Bruhat order)**.** Let $x, y \in \mathfrak{S}_n$. Then $x \geq_B y$ if and only if $x_{i,j} \geq y_{i,j}$ for all $1 \leq i \leq j \leq n$ where $x_{i,j}$ is the i-th entry in the non-decreasing rearrangement of $x_1, \ldots, x_j$ (and similarly for $y_{i,j}$).

For example, we can compute the tableau for $x = 365412$ and $y = 354216$ where row j is the non-decreasing rearrangement of $x_1, \ldots, x_j$ and $y_1, \ldots, y_j$, respectively. From the comparison of these tableaux, we see that $x \geq_B y$ by Theorem 4.5. Note that this is more difficult to determine using the description of Bruhat order based on covering relations or the diagrammatic criterion. In practice, we will mostly stick to using the definition of Bruhat order based on covering relations or the diagrammatic criterion in the coming sections.

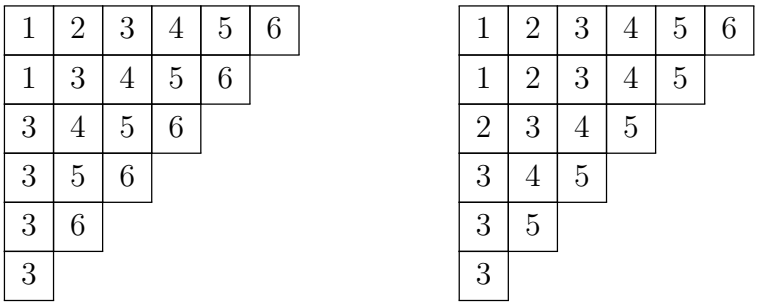

Figure 4.2: Tableau comparison of the permutations $x = 365412$ and $y = 354216$.

Chapter 5: Outcomes of interval parking functions and orders on the symmetric group

In this section we prove that outcomes of interval parking functions relate to various orders on $\mathfrak{S}_n$.

5.1 The Bruhat property

Our first main result is that outcomes of interval parking functions are related by the Bruhat order on permutations. We will use the characterization of the Bruhat order from Theorem 4.4 in our proof of Theorem 5.1. For later use, we observe that by the pigeonhole principle, it is always the case that

$$x\langle i, j\rangle \geq i - j + 1. \tag{5.1}$$

Suppose that $\mathbf{c} = (a, b)$ is an IPF, and let $x = \mathcal{O}(a)$ and $y = \mathcal{O}(b^*)^*$. Then, in this context, $x\langle i, j\rangle$ is the number of cars $1, \ldots, i$ that park at or after spot j under the parking function a.

Theorem 5.1. Suppose that $\mathbf{c} = (a, b)$ is an IPF. Let $x = \mathcal{O}(a)$ and $y = \mathcal{O}(b^*)^*$. Then $x \geq_B y$.

Proof. First, we may assume without loss of generality that $x = a$, because replacing a with x doesn't change the execution of Algorithm 4 (the i^{th} car will have to drive to spot $x(i)$ anyway, and it is able to park there because $\mathbf{c}$ is an IPF).

Fix $i, j \in [\![n]\!]$, and let $p = x\langle i, j\rangle$ and $q = y\langle i, j\rangle$. By Theorem 4.4 we wish to show that $p \geq q$. By definition of $y\langle i, j\rangle$ we have

$$\Big| y\,[\![1, i]\!] \cap [\![j, n]\!] \Big| = q \tag{5.2}$$

or equivalently

$$\Big| y^*\,[\![n - i + 1, n]\!] \cap [\![1, n + 1 - j]\!] \Big| = q. \tag{5.3}$$

Therefore, when Algorithm 1 is run on the parking function b^* with outcome y^*, the first $n - i$ cars must leave open at least q spaces in the range $[\![1, n + 1 - j]\!]$, so they cannot fill as many as $(n + 1 - j) - q + 1 = n - j - q + 2$ of them. Therefore, $b^*\,[\![1, n - i]\!]$ can contain no subset $\{v^*(1), \ldots, v^*(n - j - q + 2)\}$ such that

$$(v^*(1), \ldots, v^*(n - j - q + 2)) \leq_C (q, \ldots, n + 1 - j).$$

Equivalently, $\{b(i + 1), \ldots, b(n)\}$ can contain no subset $\{v(1), \ldots, v(n - j - q + 2)\}$ such that

$$(v(1), \ldots, v(n - j - q + 2)) \geq_C (j, \ldots, n - q + 1).$$

It follows that when Algorithm 4 is run on $\mathbf{c}$, no more than $n-j-q+1$ of the last $n-i$ cars will park in the spots $[\![j, n]\!]$. On the other hand, since $x = \mathcal{O}(\mathbf{c})$, no more than $p = x\langle i, j\rangle$ of the first i cars can park in the spots $[\![j, n]\!]$. Therefore, the total number of cars that park in $[\![j, n]\!]$ is at most

$$(n + 1 - j - q) + p = |\,[\![j, n]\!]\,| + (p - q).$$

On the other hand, exactly $|\,[\![j, n]\!]\,|$ cars park in $[\![j, n]\!]$. It follows that $p \geq q$, as desired.

■

Theorem 5.1 asserts that there is a well-defined *bioutcome* function

$$\begin{array}{rccl} \bar{\mathcal{O}}: & \mathsf{IPF}_n & \to & \{(x,y) \in \mathfrak{S}_n \times \mathfrak{S}_n : \ x \geq_B y\} \\ & (a,b) & \mapsto & (\mathcal{O}(a), \mathcal{O}(b^*)^*). \end{array} \tag{5.4}$$

We say that a pair $(x, y) \in \mathfrak{S}_n \times \mathfrak{S}_n$ is *reachable* if it is in the image of $\bar{\mathcal{O}}$; in this case we write $x \trianglerighteq_R y$. We use this notation rather than $x \geq_R y$ because reachability is not a partial order on $\mathfrak{S}_n$, as we will discuss shortly. With this notation, Theorem 5.1 asserts that all reachable pairs are related in Bruhat order.

Remark 5. If a and b^* are parking functions such that $\mathcal{O}(a) \geq_B \mathcal{O}(b^*)^*$, it does *not* follow that $\mathbf{c} = (a, b)$ is an IPF. For example, if $a = w_0$ and b is a permutation, then certainly $a = \mathcal{O}(a) \geq_B \mathcal{O}(b^*)^* = b$, but (a, b) is an IPF only if $b = w_0$ as well.

Moreover, if $x, y \in \mathfrak{S}_n$ with $x \geq_B y$, there does not necessarily exist any IPF $\mathbf{c} = (a, b)$ such that $\bar{\mathcal{O}}(\mathbf{c}) = (x, y)$. For example, when $n = 3$, take $(x, y) = (321, 213)$, so that $y^* = 132$. Then $a = 321$ is the only parking function with $\mathcal{O}(a) = x$. By Proposition 4.3(2) we must have $b \geq_C a$, so $b \in \{321, 331, 322, 332, 323, 333\}$ and $b^* \in \{321, 311, 221, 211, 121, 111\}$. But none of these parking functions have outcome $y^* = 132$.

The relation of reachability is reflexive (because $\bar{\mathcal{O}}(x, x) = (x, x)$ for all $x \in \mathfrak{S}_n$) and antisymmetric (as a consequence of Theorem 5.1). However, it is not transitive: for example, $321 \not\trianglerighteq_R 213$, as just shown, but $(321, 312) = \bar{\mathcal{O}}(312, 322)$ and $(312, 213) = \bar{\mathcal{O}}(312, 313)$ are reachable. This observation motivates the following definition.

Definition 5.1. We say that (x, y) is *pseudoreachable*, written $x \geq_P y$, if there is a sequence $x = x_0 \trianglerighteq_R x_1 \trianglerighteq_R \cdots \trianglerighteq_R x_k = y$. That is, pseudoreachability is the transitive closure of reachability. As such, it is a partial order on $\mathfrak{S}_n$, which by Theorem 5.1 is no stronger than Bruhat order.

$a \geq_C b$	Componentwise order	$\mathbb{Z}^n$
$x \geq_B y$	Bruhat order	$\mathfrak{S}_n$
$x \geq_W y$	Left weak order	
$x \trianglerighteq_R y$	Reachability (not transitive)	
$x \geq_P y$	Pseudoreachability	

Table 5.1: Summary of the orders considered in this section.

For reference, we summarize the various order-like relations that we will consider in Table 5.1.

