Title
Contributions to Stein's method and some limit theorems in probability

Author
Dey, Partha Sarathi

Publication Date
2010

Peer reviewed|Thesis/dissertation

Contributions to Stein's method and some limit theorems in probability

by

Partha Sarathi Dey

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Sourav Chatterjee, Co-chair
Professor Steven N. Evans, Co-chair
Professor David J. Aldous
Professor Lawrence C. Evans

Spring 2010

Contributions to Stein's method and some limit theorems in probability

Abstract

Contributions to Stein's method and some limit theorems in probability

by

Partha Sarathi Dey
Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Sourav Chatterjee, Co-chair
Professor Steven N. Evans, Co-chair

In this dissertation we investigate three different problems related to (1) concentration inequalities using Stein's method of exchangeable pair, (2) first-passage percolation along thin lattice cylinders and (3) limiting spectral distribution of random linear combinations of projection matrices.

Stein's method is a semi-classical tool for establishing distributional convergence, particularly effective in problems involving dependent random variables. A version of Stein's method for concentration inequalities was introduced in the Ph.D. thesis of Sourav Chatterjee to prove concentration of measure in problems involving complex dependencies such as random permutations and Gibbs measures.

In the first part of the dissertation we provide some extensions of the theory and three new applications: (1) We obtain a concentration inequality for the magnetization in the Curie-Weiss model at critical temperature (where it obeys a non-standard normalization and super-Gaussian concentration). (2) We derive exact large deviation asymptotics for the number of triangles in the Erdős-Rényi random graph $G(n,p)$ when $p \geq 0.31$. Similar results are derived also for general subgraph counts. (3) We obtain some interesting concentration inequalities for the Ising model on lattices that hold at all temperatures.

In the second part, we consider first-passage percolation across thin cylinders of the form $[0,n] \times [-h_n, h_n]^{d-1}$. We prove that the first-passage times obey Gaussian central limit theorems as long as h_n grows slower than $n^{1/(d+1)}$. We obtain appropriate moment bounds and use decomposition of the first-passage time into an approximate sum of independent random variables and a renormalization type argument to prove the result. It is an open question as to what is the fastest that h_n can grow so that a Gaussian CLT still holds. We conjecture that $n^{2/3}$ is the right answer for $d = 2$ and provide some numerical evidence for that.

Finally, in the last part we consider limiting spectral distributions of random matrices of the form $\sum_{i=1}^{k} a_i X_i M_i$ where X_i's are i.i.d. mean zero and variance one random variables, a_i's are some given sequence of real numbers with ℓ^2 norm one and

M_i's are projection matrices with dimension growing to infinity. We provide sufficient conditions under which the limiting spectral distribution is Gaussian. We also provide examples from the theory of representations of symmetric group for which our results hold.

To my family: Thakuma, Ma, Baba and Mamani.

To all my teachers.

Contents

Acknowledgments

First of all, I would like to thank my two advisors: Sourav Chatterjee and Steve Evans for their encouragement, guidance and inspiration throughout my graduate life. Apart from being great teachers and providing invaluable inputs to my research, they also helped me grow as a person, enabling me to find a career path for myself. I learned a lot about life as much as I did about mathematics from them. Their teaching will stay with me for the rest of my career. In particular, I want to thank them for their support through a personal crisis. Also many thanks to Sourav for motivating me to come to USA for higher studies and for all the nice treats at Berkeley and back at Kolkata.

I would also like to thank Professors David Aldous and Craig Evans for agreeing to be in my dissertation committee and for numerous other things I learned from these two great mathematicians. I am grateful to all my teachers at Berkeley, especially Alice Guionnet, Elchanan Mossel, Jim Pitman and Alistair Sinclair, for their invaluable teachings which was extremely useful in my later research and for being generous to me with their advice and encouragement. Many other professors have provided me with valuable advice during my Ph.D., among them: Amir Dembo, Persi Diaconis and Yuval Peres. Thank you very much for dedicating your precious time to me. Besides that, I want to thank all other professors and staff in this small wonderful Statistics department, whose door was always open for me. I am grateful to Anindita Adhikari for her help and advice during the first year. Many thanks to Angie Fong for answering all my important and not-so-important non-academic questions.

It is imperative that I thank all my fellow graduate students and postdocs who were part of my life in the last four years and without whose presence the path would not have been so smooth. I am thankful to Tonci Auntonovich, Shankar Bhamidi, Nayantara Bhatnagar, Sharmodeep Bhattacharya, Nick Crawford, Jian Ding, Jonathan Grib, Subhrosekhar Ghosh, Chris Haulk, Richard Liang, Asaf Nachmias, Kaushik Pal, Peter Ralph, Sayak Ray, Nathan Ross, Arnab Sen, Allan Sly, Frances Tong and all other friends and colleagues for making the graduate life experience so enriching. Special thanks to my classmate and roommate, Arnab Sen for always being with me through the highs and lows, for cooking so many nice dishes and for numerous nice discussions about life and math.

They say that “life is partly what we make it, and partly what it is made by the friends we choose”. I feel really blessed to have so many good friends. Especially I would like to thank two other members of my group of “three idiots” back in India, Biplab Sarkar and Dipankar Deb, who were always beside me for the last twenty three years. It is you who taught me the essence of true friendship. I am also thankful to Shirshendu Chatterjee, Samriddhi Chatterjee, Roshni Ghosh, Subarata GuhaThakurta, Sirshendu Maity, Sayantani Sarkar for making life enjoyable and lively. Many thanks to Sayantani for being there to listen to my crazy thoughts and being supportive whenever I felt lonely.

I am grateful to all of my teachers at Indian Statistical Institute, Kolkata (especially S. C. Bagchi, Arup Bose, Jayanta Ghosh, Alok Goswami, Tapas Samanta and all others) for teaching me what statistics and probability is really about. I thank Professors Probal Chaudhuri and B. V. Rao for teaching me the proper meaning of research and for showing me the real beauty in Probability. I would never have reached this place without your support.

Thanks are due to my sister Maitrei Das, brother-in-law Tamal Das and my sweet little niece Tanirika Das for making every visit to India so memorable.

I would like to say thanks to the one person who was my "roommate" for twenty-three years before coming to Berkeley, with whom I took my first step to school, who taught me the meaning of hard-work and how to lead a simple life, who told me thousands of stories every night before going to sleep, who gave me shelter after I did all those mischiefs and who always waited for me when I came late at night. I dedicate this dissertation to my loving grandmother, Mrs. Usha Rani Dey. I love you Thakuma.

Last but not the least, I thank the two persons without whom I would never have been the person I am today. They are my parents: Rina Dey and Mrinal Kanti Dey, who gave me the freedom to explore new avenues, who encouraged me to think independently, who always supported me through my life, who gave away their happiness to help me achieve my dreams and specially whose endless love provides me the strong support to go forward. I am really proud to be your son.

Chapter 1

Introduction and review of literature

In his seminal 1972 paper [103], Charles Stein introduced a method for proving central limit theorems with convergence rates for sums of dependent random variables. This has now come to be known as *Stein's method*. Over the last four decades it has become a powerful tool in approximating probability distributions and proving limit theorems with quantitative rates of convergence. Though the method is very well-developed for convergence to Poisson and Gaussian distributions, it has also been applied to various other distributions, from hypergeometric to exponential. All the various formulations of the method rely on exploiting the characterizing operator or *Stein equation* of the distribution. We defer the discussion on Stein's method with examples until Section 1.2.

On the other hand, concentration inequalities involve "good" bounds on tail probabilities, *e.g.*, on $\mathbb{P}(|f(X) - \mathbb{E}(f(X))| \geq t)$ for $t > 0$ where the distribution of X is specified and f is a "nice" function. Here we call a bound "good" if it decays to zero rapidly. The simplest useful example being Chebyshev's inequality, $\mathbb{P}(|f(X) - \mathbb{E}(f(X))| \geq t) \leq t^{-2}\operatorname{Var}(f(X))$ for $t > 0$. In many cases, concentration bounds are precursor of distributional convergence results. In fact, tightness is an important factor for proving convergence of processes. For a long time, Azuma-Hoeffding inequality [56, 4] and its relatives (bounded difference inequality [97, 98], McDiarmid's inequality [84]) remained the best possible way to obtain Gaussian type decay $e^{-ct^2}, t \geq 0$, the main ingredient being Doob's decomposition into sums of martingale difference sequences (one can view the result as a precursor of the Gaussian central limit theorem). It was subsequently used in problems from statistics, computer science and other fields, in particular machine learning and empirical process theory. The most widely used form of Azuma-Hoeffding inequality states the following:

Theorem 1.0.1 (Azuma-Hoeffding inequality [56, 4]). *Let $\{X_i : 1 \leq i \leq n\}$ be a martingale difference sequence adapted to some filtration. Suppose that there exist* *nonnegative constants* $c_1, c_2, \ldots, c_n$ *such that* $|X_i| \le c_i$ *a.s. for each* i. *Then for all* $t \ge 0$ *we have*

$$\mathbb{P}\left(\max_{1\le k\le n}\sum_{i=1}^{k} X_i \ge t\right) \le \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

However in the late nineties, starting with Talagrand's subtle use of induction argument to get strong concentration bounds for functions on product measure spaces (see [106, 107, 108]), there have been much more activities in the field of concentration bounds with higher level of sophistication. In particular, the "entropy method" of Ledoux [75] and Massart [83] (log-Sobolev and modified log-Sobolev inequalities), exponential Efron-Stein inequalities of Boucheron, Lugosi and Massart in [22], transportation cost inequalities of Marton [80, 81, 82], information theoretic inequalities of Dembo [36] are now quite well used. Talagrand's *convex distance inequality* has found applications in fields as diverse as statistics, combinatorial optimization, random matrix, spin glasses and many more. Theorem 1.0.2 shows an important and useful corollary of the convex distance inequality.