5.2 Reachability via counting fibers of the bioutcome map

Fix a pair of permutations $(x, y) \in \mathfrak{S}_n \times \mathfrak{S}_n$. How can we determine if (x, y) is reachable? More generally, what is the number $\phi(x, y) = |\bar{\mathcal{O}}^{-1}(x, y)|$ of IPFs (a, b) with bioutcome (x, y)? We can answer this enumerative question rather quickly, although the resulting formula is recursive and somewhat opaque. First, for each i, the number of possibilities $\mathsf{c}_i = \mathsf{c}_i(x, y)$ for $a(i)$ is the size of the largest block of spaces ending in $x(i)$ that are all occupied by one of the first i cars. That is,

$$\mathsf{c}_i = \mathsf{c}_i(x, y) = \max\left\{j \in [\![1, x(i)]\!] : x^{-1}(x(i) - k) \leq i \text{ for all } 0 \leq k \leq j - 1\right\}.$$

Second, given $a(1), \ldots, a(i)$, the number of possibilities for $b(i)$ is $\mathsf{d}_i = \mathsf{d}_i(x, y) = |\mathsf{D}_i(x, y)|$, where

$$\mathsf{D}_i(x, y) = \{k \in [\![0, J_i - 1]\!] : \ y(i) + k \geq x(i)\}$$

and

$$J_i = \max\{j \in [\![1, n + 1 - y(i)]\!] : \ y^{-1}(y(i) + k) \geq i \text{ for all } 0 \leq k \leq j - 1\}.$$

The definition of J_i is analogous to that of c_i: it is the size of the largest block of spaces ending in $n+1-y(i)$ that are all occupied by one of the first $n+1-i$ cars, so it is the number of possible values for b_i^* under which $\mathcal{O}(b^*) = y^*$. The additional condition $y(i)+k \geq x(i)$ in the definition of D_i ensures that (a,b) is an IPF because the upper bound on $x(i)$ given by $b(i)$ does not conflict with where the i^{th} car parks under Algorithm B.

The sequences $\mathsf{c} = (\mathsf{c}_1, \ldots, \mathsf{c}_n)$ and $\mathsf{d} = (\mathsf{d}_1, \ldots, \mathsf{d}_n)$ then determine the size of the fibers of $\bar{\mathcal{O}}$:

$$\phi(x,y) = \left|\bar{\mathcal{O}}^{-1}(x,y)\right| = \prod_{i=1}^{n} \mathsf{c}_i \mathsf{d}_i. \tag{5.5}$$

Example 5.1. Let $x = 361245$ and $y = 341256$. Then $\mathsf{c} = (1,1,1,2,4,5)$ and $\mathsf{d} = (4,1,2,1,2,1)$, so there are $2^3 4^2 5^1 = 640$ IPFs with bioutcome (x,y).

It is clear from the definition that $1 \leq \mathsf{c}_i \leq i$ for all i. On the other hand, one or more d_i may be zero. The pair (x,y) is reachable if and only if $\mathsf{d}_i > 0$ for all i; we refer to this as the *Count Criterion* for reachability. The equations given above for c_i and d_i are quite difficult to interpret. We can instead rephrase them in terms similar to the rearrangement condition for ordinary parking functions as in Proposition 4.1.

Proposition 5.2. Let $x, y \in \mathfrak{S}_n$. Let $L(i)$ denote the longest consecutive subsequence in the nondecreasing rearrangment of $y(i), \ldots, y(n)$ that starts with $y(i)$. Then $\mathsf{c}_i(x,y)$ is the length of the longest consecutive subsequence ending in $x(i)$ in the nondecreasing rearrangment of $x(1), \ldots, x(i)$. Similarly, $\mathsf{d}_i(x,y)$ is the length of the longest consecutive subsequence of $L(i)$ starting with $\max\{x(i), y(i)\}$. If no such subsequence exists, then $\mathsf{d}_i(x,y) = 0$.

The proof of Proposition 5.2 follows from unpacking the definitions of $\mathsf{c}_i, \mathsf{D}_i$, and J_i. Evidently, the largest fiber occurs when x and y both equal the identity permutation in $\mathfrak{S}_n$. In this case $\mathsf{c} = (1,2,\ldots,n)$ and $\mathsf{d} = (n, n-1, \ldots, 1)$, and the fiber size is $(n!)^2$. At the opposite end of the spectrum, if $x = y = (n, \ldots, 1)$, then $\phi(x,y) = 1$.

Perhaps a better way to think about reachability is the following criterion. If we are solely interested in reachability and not the number of IPFs that achieve a given outcome, we can rephrase reachability more directly in terms of the permutations x and y.

Theorem 5.3 (*Reachability Criterion*)**.** Let $x, y \in \mathfrak{S}_n$. Then

$$x \trianglerighteq_R y \quad \iff \quad [\![y(i), x(i)]\!] \subseteq y\,[\![i, n]\!] \quad \forall\, i \in [\![n]\!]\,. \tag{RC}$$

Proof. Let $i \in [\![n]\!]$. We will show that $\mathsf{d}_i(x, y) > 0$ if and only if $[\![y(i), x(i)]\!] \subseteq y\,[\![i, n]\!]$.

Suppose that $[\![y(i), x(i)]\!] \setminus y\,[\![i, n]\!] \neq \emptyset$. That is, there is some $m \in [\![y(i), x(i)]\!]$ such that $y^{-1}(m) < i$. Thus $J_i \leq m - y(i)$, so $y(i) + k < m \leq x(i)$ for all $k < J_i$, so $\mathsf{d}_i(x, y) = 0$.

Now assume that $[\![y(i), x(i)]\!] \subseteq y\,[\![i, n]\!]$. We wish to show that $\mathsf{D}_i \neq \emptyset$. If $y(i) \geq x(i)$, then $0 \in \mathsf{D}_i$. On the other hand, if $y(i) < x(i)$, then $m = x(i) - y(i) > 0$, and for all $0 \leq k \leq m$ we have $y^{-1}(y(i) + k) \geq i$. Therefore $J_i > m$ and $m \in \mathsf{D}_i$. ■

Remark 6. We would like to remark here that if $x \trianglerighteq_R y$, then we can intuitively think of the Reachability Criterion as implying that, if $y(i) < x(i)$, values larger than $y(i)$ appear later in y than they do in x. This can, in a very rough sense, be thought of as telling us that y is "more sorted" than x.

It is worth emphasizing that the Reachability Criterion is sufficient, but not necessary, for showing that $x \geq_P y$. For example, the pair $(x, y) = (321, 213)$ fails (*RC*) for $i = 2$, but nonetheless $x \geq_P y$.

Proposition 5.4. The sequence $\mathsf{d}(x, y)$ has the following properties.

(a) $\mathsf{d}_1 \geq 1$.

(b) For each i, if $y(i) \geq x(i)$, then $\mathsf{d}_i \geq 1$.

(c) If $x \geq_B y$, then $\mathsf{d}_n = 1$.

Proof. The first two assertions are direct consequences of (*RC*). For (a), we have

$$[\![y(1), x(1)]\!] \subseteq [\![n]\!] = y\,[\![n]\!]\,,$$

and for (b), if $y(i) \geq x(i)$ then $[\![y(i), x(i)]\!] \subseteq \{y(i)\} \subseteq y\,[\![i, n]\!]$.

For (c), if $y \leq_B x$, then $y(n) \geq x(n)$ (a consequence of the inequalities (4.3) for $i = n-1$ and all j), so $\mathsf{d}_n > 0$ by part (b). Observe that

$$J_n = \max\{j : \ y(n) + k \leq n \text{ and } y^{-1}(y(n)+k) \geq n \text{ for all } 0 \leq k \leq j-1\} = 1$$

because the conditions are true for $k = 0$ but false for $k > 0$. Therefore, $\mathsf{D}_n = \{k \in [\![0, 0]\!] : y(n) \geq x(n)\} = \{0\}$ and $\mathsf{d}_n = |\mathsf{D}_n| = 1$. ∎

5.3 Pseudoreachability order is graded

In this section, we prove that the pseudoreachability order $\geq_P$ on $\mathfrak{S}_n$ is graded by length, just like the Bruhat and weak orders. Temporarily, we will use the notation $x \rhd_R y$ to mean that $x \trianglerighteq_R y$ and $\ell(x) = \ell(y) + 1$. Note that if $x \rhd_R y$ then $x \succcurlyeq_P y$ (because $x \succcurlyeq_B y$). Our goal is to prove the converse of the last statement, which will imply that pseudoreachability is graded by length. The bulk of this is accomplished in Proposition 5.8. In broad strokes, the proof proceeds by double induction on n and $m = \ell(x) - \ell(y)$. If the last entries in x and y are equal, we can use Lemma 5.6 to project these permutations down to $\mathfrak{S}_{n-1}$ and the existence of a chain follows by the induction hypothesis. This chain can then be lifted back up to $\mathfrak{S}_n$ using Corollary 5.7. If they are not equal, then we appeal to the Bruhat order and the left weak order to construct a chain from y to x.