Theorem 1.0.2 (Talagrand [106]). *For every product probability measure* μ^n *on* $[0,1]^n$, *every convex* 1*-Lipschitz function* f *on* $\mathbb{R}^n$, *and every nonnegative real number* t, *we have*

$$\mu^n\left(|f - m(f)| \ge t\right) \le 4e^{-t^2/4}$$

where $m(f)$ *is the median of* f *under* μ^n.

We refer the reader to the excellent survey by Ledoux [75] for more results about concentration inequalities. Here we mention that concentration inequalities have also been used to understand the geometry of high dimensional spaces and groups (See e.g. [86]) and it was one of the original motivation behind the initial investigation in concentration results. While for product measure spaces the general theory works surprisingly well, for random variables with complex dependency structure, in general, concentration bounds are hard to get. Many other approaches are available which work well on particular problems.

Stein's attempts [104] at devising a version of the method for concentration inequalities did not prove fruitful. Some progress for sums of dependent random variables was made by Raič [93]. The problem was finally solved in full generality in [24] using exchangeable pair approach. The general abstract result is stated in Section 2.1. A selection of results and examples from [24] appeared in the later papers [28, 27].

In Chapter 2 of this dissertation we extend the abstract theory and work out some further examples. We also look at two other problems from first-passage percolation on lattices and random matrix theory.

1.1 Summary of the Dissertation

We now give a brief chapter by chapter description of this dissertation in the subsequent subsections. To keep the exposition simple we will avoid the abstract results and only state the simplest versions of the theorems. The main chapters of this dissertation, Chapter 2, Chapter 3 and Chapter 4, are independent of each other and may be read in any order.

1.1.1 Concentration inequalities using exchangeable pairs

In Chapter 2 we derive extension of the concentration inequalities using exchangeable pair. We also work out three new examples using the method. Let us briefly describe the examples first.

The first example being large deviation inequalities for number of triangles in Erdős-Rényi random graph. Undoubtedly the most famous combinatorial model in probability is the Erdős-Rényi random graph model $G(n, p)$, which gives a random graph on n vertices where each edge is present with probability p and absent with probability $1 - p$ independently of each other. A triangle is a set of three vertices such that all the three edges are present in the random graph. The behavior of the upper tail of subgraph counts in $G(n, p)$ is a problem of great interest in the theory of random graphs (see [17, 60, 62, 110, 70], and references contained therein). However, it is an open problem to find exact form for the tail probability depending on n, p upto second order error terms. The best upper bounds to date were obtained only recently by Chatterjee [29] for triangles and Janson, Oleszkiewicz, and Ruciński [61] for general subgraph counts. For triangles, the available results state that for a fixed $\epsilon > 0$,

$$\mathbb{P}(T_n \geq (1 + \varepsilon)n^3p^3/6) = \exp(-\Theta(n^2p^2|\log p|))$$

where T_n is the number of triangles in $G(n, p)$.

Let us briefly look at the known results about tail bounds for general subgraph counts. Let F be a finite graph. Let us denote the number of edges in F by $e(F)$ and number of vertices by $e(G)$. The quantity of interest is $X_n(F)$, the number of copies of F in the Erdős-Rényi random graph $G(n, p)$. We need to define few quantities first before stating the results. Define

$$m(F) := \max\left\{\frac{e(H)}{v(H)} \mid H \subseteq F, v(H) > 0\right\}$$
$$\text{and } \Phi_n(F) := \min\left\{\mathbb{E}[X_n(H)] \mid H \subseteq F, e(H) > 0\right\}.$$

A graph F is called *balanced* if $m(F) = e(F)/v(F)$. The importance of $m(F)$ comes from the fact that

$$\mathrm{Var}(X_n(F)) \approx (1-p)\frac{\mathbb{E}[X_n(F)]^2}{\Phi_n(F)}$$

and $\Phi_n(F) \to \infty$ iff $np^{m(F)} \to \infty$. A result of Ruciński [96] states that $np^{m(F)} \to \infty$ and $n^2(1-p) \to 0$ as $n \to \infty$ is a necessary and sufficient condition for Gaussian CLT for normalized $X_n(F)$. The difficult part is to correctly bound the upper tail, since for the lower tail one can find a strong bound easily (see [60]). One can easily check using FKG inequality that the the bound is best possible as long as p stays away from one.

Theorem 1.1.1. *Let F be a fixed graph. Let $X_n(F)$ be the number of copies of F in the Erdős-Rényi random graph $G(n,p)$. Then for any $\varepsilon > 0$ we have*

$$\mathbb{P}(X_n(F) \le (1-\varepsilon)\,\mathbb{E}[X_n(F)]) \le \exp(-c(\varepsilon)\Phi_n(F))$$

for all n, p for some constant $c(\varepsilon) > 0$ depending on ε.

Now to state the results for upper tail bound for $X_n(F)$, we need two more quantities. For two graphs H, F define

$$N(F,H) := \text{ number of copies of } H \text{ in } F$$
$$\text{and } N(n,m,H) := \max\{N(F,H) \mid v(F) \le n, e(F) \le m\}.$$

Finally consider

$$M_F^*(n,p) := \begin{cases} \max\left\{m \mid \text{For all } H \subseteq F, N(n,m,H) \le n^{v(H)}p^{e(H)}\right\} & \text{if } p \ge n^{-2} \\ 1 & \text{otherwise.} \end{cases}$$

Now the best known bound for the upper tail for general subgraph count says the following:

Theorem 1.1.2 (Theorem 1.2 in [61])**.** *For every graph F and every $\varepsilon > 0$ there exist positive real numbers $c(\varepsilon, F), C(\varepsilon, F)$ such that for all $n \ge v(F)$ and $p \in (0,1)$ we have*

$$\mathbb{P}(X_n(F) \ge (1+\varepsilon)\,\mathbb{E}[X_n(F)]) \le \exp\left(-c(\varepsilon,F)M_F^*(n,p)\right)$$

and, provided $(1+\varepsilon)\,\mathbb{E}[X_n(F)] \le N(K_n, G)$,

$$\mathbb{P}(X_n(F) \ge (1+\varepsilon)\,\mathbb{E}[X_n(F)]) \ge \exp\left(-C(\varepsilon,F)M_F^*(n,p)|\log p|\right)$$

where K_n is the complete graph on n vertices.

Let $\Delta(F)$ denote the maximum degree of F. Then,

$$M_F^*(n,p) = \Theta(n^2 p^{\Delta(F)})$$

as long as $p \gg n^{-1/\Delta(F)}$ (see [62]). We investigate the behavior of $\log \mathbb{P}(X_n(F) \ge (1+\varepsilon)\,\mathbb{E}[X_n(F)])$ when ε and p are fixed.

In Theorem 2.3.4 we prove a large deviation result for the number of triangles in $G(n,p)$ which gives explicit rate parameters. Let us define the function $I(\cdot,\cdot)$ on $(0,1)\times(0,1)$ as $I(r,s) := r\log(r/s) + (1-r)\log((1-r)/(1-s))$ which is the relative entropy of Bernoulli(r) w.r.t. Bernoulli(s) measure. The function $I(\cdot,\cdot)$ appears as the large deviation rate function for number of edges in $G(n,p)$. We prove the following result:

Theorem 1.1.3. *Let T_n be the number of triangles in $G(n,p)$, where $p > p_0$ where $p_0 = 2/(2+e^{3/2}) \approx 0.31$. Then for any $r \in (p,1]$,*

$$\mathbb{P}(T_n \geq n^3r^3/6) = e^{-\frac{1}{2}n^2 I(r,p)(1+o(1))}.$$

Moreover, even if $p \leq p_0$, there exist p', p'' such that $p < p' \leq p'' < 1$ and the same result holds for all $r \in (p,p') \cup (p'',1]$.

The result is a nontrivial consequence of Stein's method for concentration inequalities and involves analyzing the tilted measure, which in this case leads to what is known as an 'exponential random graph', a little studied object in the rigorous literature. Clearly, our result gives a lot more in the situations where it works (see Figure 1). The method of proof can be easily extended to prove similar results for general subgraph counts and are discussed in Section 2.3.3. However, there is an obvious incompleteness in Theorem 2.3.4 (and also for general subgraphs counts), namely, that it does not work for all (p,r). It is an interesting open problem to solve the large deviation problem for the whole region. Here we mention that, in a recent article in preparation, Chatterjee and Varadhan [31] have obtained the large deviation rate function in the full regime using Szemerédi regularity lemma.

In Section 2.3.1 we prove a super-Gaussian concentration inequality for critical Curie-Weiss model. The 'Curie-Weiss model of ferromagnetic interaction' at inverse temperature β and zero external field is given by the following Gibbs measure on $\{+1,-1\}^n$. For a typical configuration $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n) \in \{+1,-1\}^n$ the probability of $\boldsymbol{\sigma}$ is given by

$$Z_\beta^{-1} \exp\left(\beta \sum_{i<j} \sigma_i\sigma_j/n\right)$$

where Z_β is the normalizing constant. It is well known that the Curie-Weiss model shows a phase transition at $\beta_c = 1$. Using concentration inequalities for exchangeable pairs it was proved in [24] that for all $\beta \geq 0, n \geq 1, t \geq 0$ we have

$$\mathbb{P}\left(\sqrt{n}|m - \tanh(\beta m)| \geq t + \beta/\sqrt{n}\right) \leq 2e^{-t^2/(4+4\beta)}.$$

It is known that at $\beta = 1$ as $n \to \infty$, $n^{1/4}m(\boldsymbol{\sigma})$ converges to the probability distribution on $\mathbb{R}$ having density proportional to $\exp(-t^4/12)$ (see Simon and Griffiths [100]). The following concentration inequality stated in Proposition 2.3.1 and derived using Theorem 2.2.2, fills the gap in the tail bound at the critical point.

Theorem 1.1.4. *Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at the critical temperature $\beta = 1$. Then, for any $n \geq 1$ and $t \geq 0$ the magnetization satisfies*

$$\mathbb{P}(n^{1/4}|m(\boldsymbol{\sigma})| \geq t) \leq 2e^{-ct^4}$$

where $c > 0$ is an absolute constant.