We have already shown that pseudoreachability order is no stronger than Bruhat order $\geq_B$. We next show that it is no weaker than left weak order $\geq_W$.

Proposition 5.5. If $x \gtrdot_W y$, then $x \rhd_R y$.

Proof. Suppose that $x \gtrdot_W y$, i.e., that $x = s_a y$, where $j = y^{-1}(a) < y^{-1}(a+1) = k$. Then Proposition 5.4(b) implies that $\mathsf{d}_i(x, y) > 0$ for all $i \in [\![n]\!] \setminus \{j\}$. Meanwhile $[\![y(j), x(j)]\!] = \{a, a+1\} = \{y(j), y(k)\} \subseteq [\![y(j), y(n)]\!]$, so (*RC*) implies that $\mathsf{d}_j(x, y) > 0$ as well. ■

For each $x \in \mathfrak{S}_n$, let $\hat{x}$ be the permutation in $\mathfrak{S}_{n-1}$ defined by

$$\hat{x}(i) = \begin{cases} x(i) & \text{if } x(i) < x(n), \\ x(i) - 1 & \text{if } x(i) > x(n). \end{cases} \tag{5.6}$$

Lemma 5.6. Let $x, y \in \mathfrak{S}_n$ with $x(n) = y(n)$. Then $x \trianglerighteq_R y$ if and only if $\hat{x} \trianglerighteq_R \hat{y}$.

Proof. By (*RC*), the proof reduces to showing that

$$[\![y(i), x(i)]\!] \subseteq y\,[\![i, n]\!] \qquad \forall i \in [\![n]\!] \tag{5.7a}$$

if and only if

$$[\![\hat{y}(i), \hat{x}(i)]\!] \subseteq \hat{y}\,[\![i, n]\!] \qquad \forall i \in [\![n-1]\!]\,. \tag{5.7b}$$

($\Longrightarrow$) Assume that (5.7a) holds. Let $i \in [\![n-1]\!]$ and $a \in [\![\hat{y}(i), \hat{x}(i)]\!]$. There are two cases to consider.

Case 1a: $a < y(n)$. Then $\hat{y}(i) \leq a < y(n)$, so $\hat{y}(i) = y(i)$ (since (5.6) implies that if $\hat{y}(i) = y(i) - 1$ then $\hat{y}(i) \geq y(n)$). Thus

$$[\![\hat{y}(i), a]\!] = [\![y(i), a]\!] \subseteq [\![y(i), x(i)]\!] \subseteq y\,[\![i, n]\!]$$

because $a \leq \hat{x}(i) \leq x(i)$, and by (5.7a). Therefore $a = y(k) = \hat{y}(k)$ for some $k \in [\![i, n-1]\!]$.

Case 1b: $a \geq y(n)$. Then, since $\hat{y}(i) \geq y(i) - 1$ and $x(i) \geq \hat{x}(i) \geq y(n)$, $a \in [\![\hat{y}(i), \hat{x}(i)]\!]$ implies that $a \in [\![y(i) - 1, x(i) - 1]\!]$, i.e., $y(i) \leq a + 1 \leq x(i)$. By (5.7a) there is some $k \in [\![i, n]\!]$ such that $a + 1 = y(k)$. In fact $k \neq n$ (since $a + 1 > y(n)$), so $\hat{y}(k) = y(k) - 1 = a$ and so $a \in \hat{y}\,[\![i, n-1]\!]$.

In both cases we have proved (5.7b).

($\Longleftarrow$) Assume that (5.7b) holds. It is immediate that (5.7a) holds when $i = n$, so fix $i \in [\![n-1]\!]$ and $a \in [\![y(i), x(i)]\!]$. We wish to show that $a = y(k)$ for some $k \in [\![i, n]\!]$. This is clear if $a = y(n)$, so assume $a \neq y(n)$.

Case 2a: $a < y(n)$. Since $a \in [\![y(i), x(i)]\!]$, either $a = x(i)$ or $a < x(i)$. If $a = x(i)$, then $a = x(i) = \hat{x}(i)$. If $a < x(i)$, then $a \leq \hat{x}(i)$ since $\hat{x}(i) \geq x(i) - 1$. In either case,

$$[\![y(i), a]\!] = [\![\hat{y}(i), a]\!] \subseteq [\![\hat{y}(i), \hat{x}(i)]\!] \subseteq \hat{y}\,[\![i, n-1]\!]\,.$$

Thus $a = \hat{y}(k) = y(k)$ for some $k \in [\![i, n-1]\!]$.

Case 2b: $a > y(n)$. Since $a \in [\![y(i), x(i)]\!]$, either $a = y(i)$ or $a > y(i)$. If $a = y(i)$, then $a - 1 = y(i) - 1 = \hat{y}(i)$ since $y(i) > y(n)$. If $a > y(i)$, then we know that $a - 1 \geq \hat{y}(i)$ since $y(i) \geq \hat{y}(i)$. It follows that $a - 1 \in [\![\hat{y}(i), \hat{x}(i)]\!]$, so, by (5.7b), there is some $k \in [\![i, n-1]\!]$ such that $a - 1 = \hat{y}(k) \geq y(n)$. Therefore, $a = y(k)$.

In both cases we have proved (5.7a). ■

Corollary 5.7. Let $x, y \in \mathfrak{S}_n$ with $x(n) = y(n)$. Then $x \rhd_R y$ if and only if $\hat{x} \rhd_R \hat{y}$.

Proof. The definition of $\hat{x}$ implies that

$$\ell(\hat{x}) = \ell(x) - (n - x(n)), \tag{5.8}$$

which together with Lemma 5.6 produces the desired result. ■

Proposition 5.8. Let $x, y \in \mathfrak{S}_n$ such that $x \trianglerighteq_R y$, and let $m = \ell(x) - \ell(y)$. Then there exists a chain

$$y = x_0 \lhd_R x_1 \lhd_R \cdots \lhd_R x_m = x. \tag{5.9}$$

Proof. The proof proceeds by double induction on n and m. The conclusion is trivial when $n \leq 2$ or $m \leq 1$. Accordingly, let $n > 2$ and $m > 1$, and assume inductively that the theorem holds for all $(n', m') <_C (n, m)$.

First, suppose that $x(n) = y(n)$. Then $\hat{x} \trianglerighteq_R \hat{y}$ by Lemma 5.6 where $\hat{x}, \hat{y}$ are defined by (5.6). Moreover, $\ell(\hat{x}) - \ell(\hat{y}) = \ell(x) - \ell(y) = m$ by (5.8). Therefore, by the induction hypothesis, there is a chain $\hat{y} = \hat{x}_0 \lhd_R \hat{x}_1 \lhd_R \cdots \lhd_R \hat{x}_m = \hat{x}$ in $\mathfrak{S}_{n-1}$, which by Corollary 5.7 can be lifted to a chain of the form (5.9).

Second, suppose that $x(n) \neq y(n)$. Since $x \geq_B y$ by Theorem 5.1, in fact $x(n) < y(n)$ (as noted in the proof of Proposition 5.4(c)). Let $p = y(n) - 1$; then $p \in [\![1, n-1]\!]$, so we may set $q = y^{-1}(p)$ and $z = s_p y = y t_{q,n}$. Then $z \gtrdot_W y$ and so $z \rhd_R y$ by Proposition 5.5. We will show that $x \trianglerighteq_R z$ using (*RC*).

Case 1: $1 \leq i \leq q$. Then $[\![z(i), x(i)]\!] \subseteq [\![y[i], x(i)]\!]$ and $y\,[\![i, n]\!] = z\,[\![i, n]\!]$, so $\mathsf{d}_n(x, y) \geq 1$ implies $\mathsf{d}_n(x, z) \geq 1$.

Case 2: $q < i < n$. Then $p = y(q) \notin y\,[\![i, n]\!]$, so by (*RC*) $p \notin [\![y(i), x(i)]\!]$. Thus $p + 1 \notin [\![y(i) + 1, x(i) + 1]\!]$, and certainly $p + 1 = y(n) \neq y(i)$. Thus $[\![y(i), x(i)]\!] \subseteq y\,[\![i, n]\!] \setminus \{y(n)\} = y\,[\![i, n-1]\!]$ and

$$[\![z(i), x(i)]\!] = [\![y(i), x(i)]\!] \subseteq y\,[\![i, n-1]\!] = z\,[\![i, n-1]\!] \subseteq z\,[\![i, n]\!]$$

so again $\mathsf{d}_n(x, z) \geq 1$.