Here we may remark that such a concentration inequality probably cannot be obtained by application of standard off-the-shelf results (e.g. those surveyed in Ledoux [75], the famous results of Talagrand [106] or the recent breakthroughs of Boucheron, Lugosi and Massart [22]), because they generally give Gaussian or exponential tail bounds. There are several recent remarkable results giving tail bounds different from exponential and Gaussian (see [14, 74, 9, 45, 33, 15, 49, 50]). However, it seems that none of the techniques given in these references would lead to the above result. We also look at general critical Curie-Weiss models. In Section 2.3.4, we derive some interesting concentration bounds for Ising model on d-dimensional square lattices.

1.1.2 First-passage percolation

In 1965, Hammersley and Welsh [54] introduced first-passage percolation to model the spread of fluid through a randomly porous media. The model is defined as follows. Consider the d-dimensional cubic lattice $\mathbb{Z}^d$ and the edge set E consisting of nearest neighbor edges. With each edge $e \in E$ is associated an independent nonnegative random variable ω_e distributed according to a fixed distribution F. The random variable ω_e represents the amount of time it takes the fluid to pass through the edge e. For a finite path $\mathcal{P}$ in $\mathbb{Z}^d$ define

$$\omega(\mathcal{P}) := \sum_{e \in \mathcal{P}} \omega_e$$

as the passage time for $\mathcal{P}$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^d$, the *first-passage time* $a(\boldsymbol{x}, \boldsymbol{y})$ is defined as the minimum passage time over all paths from $\boldsymbol{x}$ to $\boldsymbol{y}$. Intuitively $a(\boldsymbol{x}, \boldsymbol{y})$ is the first time the fluid will appear at $\boldsymbol{y}$ if a source of water is introduced at the vertex $\boldsymbol{x}$ at time 0. We postpone the discussion about known results until Section 3.1.

Convergence to the Tracy-Widom law is known for *directed* last-passage percolation in $\mathbb{Z}^2$ under very special conditions, but the techniques do not carry over to the undirected case. Naturally, one may expect that convergence to something like the Tracy-Widom distribution may hold for undirected first-passage percolation also, but surprisingly, this does not seem to be the case. Here we mention that, in fact, almost no nontrivial distributional result is known for undirected first-passage percolation.

In Chapter 3 we consider first-passage percolation on $\mathbb{Z}^d$ with height restricted by an integer h (which is allowed to grow with n). We define

$$a_n(h) := \inf\{\omega(\mathcal{P}) \mid \mathcal{P} \text{ is a path from } \mathbf{0} \text{ to } n\boldsymbol{e}_1 \text{ in } \mathbb{Z} \times \{-h, -h+1, \ldots, h\}^{d-1}\}$$

where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$. Informally, $a_n(h)$ is the minimal passage time over all paths which deviate from the straight line path joining the two end points by a distance at most h. Given the dimension d, we consider a non-degenerate distribution F supported on $[0, \infty)$ for which we have $F(\lambda) < p_c(d)$ where λ is the smallest point in the support of F and $p_c(d)$ is the critical probability for Bernoulli bond percolation in $\mathbb{Z}^d$. Standard result gives that

$$\nu(\boldsymbol{e}_1) := \lim_{n\to\infty} \mathbb{E}[a(\mathbf{0}, n\boldsymbol{e}_1)]/n \tag{1.1}$$

exists and is positive when $F(0) < p_c(d)$. In Theorem 3.1.2 we proved that for cylinders that are 'thin' enough, a Gaussian CLT holds for $a_n(h)$ after proper centering and scaling. Let $\mu_n(h_n)$ and $\sigma_n^2(h_n)$ be the mean and variance of $a_n(h_n)$.

Theorem 1.1.5. *Let F be as above. Suppose $\mathbb{E}[\omega^p] < \infty$ for all $p < \infty$. Let $\{h_n\}_{n\geq 1}$ be a sequence of integers satisfying $h_n = o(n^\alpha)$ where $\alpha < 1/(d+1)$. Then we have*

$$(a_n(h_n) - \mu_n(h_n))/\sigma_n(h_n) \xrightarrow{\text{w}} N(0,1) \text{ as } n \to \infty.$$

When $h_n \to \infty$ as $n \to \infty$, $\lim_{n\to\infty} \mu_n(h_n)/n = \nu(\boldsymbol{e}_1)$, where $\nu(\boldsymbol{e}_1)$ is defined as in (1.1)*. Moreover, we have $c_1 n h_n^{-d+1} \leq \sigma_n^2(h_n) \leq c_2 n$ for some positive absolute constants c_i depending only on d and F.*

The main idea behind Theorem 3.1.2 is to decompose $a_n(h_n)$ as an "approximate" sum of i.i.d. random variables. The CLT is relatively easier to prove when $h_n = o(n^{1/(3d-1)})$. However, using a blocking technique, which is reminiscent of the "renormalization group" method, by successively breaking into smaller cylinders, we finally extend the growth rate of h_n to $o(n^{1/(d+1)})$. In fact Theorem 3.1.2 give rise to a new exponent $\gamma(d)$ defined as

$$\gamma(d) := \sup\{\alpha : (a_n(n^\alpha) - \mu_n(n^\alpha))/\sigma_n(n^\alpha) \xrightarrow{\text{w}} N(0,1) \text{ as } n \to \infty\}.$$