Case 3: $i = n$. Then $x(n) \leq y(n) - 1 = z(n)$, so $\mathsf{d}_n(x, z) \geq 1$ by Proposition 5.4(b).

Taken together, the three cases imply $x \trianglerighteq_R z$. By induction there is a chain $z = x_1 \lhd_R \cdots \lhd_R x_m = x$, and appending $y = x_0$ produces a chain of the form (5.9). ∎

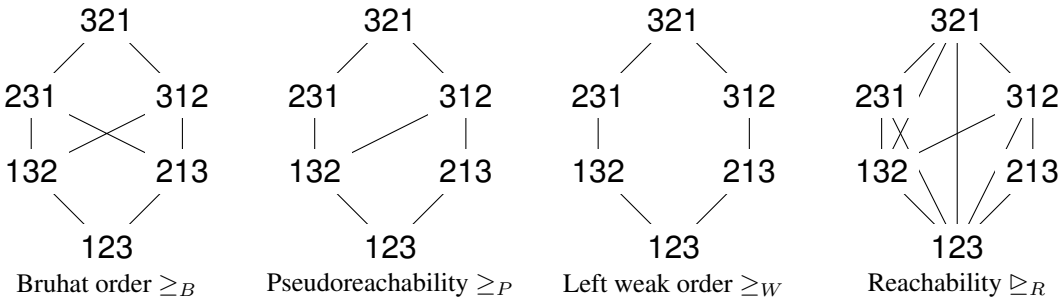

Figure 5.1: Bruhat, pseudoreachability, left weak order, and reachability on $\mathfrak{S}_3$.

Theorem 5.9. Pseudoreachability order is graded by length.

Proof. The definition of pseudoreachability as the transitive closure of reachability order implies that if $x_0 <_P \cdots <_P x_m$ is a maximal chain, then in fact each $x_{i-1} \trianglelefteq_R x_i$ for all i. Now, maximality together with Proposition 5.8 implies in turn that in fact $x_{i-1} \triangleleft_R x_i$. ■

For comparison, the Hasse diagrams of Bruhat, pseudoreachability, and left weak orders on $\mathfrak{S}_3$ are shown in Figure 5.1, together with the reachability relation (which is reflexive and antisymmetric, but not transitive). The three partial orders on $\mathfrak{S}_4$ are shown in Figure 5.2.

5.4 Pseudoreachability coincides with Armstrong's sorting order

The theory of normal forms in a Coxeter system was introduced by du Cloux [30] and is described in [13, §3.4]. We sketch here the facts we will need; see especially [13, Example 3.4.3], which describes normal forms in the symmetric group in terms of bubble-sorting. Let $\sigma_k = s_1 \cdots s_k$ and $\omega_n = \sigma_{n-1} \cdots \sigma_1$; then ω_n is a reduced word for $w_0 \in \mathfrak{S}_n$. Every $x \in \mathfrak{S}_n$ has a unique *conormal form*: a reduced word $N(w)$ of the form $v_{n-1}v_{n-2} \cdots v_2 v_1$, where $v_k = s_j s_{j+1} \cdots s_k$ is a suffix of σ_k. The conormal form is the reverse of the lexicographically first reduced word for x^{-1} (that is, of the normal form of

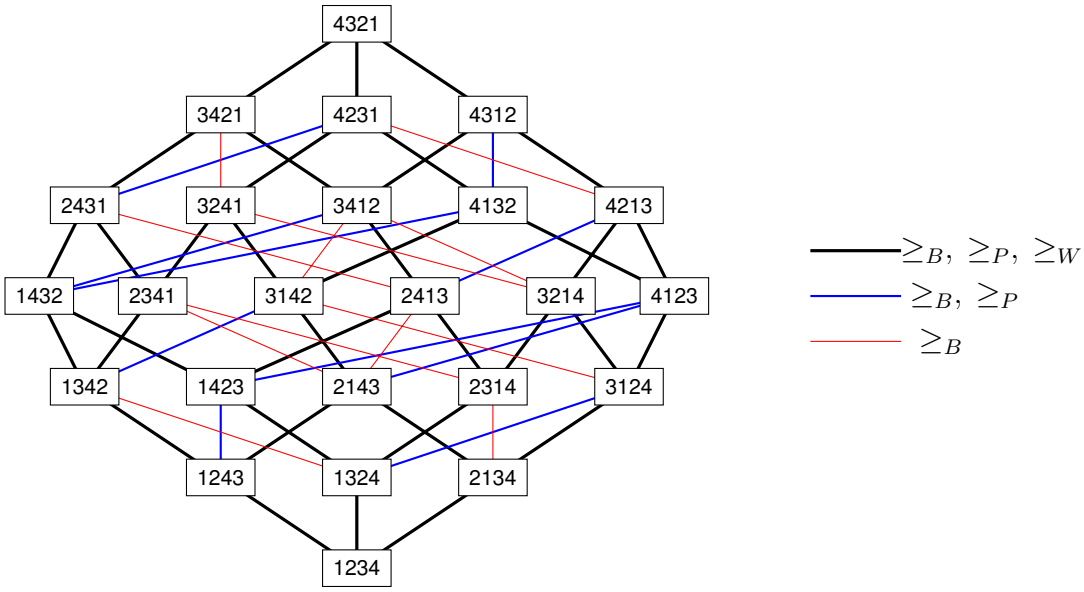

Figure 5.2: Bruhat, pseudoreachability, and left weak order on $\mathfrak{S}_4$.

x^{-1}, as described in [13]). Thus x is characterized by the sequence

$$\lambda(x) = (\lambda_{n-1}(x), \ldots, \lambda_1(x)) = (|v_{n-1}|, \ldots, |v_1|) \in [\![0, n-1]\!] \times [\![0, n-2]\!] \times \cdots \times [\![0, 1]\!] .$$

We now give an example of building the conormal form of a permutation.

Example 5.2. Let $x = 43521$. First find $x^{-1} = 54213$. Then begin by recording the sequence of transpositions to place 5 in the fifth place for x^{-1}. Then do the same for 4 into the fourth place, and so on. Table 5.2 illustrates the process to find that the conormal form of the permutation x is $s_1 s_2 s_3 s_4 s_1 s_2 s_3 s_1$ and $\lambda(x) = (4, 3, 0, 1)$.

Armstrong [9] defined a general class of *sorting orders* on a Coxeter system (W, S): one fixes $w \in W$ and chooses a reduced word ω (the "sorting word") for $w \in W$, then partially orders all group elements expressible as a subword of ω by inclusion between their lexicographically first such expressions. Armstrong proved that for every reduced word for the top element of a finite Coxeter group, the sorting order is a distributive lattice

Permutation	Transpositions to sort x^{-1}
54213	x^{-1}
45213	$x^{-1}s_1$
42513	$x^{-1}s_1s_2$
42153	$x^{-1}s_1s_2s_3$
42135	$x^{-1}s_1s_2s_3s_4$
24135	$x^{-1}s_1s_2s_3s_4s_1$
21435	$x^{-1}s_1s_2s_3s_4s_1s_2$
21345	$x^{-1}s_1s_2s_3s_4s_1s_2s_3$
12345	$x^{-1}s_1s_2s_3s_4s_1s_2s_3s_1$

Table 5.2: Example of the permutation $x = 43521$ and the calculation of its conormal form.

intermediate between the weak and Bruhat orders. In the case that $W = \mathfrak{S}_n$ and $\omega = \omega_n$, the sorting order is equivalent to comparing $\lambda(x)$ and $\lambda(y)$ componentwise, hence is isomorphic to $C_2 \times \cdots \times C_n$, where C_i denotes a chain with i elements.

Remark 7. Given a permutation x and its conormal form $\lambda(x) = (|w_{n-1}|, \ldots, |w_1|)$, it may be convenient to think of $\sum |w_i| = |w_1| + \cdots + |w_{n-1}|$ as a measure of how sorted the permutation is. A value of $\sum |w_i|$ close to 0 indicates that the permutation is close to sorted, whereas a value close to $\binom{n}{2}$ indicates that the permutation is very disordered. Furthermore, for two permutations x and y, it may be intuitive to think that $\lambda(y) \leq \lambda(x)$ implies that y can be sorted with a certain subset of the moves that were used to sort x. One could say then that y is "more sorted" than x.

Lemma 5.10. Let $x, y \in \mathfrak{S}_n$ with $x(n) = y(n) = n$, and let $v = s_j s_{j+1} \cdots s_{n-1}$ be a suffix of $s_1 \ \cdots \ s_{n-1}$. Then $x \trianglerighteq_R y$ if and only $vx \trianglerighteq_R vy$.