Clearly we have $\gamma(d) \geq 1/(d+1)$ for F having all moments finite and satisfying the conditions in Theorem 3.1.2. Is $\gamma(d)$ actually equal to $1/(d+1)$? There are indications that this is not true. In Section 3.6 we provide some heuristic justifications for that. In Section 3.9 we provide some numerical results in support of the following conjecture:

Conjecture 1.1.6. *For $d = 2$, we have $\gamma(d) = 2/3$ and $\sigma_n^2(h_n) = \Theta(n h_n^{-1/2})$.*

One of the future project is to prove Central limit theorem upto $n^{2/3}$ and extend the idea to passage times involving monotone paths.

1.1.3 Spectra of random linear combination of projection matrices

For a symmetric $n \times n$ matrix A, let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ denote its eigenvalues arranged in nonincreasing order. The spectral measure Λ_A of A is defined as the empirical measure of its eigenvalues which puts mass $1/n$ to each of its eigenvalues, *i.e.,*

$$\Lambda_A = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}$$

where δ_x is the dirac measure at x. In particular when the matrix A is random we have a random spectral measure corresponding to A.

In his seminal paper [111] Wigner proved that the spectral measure for a large class of random matrices converges to the semi-circular law, as the dimension grows to infinity. Much work has since been done on various aspects of eigenvalues for different ensembles of large real symmetric or complex hermitian random matrices, random matrices coming from Haar measure on classical groups (e.g., orthogonal, unitary, simplectic group). Some of the results are surveyed in [53, 85]. Many new results have been proved in the last few years for understanding liming spectral distribution of large random matrices having complicated algebraic structure. In [23] the authors considered the spectra of large random Hankel, Markov and Toeplitz matrices which was inspired by an open problem in [5] (see also [55]). Recently, in [43] the author considered linear combinations of matrices defined via representations and coxeter generators of the symmetric group.

In many of the examples the random matrix can be written a linear function $\sum_\alpha X_\alpha M_\alpha^{(n)}$ of i.i.d. random variables $\{X_\alpha\}$ where $M_\alpha^{(n)}$'s are deterministic matrices. For example Wigner matrices can be written as $\sum_{i\le j} X_{ij} M_{ij}^{(n)}$ where $M_{ij}^{(n)}$ is the $n\times n$ matrix with 1 at the (i,j) and (j,i)-th position and zero everywhere else.

In Chapter 4, we investigate the case when $M_\alpha^{(n)}$ is a projection matrix (or a affine transform of a projection matrix). Recall that a projection matrix P satisfies $P = P^* = P^2$. The Markov random matrix example in [23] and the result in [43] fall in this category.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. real random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}(X_1^2) = 1$. Given n, suppose we have $k = k(n)$ many $n \times n$ symmetric matrices $M_1^{(n)}, M_2^{(n)}, \ldots, M_k^{(n)}$. For simplicity, we assume that all $M_i^{(n)}$'s are projection matrices for $i = 1, 2, \ldots, k$. Now consider the random matrix

$$A_n = \sum_{i=1}^{k} a_i^{(n)} X_i M_i^{(n)}$$

where $\{a_i^{(n)}\}$ is a sequence of nonnegative real numbers. Let Λ_n be the spectral measure of A_n. Clearly Λ_n is a random measure on $\mathbb{R}$. In Lemma 4.2.1 we provide simple conditions under which universality holds.

We assume that $\mu_k(n) := \mathrm{Tr}(M_{i_1}^{(n)} M_{i_2}^{(n)} \cdots M_{i_k}^{(n)})$ depends only on k, n when $i_1, i_2, \ldots, i_k$'s are distinct integers such that $M_{i_1}^{(n)}, M_{i_2}^{(n)}, \ldots, M_{i_k}^{(n)}$ commute with each other. Our main theorem (Theorem 4.2.4) says that:

Theorem 1.1.7. *Assume that*

$$\sum_{i=1}^{k(n)} \big(a_i^{(n)}\big)^2 = 1$$

and

$$\max_{1\le i\le k(n)} |a_i^{(n)}| \to 0, \quad \sum_{(i,j)\in E_n} \big(a_i^{(n)} a_j^{(n)}\big)^2 \to 0 \text{ as } n\to\infty$$

where $E_n := \{(i,j) : M_i^{(n)}$ *does not commute with* $M_j^{(n)}\}$. *Also assume that*

$$\frac{\mu_1(n)}{n} \to \theta \text{ and } \frac{\mu_2(n)}{n} \to \theta^2 \text{ as } n\to\infty$$

for some real number $\theta \in [0,1]$. *Let* Λ_n *be the empirical spectral distribution of*

$$A_n = \sum_{i=1}^{k(n)} a_i^{(n)} Z_i M_i^{(n)}$$

where Z_i*'s are i.i.d. standard Gaussian random variables. Then* Λ_n *converges in distribution (with respect to the topology of weak convergence of probability measures on* $\mathbb{R}$*) to a random distribution* Λ_∞ *in probability where* $\Lambda_\infty = \nu_Z, Z$ *is* $N(0,1)$ *and* ν_z *is the distribution* $N(\theta z, \theta(1-\theta))$.