Proof. If $v = e$, there is nothing to prove. Otherwise, by (*RC*), it suffices to show that for every $i \in [\![n]\!]$, we have

$$[\![y(i), x(i)]\!] \subseteq y\,[\![i, n]\!] \tag{5.10a}$$

if and only if

$$[\![(vy)(i), (vx)(i)]\!] \subseteq vy\,[\![i, n]\!]\,. \tag{5.10b}$$

This is clear if $i = n$, so we assume henceforth that $i \neq n$. Moreover,

$$v(k) = \begin{cases} k & \text{if } k < j, \\ k+1 & \text{if } j \leq k < n, \\ j & \text{if } k = n \end{cases} \qquad \text{and} \qquad v^{-1}(k) = \begin{cases} k & \text{if } k < j, \\ n & \text{if } k = j, \\ k-1 & \text{if } k > j. \end{cases}$$

In particular, if $i \neq n$, then $x(i) > y(i)$ if and only if $v(x(i)) > v(y(i))$. We assume from now on that these two equivalent conditions hold, since, if both fail, then (5.10a) and (5.10b) are both trivially true. The proofs of the two directions now proceed very similarly.

(5.10a) $\implies$ (5.10b): There are three cases.

Case 1a: $j > x(i)$. Then v fixes $[\![1, x(i)]\!]$ pointwise, so

$$[\![(vy)(i), (vx)(i)]\!] = v\,[\![y(i), x(i)]\!] \subseteq vy\,[\![i, n]\!]$$

by applying v to both sides of (5.10a).

Case 1b: $y(i) < j \leq x(i)$. Then $(vx)(i) = x(i) + 1$ and $(vy)(i) = y(i)$, so

$$\begin{aligned} [\![(vy)(i), (vx)(i)]\!] &= [\![y(i), j-1]\!] \cup \{j\} \cup [\![j+1, x(i)+1]\!] \\ &= v\,[\![y(i), j-1]\!] \cup \{v(n)\} \cup v\,[\![j, x(i)]\!] \\ &= v\,([\![y(i), x(i)]\!] \cup \{y(n)\}) \\ &\subseteq vy\,[\![i, n]\!] \end{aligned}$$

establishing (5.10b).

Case 1c: $j \leq y(i)$. Similarly to Case 1a, we have

$$[\![(vy)(i), (vx)(i)]\!] = [\![y(i)+1, x(i)+1]\!] = v\,[\![y(i), x(i)]\!] \subseteq vy\,[\![i, n]\!]\,,$$

as desired.

(5.10b) $\implies$ (5.10a): Applying v^{-1} to both sides of (5.10b) gives

$$v^{-1}\,[\![vy(i), vx(i)]\!] \subseteq y\,[\![i, n]\!]\,,$$

so in order to prove (5.10a) It is enough to show that

$$[\![y(i), x(i)]\!] \subseteq v^{-1}\,[\![vy(i), vx(i)]\!] \tag{5.11}$$

Moreover, the earlier assumption $i \neq n$ implies that $vx(i) \neq j$ and $vy(i) \neq j$.

Case 2a: $j > vx(i)$. Then v^{-1} fixes the set $[\![1, vx(i)]\!]$ pointwise, so in particular $[\![y(i), x(i)]\!] = [\![vy(i), vx(i)]\!] = v^{-1}\,[\![vy(i), vx(i)]\!]$, establishing (5.11).

Case 2b: $vy(i) < j < vx(i)$. Then $y(i) = vy(i)$ and $x(i) = vx(i) - 1$, so

$$\begin{aligned}
[\![y(i), x(i)]\!] &= [\![vy(i), j-1]\!] \cup [\![j, vx(i)-1]\!] \\
&= v^{-1}\,[\![vy(i), vy(n)-1]\!] \cup v^{-1}\,[\![vy(n)+1, vx(i)]\!] \\
&\subseteq v^{-1}\,[\![vy(i), vx(i)]\!]\,.
\end{aligned}$$

Case 2c: $j < vy(i)$. Then $[\![y(i), x(i)]\!] = [\![vy(i)-1, vx(i)-1]\!] = v^{-1}\,[\![vy(i), vx(i)]\!]$, again implying (5.11). ■

Theorem 5.11. The pseudoreachability order coincides with the bubble-sort order.

Proof. It suffices to show that the two partial orders have the same covering relations, i.e., that

$$x \succcurlyeq_P y \quad \iff \quad \lambda(x) \succcurlyeq_C \lambda(y).$$

We induct on n; the base case $n = 1$ is trivial. Let $x, y \in \mathfrak{S}_n$ with $n > 1$, and let their conormal forms be

$$x = u\bar{x} = (s_i \cdots s_{n-1})\bar{x}, \qquad y = v\bar{y} = (s_j \cdots s_{n-1})\bar{y}$$

where $i = x(n) = n - \lambda_{n-1}(x)$ and $j = y(n) = n - \lambda_{n-1}(y)$.

($\Longleftarrow$) Suppose that $\lambda(x) \succcurlyeq_C \lambda(y)$. Then either $i = j - 1$ or $i = j$. If $i = j - 1$, then $\lambda(\bar{x}) = \lambda(\bar{y})$, so $\bar{x} = \bar{y}$ and $x = s_i y$, which by Proposition 5.5 implies $x \succcurlyeq_P y$. If $i = j$, then $\lambda(\bar{x}) \succcurlyeq_C \lambda(\bar{y})$. Then $\bar{x} \succcurlyeq_P \bar{y}$ by induction, so $x = v\bar{x} \succcurlyeq_P v\bar{y} = y$ by Lemma 5.10.

($\Longrightarrow$) Suppose that $x \succcurlyeq_P y$. Then $x \succcurlyeq_B y$ by Theorem 5.1, so $i \leq j$ (as noted in the proof of Proposition 5.4).

If $i < j$, then v is a proper suffix of u. By the definition of Bruhat order it must be the case that $x = yt_{a,b}$ for some $a < b$; in fact $b = n$ (otherwise $x(n) = y(n)$). Then $x(n) = y(a)$ and $x(a) = y(n)$, and $x(k) = y(k)$ for $k \notin \{a, n\}$. Moreover, $y(a) < x(a)$ (since $x \succcurlyeq_B y$ and not vice versa). On the other hand, if $y(a) \leq x(a) - 2$, so that $y(a) < c < x(a) = y(n)$ for some c, then by (*RC*) $c = y(k)$ for some $k \in [\![a + 1, n - 1]\!]$, and in particular x has at least three more inversions than y — not only (a, n), but also (a, k) and (k, n), which contradicts the assumption $x \succcurlyeq_P y$. Therefore $y(a) = x(a) - 1$, i.e., $x(n) = y(n) - 1$. We conclude that $x = s_i y$, so $\lambda(x) \succcurlyeq_C \lambda(y)$ using the conormal forms above.

If $i = j$, then $u = v$, so $\bar{x} \succcurlyeq_P \bar{y}$ by Prop. 5.10. By induction $\lambda(\bar{x}) \succcurlyeq_C \lambda(\bar{y})$, and prepending $n - i$ gives $\lambda(x) \succcurlyeq_C \lambda(y)$ as well. ■

5.5 Reachability via pattern avoidance

In this section, we establish two sufficient conditions for reachability using pattern avoidance. Let $\pi \in \mathfrak{S}_n$ and $\sigma \in \mathfrak{S}_m$, where $m \leq n$. A *σ-pattern* is a subsequence $\pi(i_1), \ldots, \pi(i_m)$ in the same relative order as σ, i.e., such that $1 \leq i_1 < \cdots < i_m \leq n$ and $\pi(i_j) < \pi(i_k)$ if and only if $\sigma(j) < \sigma(k)$. If π contains no σ-pattern then we say that π *avoids* σ.

Theorem 5.12. If $x \geq_B y$ and y avoids 213, then $x \trianglerighteq_R y$.

Proof. Suppose that $x \geq_B y$ and y avoids 213, but $x \not\trianglerighteq_R y$. Let i be any index such that $\mathsf{d}_i(x, y) = 0$. By Proposition 5.4 we know that $1 < i < n$ and that $y(i) < x(i)$. In particular, $m \neq i$, where $m = y^{-1}(x(i))$; that is, $y(m) = x(i)$.

First, suppose that $m > i$. We claim that there exists some $u < i$ such that $y(i) < y(u) < y(m)$. Otherwise, $J_i \geq y(m) - y(i) + 1$, and then $k = y(m) - y(i)$ has the properties $k < J_i$ and $y(i) + k = y(m) = x(i)$, so $k \in \mathsf{D}_i(x, y)$, contradicting the assumption $\mathsf{d}_i(x, y) = 0$. Therefore $y(u), y(i), y(m)$ is a 213-pattern.