In Section 4.2 we describe the main results of Chapter 4. The proof uses moment method and Malliavin calculus. We will provide several examples from representation theory of symmetric groups in Section 4.3 and some generalization in Section 4.4.

In the next section we briefly describe the concept of Stein's method using the example of magnetization in critical Curie-Weiss model.

1.2 Stein's method

For two random variables X and Z, the most natural and popular way of measuring the distance between them is to consider a class of functions $\mathcal{F}$ and consider the distance

$$d_{\mathcal{F}}(X,Z) = \sup_{f\in\mathcal{F}} |\,\mathbb{E}[f(X) - f(Z)]|.$$

Various choices of family $\mathcal{F}$ lead to different notions of distances between two probability measures. Famous examples of such distances include Total variation distance, Kolmogorov distance, Wasserstein distance and so on.

Stein's revolutionary idea [103] was that instead of bounding the difference for every function $f \in \mathcal{F}$ break the problem into several manageable independent parts and use the properties of X and Z that will imply their closeness in distribution.

(a) The first step is, to construct an operator T_0 defined on an appropriate function space $\mathcal{H}_Z$ that characterizes the distribution of Z in the sense that for some random variable W, $\mathbb{E}[T_0 f(W)] = 0$ for all $f \in \mathcal{H}_Z$ implies W and Z have the same distribution. The operator T_0 is called the *Stein operator*. For example, if Z has a standard normal distribution, then

$$(T_0 f)(x) = f'(x) - xf(x) \text{ for } f \in \mathcal{D}$$

where $\mathcal{D}$ = set of all locally absolutely continuous functions, is a Stein operator.

(b) Similarly we construct an operator T on some function space $\mathcal{H}_X$ such that $\mathbb{E}[Tf(X)] = 0$ for all $f \in \mathcal{H}_X$. If we think of X as sample version of Z, then T can be viewed as a sample version of T_0.

(c) Finally, one studies the properties of the pseudo-inverse U of T_0, if it exists, such that $T_0 U(f) = f - \mathbb{E}f(Z)$ for all $f \in \mathcal{F}$ and $U(\mathcal{F}) \subseteq \mathcal{H} := \mathcal{H}_Z \cap \mathcal{H}_X$.

(d) Now, since

$$\begin{aligned}
|\,\mathbb{E}[f(X) - f(Z)]| &= |\,\mathbb{E}[T_0 U f(X)]| \\
&= |\,\mathbb{E}(T_0 - T)Uf(X)| \le \sup_{g \in \mathcal{H}} |\,\mathbb{E}[(T_0 - T)g(X)]|
\end{aligned}$$

for $f \in \mathcal{F}$, the job boils down to showing that the operators T and T_0 are "close" when restricted to the set $\mathcal{H}$. And in most of the cases this is the hardest part to analyze.

Note that if the distribution of Z is the equilibrium distribution of a stationary reversible Markov process with infinitesimal generator $\mathcal{A}$, then $\mathcal{A}$ is a Stein operator for Z. So the natural thing to consider is to construct a reversible Markov chain with generator $\mathcal{B}$ and having stationary distribution given by the "sample" X and prove convergence of $\mathcal{B}$ to $\mathcal{A}$ in appropriate sense to prove process convergence. However, proving convergence for the equilibrium distribution is much more simpler than proving convergence for the whole process. The simplicity of Stein's method of exchangeable pair comes from the fact that it uses only one step of the reversible Markov chain (which gives an exchangeable pair) to prove convergence.

In the exchangeable pair approach the "sample" operator T is created using an exchangeable pair. First construct a random variable X' such that (X, X') is an exchangeable pair. Suppose both X, X' takes values in $\mathcal{X}$. Then find an operator α such that for any suitable real valued function $g : \mathcal{X} \to \mathbb{R}$, $\alpha g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is an antisymmetric function (that is, $(\alpha g)(x, x') = (\alpha g)(x', x)$). Then, by antisymmetry, the operator

$$Tg(x) = \mathbb{E}[(\alpha g)(X, X')|X = x]$$

gives a "sample" characterizing operator and the problem boils down to bounding

$$\sup_x |(T - T_0)g(x)|$$

for $g \in U\mathcal{F}$.

There are other variations of Stein's method that exploit the characterizing operator in different ways, for example the zero bias transformation popularized by Goldstein [47, 46], the size bias coupling [7, 8, 48], dependency graph approach of Arratia, Goldstein and Gordon [2, 3], and other ad hoc methods [18, 34], but we shall not discuss those here. For further discussion and exposition on Stein's method of exchangeable pair we refer to the monograph [37].

1.2.1 Exact convergence rate in critical Curie-Weiss model

We illustrate the concept using the example of magnetization in critical Curie-Weiss model and finding the *exact* rate of convergence w.r.t. Wasserstein distance. An upper bound for the convergence rate w.r.t. kolmogorov distance is given in [30] (see also [39]).