Second, suppose that $m < i$. If $y(k) > y(m)$ for some $k > i$, then $y(m), y(i), y(k)$ is a 213-pattern. On the other hand, suppose that $y(k) < y(m) = x(i)$ for all $k > i$ (hence for all $k \geq i$). Then

$$\begin{aligned} \{k \in [\![i, n]\!] : \ y(k) < x(i)\} &= [\![i, n]\!] \\ &\supsetneq [\![i+1, n]\!] \\ &\supseteq \{k \in [\![i, n]\!] : \ x(k) < x(i)\} \subseteq [\![i+1, n]\!] \end{aligned}$$

so

$$|\{k \in [\![i, n]\!] : \ y(k) < x(i)\}| > |\{k \in [\![i, n]\!] : \ x(k) < x(i)\}|$$

$$\therefore \quad |\{k \in [\![1, i-1]\!] : \ y(k) < x(i)\}| < |\{k \in [\![1, i-1]\!] : \ x(k) < x(i)\}|$$

$$\therefore \quad |\{k \in [\![1, i-1]\!] : \ y(k) \geq x(i)\}| > |\{k \in [\![1, i-1]\!] : \ x(k) \geq x(i)\}| \,.$$

That is, $y\langle i-1, x(i)\rangle > x\langle i-1, x(i)\rangle$, contradicting the assumption $x \geq_B y$. ■

Theorem 5.12 partially answers the question of when the converse of Theorem 5.1 holds, i.e., which Bruhat relations are also relations in pseudoreachability order. We next study if there is an analogous condition on x, rather than y, that suffices for reachability. One such condition that allows us to restrict x instead of y is to ensure that only very few entries $x(i)$ are large with respect to i.

Lemma 5.13. Let $x \in \mathfrak{S}_n$. The following conditions are equivalent:

1. $x^{-1}(i) \leq i+1$ for all $i \in [\![n]\!]$.
2. $x\langle j, j\rangle = 1$ for all $j \in [\![n]\!]$.
3. x avoids both 231 and 321.
4. x is of the form $s_{i_1} \cdots s_{i_k}$, where $n-1 \geq i_1 > \cdots > i_k \geq 1$.

The number of these permutations is 2^{n-1}, which is easiest to see from condition (4). Conditions (1) and (3) were mentioned by J. Arndt (June 24, 2009) and M. Riehl (August 5, 2014) respectively in the comments on sequence A000079 in [64]. Accordingly, we will call a permutation satisfying the condition of Lemma 5.13 an *AR permutation* (for Arndt–Riehl).

Proof. $(1) \iff (2)$: Formula (4.3) implies that

$$\begin{aligned} \forall j \in [\![n]\!] : \ x\langle j, j\rangle = 1 &\iff \forall j \in [\![n]\!] : \ [\![1, j-1]\!] \subseteq x\,[\![1, j]\!] \\ &\iff \forall j \in [\![n]\!] : \ x^{-1}\,[\![1, j-1]\!] \subseteq [\![1, j]\!] \end{aligned}$$

$$\iff \forall i \in [\![n]\!] : \; x^{-1}[\![1,i]\!] \subseteq [\![1,i+1]\!]$$

since the last two statements differ only by the trivially true cases $i = 0$ and $i = n$.

(3) $\iff$ (1): Condition (3) holds if and only if no digit $i \in [\![n]\!]$ occurs later than position $i + 1$, but this is precisely condition (1).

(4) $\iff$ (1)/(3): Let Y_n be the set of permutations in $\mathfrak{S}_n$ satisfying the equivalent conditions (1) and (3), and let Z_n be the set satisfying condition (4). For $n \leq 2$ we evidently have $Y_n = Z_n = \mathfrak{S}_n$. For $n \geq 3$, we proceed by induction. Observe that $Z_n = Z_{n-1} \cup s_{n-1}Z_{n-1}$, and that left-multiplication by s_{n-1} (i.e., swapping the locations of $n-1$ and n) does not affect condition (1), which is always true for $i \in \{n-1, n\}$. Therefore $Z_n \subseteq Y_n$.

On the other hand, if $w \in Y_n$ then $w_n \in \{n-1, n\}$, otherwise w_n, together with the digits $n-1$ and n, would form a 231- or 321-pattern. Therefore, $w'_n = n$, where either $w' = w$ or $w' = s_{n-1}w$. By induction $w' \in Z_{n-1}$, so $w \in Z_n$ as desired. ■

Corollary 5.14. If x is AR and $y \leq_B x$, then y is AR as well.

Proof. Lemma 5.13 asserts that $x\langle i, i\rangle = 1$ for all $i \in [\![n]\!]$. Since $y \leq_B x$, $y\langle i, i\rangle = 1$ or 0, but the latter could not happen by the pigeonhole principle. ■

An *exceedance* of a permutation $x \in \mathfrak{S}_n$ is an index $k \in [\![n]\!]$ such that $x(k) > k$.

Lemma 5.15. Let $x \in \mathfrak{S}_n$ be an AR permutation. Suppose that k is an exceedance of x, and let $i = x(k)$. Then $x(j) = j - 1$ for all $j \in [\![k+1, i]\!]$.

Proof. The argument of Lemma 5.13 implies that $[\![1, k-1]\!] \subseteq x[\![1, k]\!]$; however, since $x(k) > k$ we have in fact $[\![1, k-1]\!] = x[\![1, k-1]\!]$.

Now let $j \in [\![k+1, i]\!]$. Lemma 5.13 also asserts that $x\langle j, j\rangle = \#A_j = 1$, where $A_j = \{m \in [\![j]\!] : \; x(m) \geq j\}$. Certainly $k \in A_j$, so $j \notin A_j$, that is, $x(j) < j$. But since $x(j) \geq k$ for each such j, we can infer in turn that $x(k+1) = k$, $x(k+2) = k+1$, …, $x(i) = i - 1$. ■

Theorem 5.16. If $x \geq_B y$ and x is AR, then $x \trianglerighteq_R y$.

Proof. Suppose that $x \geq_B y$ and x is AR, but $x \not\trianglerighteq_R y$. Let i be some index such that $\mathsf{d}_i(x, y) = 0$. By (*RC*), there exists $j < i$ such that

$$y(i) < y(j) \leq x(i). \tag{5.12}$$

By Lemma 5.13, $x\langle i, i\rangle = 1$; that is, there exists some (unique) $k \leq i$ such that $x(k) \geq i$. First, suppose that $k = i$. Then $x\langle i-1, i\rangle = 0$, and $y\langle i-1, i\rangle = 0$ as well because $y \leq_B x$. Hence $y[\![1, i-1]\!] = [\![i-1]\!]$. But then (5.12) implies that $y(i) < y(j) \leq i-1$ as well, a contradiction. Second, suppose that $k < i$. Then $y(i) < x(i) < i$ by Lemma 5.15, so $y(i) \leq i-2$. Set $k = y(i)$; then $y^{-1}(k) = i \geq k+2$. But then y is not AR, which violates Corollary 5.14. ■

5.6 Counting reachable pairs and open questions

Let $r(n) = |\{(x, y) \in \mathfrak{S}_n \times \mathfrak{S}_n : x \trianglerighteq_R y\}|$ be the number of reachable pairs in $\mathfrak{S}_n$. Explicit computation (using Python) reveals that the sequence $r(1), r(2), \ldots$ begins

$$1,\ 3,\ 17,\ 151,\ 1901,\ 31851,\ 680265,\ 17947631,\ \ldots$$

which matches OEIS sequence A145081. Accordingly, we conjecture that this sequence gives the values of $r(n)$ for all integers n. The OEIS entry does not give a combinatorial interpretation for this sequence; rather, the description is as follows. Consider a family of power series $F(t, x)$ for $t = 0, 1, 2, \ldots$ that satisfy $F(t, 0) = 1$ and

$$F_x(t, x) = tF(t, x)F(t+1, x), \tag{5.13}$$

where $F_x = \frac{d}{dx}(F)$. We note that the OEIS also gives two other equivalent recurrences in the form of integral equations. If we interpret $F(t,x)$ as an exponential generating function

$$F(t,x) = \sum_{n=0}^{\infty} R_n(t)\frac{x^n}{n!}$$

then the functional recurrence (5.13) can be transformed into the recurrence

$$R_0(t) = 1, \qquad R_{n+1}(t) = t\sum_{i=0}^{n}\binom{n}{i}R_i(t)R_{n-i}(t+1) \tag{5.14}$$

which is convenient for explicit calculation (in particular, the $R_n(t)$ are polynomials). The table of values for $R_n(t)$ for $t = 1, 2, 3, \ldots$ and $n = 0, 1, 2, \ldots$ is given by OEIS sequence A145080. The sequence $R_0(1), R_1(1), R_2(1), \ldots$ is OEIS A145081. The first several of the polynomials $R_n(t)$ are as follows.

$$\begin{aligned}
R_0(t) &= 1,\\
R_1(t) &= t,\\
R_2(t) &= 2t^2 + t,\\
R_3(t) &= 6t^3 + 8t^2 + 3t,\\
R_4(t) &= 24t^4 + 58t^3 + 52t^2 + 17t,\\
R_5(t) &= 120t^5 + 444t^4 + 680t^3 + 506t^2 + 151t.
\end{aligned}$$

Conjecture 5.1. For all $n \geq 1$ we have $r(n) = R_n(1)$.

In order to use this recurrence to prove the conjecture, it appears necessary to either find a combinatorial interpretation for the entire table of numbers $R_n(t)$, or transform (5.14) into a single-term recurrence for $R_n(1)$.

Problem 5.1. Find a statistic f on reachable pairs of permutations in $\mathfrak{S}_n$ such that

$$R_n(t) = \sum_{x \trianglerighteq_R y} t^{f(x,y)}.$$

The form of (5.14) somewhat resembles other recurrences that arise in the theory of parking functions; see, e.g., [29, Corollary 1] and [2, Corollary 2], both of which count the number of parking functions in which certain entries are specified in advance.

We now state and prove closed-form descriptions for some of the extreme coefficients of the polynomials $R_n(t)$ and an identity for the value of $R_n(-1)$.

Proposition 5.17. Let $[t^m]R_n(t)$ denote the coefficient of t^m in $R_n(t)$. Then

1. $R_n(-1) = (-1)^n$ for $n \geq 0$.
2. $[t^n]R_n(t) = n!$ for $n \geq 1$.
3. $[t^1]R_n(t) = R_{n-1}(1)$ for $n \geq 1$.
4. $R_n(0) = 0$ for $n \geq 1$.

Proof. We will prove (1) and (2) by induction on n, while (3) and (4) follow by direct computation. Note that $\deg(R_n(t)) = n$.

1. Since $F(t, 0) = 1$, it follows that $R_0(t) = 1$, so $R_0(-1) = 1$. Assume inductively that $R_{n-1}(-1) = (-1)^{n-1}$. Using the recurrence (5.14), $R_0(0) = 1$ and $R_m(0) = 0$ for all $m \geq 1$. Therefore, again using (5.14),

$$R_n(-1) = -\sum_{i=0}^{n-1} \binom{n-1}{i} R_i(-1) R_{n-i-1}(0) = -R_{n-1}(-1) = (-1)^n.$$

2. From (1), $[t]R_1(t) = 1$. Assume inductively that $[t^m]R_m(t) = m!$ for all $m < n$, then by (5.14),

$$\begin{aligned}[t^n]R_n(t) &= \sum_{i=0}^{n-1} \binom{n-1}{i} [t^{n-1}](R_i(t)R_{n-i-1}(t+1)) \\ &= \sum_{i=0}^{n-1} \binom{n-1}{i} i!(n-i-1)! = n!\end{aligned}$$

3. Since $R_1(t) = t$, $[t]R_1(t) = 1 = R_0(1)$. More generally, using recurrence (5.14),

$$[t]R_n(t) = \sum_{i=0}^{n-1} \binom{n-1}{i} [t^0](R_i(t)R_{n-i-1}(t+1)) = [t^0]R_{n-1}(t+1) = R_{n-1}(1).$$

The second to last equality follows since t is a factor of $R_i(t)$ for $i > 1$ and the last equality follows since the constant term in $R_{n-1}(t+1)$ is the sum of the coefficients in $R_{n-1}(t)$.

4. This is immediate from (5.14). ■

In addition, the second-leading coefficients of $R_n(t)$ appear to have the following combinatorial interpretation. The *Eulerian number of the second kind* $\left\langle\!\left\langle {n \atop k} \right\rangle\!\right\rangle$ is the number of rearrangements of $(1, 1, 2, 2, \ldots, n, n)$ with k ascents such that every number between the two occurrences of m is less than m, for all $m \in [\![n]\!]$.

Conjecture 5.2. For every integer n, we have $[t^{n-1}]R_n(t) = \left\langle\!\left\langle {n+1 \atop n-1} \right\rangle\!\right\rangle$ (OEIS sequence A002538).

We have verified this conjecture computationally for $n \leq 100$. Given the relationship between reachability and Bruhat order, we find it intriguing that the numbers $\left\langle\!\left\langle {n+1 \atop n-1} \right\rangle\!\right\rangle$ also count the edges (i.e., covering relations) in Bruhat order on $\mathfrak{S}_{n+1}$.

We briefly present another possible direction of investigation that relates the problem of counting reachable pairs of bioutcomes of interval parking functions to count-

ing ordinary parking functions. We note that this line of investigation was suggested through communications between Jeremy L. Martin and Mei Yin with Richard Stanley. Let $H(x) = xF(t,x)$, regarded as a power series in x with coefficients in $\mathbb{C}[t]$, and let $G(x)$ be the compositional inverse of $H(x)$ (see [67, §5.4]). Expanding $G(x)$ as an exponential generating function

$$G(x) = \sum_{n=0}^{\infty} S_n(t) \frac{x^{n+1}}{n!} \tag{5.15}$$

it appears that $S_n(t)$ is a polynomial of degree n in t, with integer coefficients that alternate in sign. The first several of these polynomials are as follows.

$$\begin{aligned}
S_0(t) &= 1, \\
S_1(t) &= -t, \\
S_2(t) &= 2t^2 - t, \\
S_3(t) &= -6t^3 + 7t^2 - 3t, \\
S_4(t) &= 24t^4 - 46t^3 + 38t^2 - 17t, \\
S_5(t) &= -120t^5 + 326t^4 - 400t^3 + 299t^2 - 151t.
\end{aligned}$$

By Proposition 5.17 (1), $R_n(-1) = (-1)^n$, so $H(x)|_{t=-1} = xe^{-x}$. Since the compositional inverse of xe^{-x} is $\sum_{n\geq 1} n^{n-1}x^n/n!$ [67, Example 5.4.4, page 43], it follows easily from (5.15) that $S_n(-1) = (n+1)^{n-1}$, the number of parking functions of length n. The extreme coefficients are familiar: $[t^n]S_n(-t) = n!$ and $[t]S_n(-t) = r(n-1)$. Furthermore, the coefficient of t^{n-1} in $S_n(-t)$ appears to match OEIS sequence A067318.

Problem 5.2. Find a statistic g on parking functions of length n such that

$$S_n(-t) = \sum_{p \in \mathsf{PF}_n} t^{g(p)}.$$

Of course, parking functions might be replaced with any of the various other combinatorial objects enumerated by $(n+1)^{n-1}$, such as labelled trees on $n+1$ vertices.

Lastly, we ask the following about reachability and pattern avoidance.

Problem 5.3. Can reachability of the pair $x \trianglerighteq_R y$ be fully characterized in terms of pattern avoidance conditions coupled with the assumption that $x \geq_B y$?

BIBLIOGRAPHY

[1] E. Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18:1–86, 2018.

[2] A. Adeniran, S. Butler, G. Dorpalen-Barry, P.E. Harris, C. Hettle, Q. Liang, J.L. Martin, and H. Nam. Enumerating parking completions using Join and Split. *Electron. J. Combin.*, 27(2):Paper #P2.44, 19 pp., 2020.

[3] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

[4] D. Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11:581–598, 1981.

[5] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs, 2009.

[6] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electronic Journal of Probability*, 12:1454–1508, 2007.

[7] D. Aldous and J.M. Steele. *The objective method: Probabilistic combinatorial optimization and local weak convergence*. Springer, Berlin, 2004.

[8] D. Aristoff and L. Zhu. Asymptotic structure and singularities in constrained directed graphs. *Stochastic Process. Appl.*, 125:4154–4177, 2015.

[9] D. Armstrong. The sorting order on a Coxeter group. *J. Combin. Theory Ser. A*, 116(8):1285–1305, 2009.

[10] A.D. Barbour and A. Röllin. Central limit theorems in the configuration model. *arXiv: 1710.02644*, 2017.

[11] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electronic Journal of Probability*, 6:1–13, 2001.

[12] S. Bhamidi, G. Bresler, and A. Sly. Mixing time of exponential random graphs. *Ann. Appl. Probab.*, 21:2145–2170, 2011.

[13] A. Björner and F. Brenti. *Combinatorics of Coxeter groups*, volume 231 of *Graduate Texts in Mathematics*. Springer, New York, 2005.

[14] C. Borgs, J. Chayes, H. Cohn, and Y. Zhao. An l^p theory of sparse graph convergence i. limits, spars random graph models, and power law distributions. *arXiv:1401.2906*, 2014.

[15] C. Borgs, J. Chayes, H. Cohn, and Y. Zhao. An l^p theory of sparse graph convergence ii. ld convergence, quotients, and right convergence. *arXiv: 1408.0744*, 2014.

[16] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi. Counting graph homomorphisms. *Topics in Discrete Mathematics*, 26:315–371, 2006.

[17] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs i. subgraph frequencies, metric properties and testing. *Adv. Math.*, 219:1801–1851, 2008.

[18] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs ii. multiway cuts and statistical physics. *Ann. of Math.*, 176:151–219, 2012.

[19] R. Bubley and M. Dyer. Path coupling: a technique for proving rapid mixing in markov chains. *FOCS*, pages 223–231, 1997.

[20] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396–450, 2016.

[21] S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41:2428–2461, 2013.

[22] S. Chatterjee and S.R.S. Varadhan. The large deviation principle for the erdos-renyi random graph. *European J. Combin.*, 32:1000–1017, 2011.

[23] E. Colaric, R. DeMuse, J. L. Martin, and M. Yin. Interval parking functions. *Adv. Appl. Math.*, 123:102129, 2021.

[24] S.J. Cranmer and B.A. Desmarais. Inferential network analysis with exponential random graph models. *Pol. Anal.*, 19:66–86, 2011.

[25] R. DeMuse, T. Easlick, and M. Yin. Mixing time of vertex-weighted exponential random graphs. *J. Comput. Appl. Math*, 362:443–459, 2019.

[26] R. DeMuse, D. Larcomb, and M. Yin. Phase transitions in edge-weighted exponential random graphs: near-degeneracy and universality. *J. Stat. Phys.*, 171:127–144, 2018.

[27] R. DeMuse and M. Yin. Dimension reduction in vertex-weighted exponential random graphs. *Physica A*, 561:125289, 2021.

[28] P. Diaconis and A. Hicks. Probabilizing parking functions. *Adv. Appl. Math.*, 89:125–155, 2017.

[29] P. Diaconis and A. Hicks. Probabilizing parking functions. *Adv. in Appl. Math.*, 89:125–155, 2017.

[30] F. du Cloux. A transducer approach to Coxeter groups. *J. Symbolic Comput.*, 27(3):311–324, 1999.

[31] R. Ehrenborg. Parking cars of different sizes. *Amer. Math. Monthly*, 123(10):1045–1048, 2016.

[32] R. Eldan and R. Gross. Exponential random graphs behave like mixtures of stochastic block models. *Ann. Appl. Probab.*, 28:3698–3735, 2018.

[33] P. Erdős and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[34] C.M. Fortuin and P.W. Kasteleyn. On the random cluster model i. introduction and relation to other models. *Physica*, 57:536–564, 1971.

[35] O. Frank and D. Strauss. Markov graphs. *J. Amer. Statist. Assoc.*, 81:832–842, 1986.

[36] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19:175–220, 1999.

[37] R.B. Griffiths, C.-Y Weng, and J.S. Langer. Relaxation times for metastable states in the mean-field model of a ferromagnet. *Phys. Rev.*, 149:301–305, 1966.

[38] D. Hoover. Row-column exchangeability and a generalized model for probability. In G. Koch and F. Spizzichino, editors, *Exchangeability in Probability and Statistics*, pages 281–291. Elsevier, 1982.

[39] C. Jiao, T. Wang, J. Liu, H. Wu, F. Cui, and X. Peng. Using exponential random graph models to analyze the character of peer relationship networks and their effects on the subjective well-being of adolescents. *Front. Psychol.*, 8:583, 2017.

[40] J. Jonasson. The random triangle model. *J. Appl. Prob.*, 36:852–867, 1999.

[41] R. Kenyon, C. Radin, K. Ren, and L. Sadun. Multipodal structure and phase transitions in large constrained graphs. *arXiv: 1405.0599*, 2014.

[42] R. Kenyon and M. Yin. On the asymptotics of constrained exponential random graphs. *J. Appl. Probab.*, 54:165–180, 2017.

[43] A.G. Konheim and B. Weiss. An occupancy discipline and applications. *SIAM J. Appl. Math.*, 14(6):1266–1274, 1966.

[44] D.A. Levin, M.J. Luczak, and Y. Peres. Glauber dynamics for the mean-field ising model: cut-off, critical power law, and metastability. *Probab. Theory Relat. Fields*, 146:223–265, 2010.

[45] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, 2008.

[46] L. Lovász. *Large networks and graph limits*. AMS Colloquium Publications, Providence, 2012.

[47] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B.*, 96:933–957, 2006.

[48] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. *Random Structures and Algorithms*, 47:109–146, 2015.

[49] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. *Random Structures and Algorithms*, 50:420–436, 2017.

[50] R. Lyons. Asymptotic enumeration of spanning trees. *Combin. Probab. Comput.*, 14:491–522, 2005.

[51] C. Obando and F. De Vico Fallani. A statistical model for brain networks inferred from large-scale electrophysiological signals. *J. Royal Soc. Interface*, 14, 2017.

[52] J. Park and M. Newman. Solution of the two-star model of a network. *Phys. Rev. E*, 70:066146, 2004.

[53] J. Park and M. Newman. Solution for the properties of a clustered network. *Phys. Rev. E*, 72:026136, 2005.

[54] M. Penrose. *Random geometric graphs*. Oxford University Press, 2003.

[55] A. Postnikov and B. Shapiro. Trees, parking functions, syzygies, and deformations of monomial ideals. *Trans. Amer. Math. Soc.*, 356:3109–3142, 2004.

[56] R. Pyke. The supremum and infimum of the Poisson process. *Ann. Math. Statist.*, 30:568–576, 1959.

[57] C. Radin, K. Ren, and L. Sadun. The asymptotics of large constrained graphs. *J. Phys. A: Math. Theor.*, 47:175001, 2014.

[58] C. Radin and L. Sadun. Phase transitions in a complex network. *J. Phys. A: Math. Theor.*, 46:305002, 2013.

[59] C. Radin and L. Sadun. Singularities in the entropy of asymptotically large simple graphs. *J. Stat. Phys.*, 158:853–865, 2015.

[60] C. Radin and M. Yin. Phase transitions in exponential random graphs. *Ann. Appl. Probab.*, 23:2458–2471, 2013.

[61] F. Rassoul-Agha and T. Seppäläinen. *A course on large deviations with an introduction to Gibbs measures*, volume 162 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2015.

[62] J. Riordan. Ballots and trees. *J. Combinatorial Theory*, 6:408–411, 1969.

[63] G.L. Robins, P.E. Pattison, Y. Kalish, and D. Lusher. An introduction to exponention random graph p* models for social networks. *Soc. Netw.*, 29:173–191, 2007.

[64] N.J.A. Sloane. The On-Line Encyclopedia of Integer Sequences, 2020. Published electronically at https://oeis.org.

[65] R.P. Stanley. Hyperplane arrangements, interval orders, and trees. *Proc. Nat. Acad. Sci.*, 93(6):2620–2625, 1996.

[66] R.P. Stanley. Parking functions and noncrossing partitions. *Electron. J. Combin.*, 4, 1997.

[67] R.P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.

[68] S. Strogatz and D. Watts. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[69] A.C.D. van Enter, R. Fernández, F. den Hollander, and F. Redig. Possible loss and recovery of gibbsianness during the stochastic evolution of gibbs measures. *Commun. Math. Phys.*, 226:101–130, 2002.

[70] S. Wasserman and P. Pattison. Logit models and logistic regressions for social networks. *Psychometrika*, 61:401–425, 1996.

[71] C.H. Yan. Parking functions. In *Handbook of enumerative combinatorics*, Discrete Math. Appl. (Boca Raton), pages 835–893. CRC Press, Boca Raton, FL, 2015.

[72] M. Yin. Critical phenomena in exponential random graphs. *J. Stat. Phys.*, 153:1008–1021, 2013.

[73] M. Yin. Phase transitions in edge-weighted exponential random graphs. *arXiv: 1607.04084*, 2016.

[74] Y. Zhao. On the bruhat order of the symmetric group and its shellability. Unpublished paper, 2007.

[75] R.K.P. Zia, E.F. Redish, and S. McKay. Making sense of the legendre transform. *Amer. J. Phys.*, 77:614–622, 2009.