First we recall the definition of critical Curie-Weiss model from Subsection 1.1.1. The critical Curie-Weiss model of ferromagnetic interaction at zero external field is given by the following gibbs measure on $\{+1,-1\}^n$. For a typical configuration $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n) \in \{+1,-1\}^n$ the probability of $\boldsymbol{\sigma}$ is given by

$$\mu_n(\boldsymbol{\sigma}) := Z_n^{-1} \exp\left(\frac{1}{n}\sum_{i<j} \sigma_i\sigma_j\right).$$

where Z_n is the normalizing constant. Define the magnetization as $m(\boldsymbol{\sigma}) = \frac{1}{n}\sum_{i=1}^n \sigma_i$. Consider the random variable $X_n = n^{1/4} m(\boldsymbol{\sigma})$ where $\boldsymbol{\sigma} \sim \mu_n$. It is known that X_n converges in distribution to Z as $n \to \infty$ where Z has density proportional to $\exp(-t^4/12)$ (see Simon and Griffiths [100]). As stated earlier, in Section 2.3.1 we will prove a super-Gaussian concentration inequality for X_n. Here we consider the rate of convergence w.r.t. Wasserstein distance:

$$d_{\mathcal{W}}(X_n, Z) = \sup_{g:\sup_{x\in\mathbb{R}} |g'(x)| \le 1} |\mathbb{E}(g(X_n) - g(Z))|.$$

We show that,

Lemma 1.2.1. *There exists a constant $a \in (0,\infty)$ such that,*

$$n^{1/2} d_{\mathcal{W}}(X_n, Z) \to a$$

as $n \to \infty$.

Proof. Here we have $\mathcal{F} = \{g : \mathbb{R} \to \mathbb{R} \mid g \text{ is 1-Lipschitz}\}$. It is easy to check that the operator T_0 acting on functions in $\mathcal{F}$ by

$$T_0 f(x) = f'(x) - x^3 f(x)/3$$

is a Stein operator for the distribution of Z. Also the operator U defined by

$$Ug(x) := e^{x^4/12} \int_{\infty}^{x} (g(y) - \mathbb{E}(g(Z)))e^{-y^4/12}\, dx$$

gives the pseudo-inverse of T_0 in the sense that $(T_0 U)g = g - \mathbb{E}(g(Z))$ for all $g \in \mathcal{F}$. An analytical calculation (or see Lemma 4.1 in [30]) shows that

$$U\mathcal{F} \subseteq \mathcal{H} := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is twice differentiable}, \sup_{x\in\mathbb{R}}(|f'(x)|+|f'(x)|+|f''(x)|) \leq c\}$$

for a constant $c < \infty$.

Now we construct the sample operator T_n using exchangeable pair. Let $\boldsymbol{\sigma}'$ be obtained from $\boldsymbol{\sigma}$ by one step of the heat-bath Glauber dynamics: A coordinate I is chosen uniformly at random from $\{1, 2, \ldots, n\}$, and σ_I is replace by σ_I' drawn from the conditional distribution of the I-th coordinate given $\{\sigma_j : j \neq I\}$. An easy computation gives that $\mathbb{E}(\sigma_i|\{\sigma_j, j \neq i\}) = \tanh(m_i)$ where $m_i = m_i(\boldsymbol{\sigma}) = n^{-1}\sum_{j\neq i}\sigma_j$ for all $i = 1, 2, \ldots, n$. Now define $X_n' = n^{1/4}m(\boldsymbol{\sigma}')$. Clearly (X_n, X_n') is an exchangeable pair and we have $X_n' - X_n = n^{-3/4}(\sigma_I' - \sigma_I)$ where I is uniform over $\{1, 2, \ldots, n\}$.

Given a function $f \in \mathcal{H}$, define the function $F : \mathbb{R} \to \mathbb{R}$ by

$$F(x) = \int_0^x f(y)dy \text{ for } x \in \mathbb{R}$$

so that $F' = f$. Note that f is twice-differentiable. We define

$$T_n f(x) := n^{3/2}\, \mathbb{E}[F(X_n') - F(X_n)|X_n = x].$$

By Taylor approximation (and the fact that $|X_n - X_n'| \leq 2n^{-3/4}$ a.s.) we have

$$\begin{aligned} &T_n f(x) \\ &= n^{3/2} f(x)\, \mathbb{E}(X_n' - X_n|X_n = x) + \frac{n^{3/2} f'(x)}{2}\, \mathbb{E}((X_n' - X_n)^2|X_n = x) + R \end{aligned} \tag{1.2}$$

where $|R| \leq n^{3/2}/6 \times (2n^{-3/4})^3 \sup_{x\in\mathbb{R}} |f''(x)| \leq Cn^{-3/4}$ for some constant C. After explicit calculation and substituting the conditional means we have

$$n^{3/2}\, \mathbb{E}[X_n' - X_n|X_n] = n^{3/4}\, \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \tanh(m_i(\boldsymbol{\sigma})) - m(\boldsymbol{\sigma})\middle| m(\boldsymbol{\sigma})\right]. \tag{1.3}$$

We now expand the hyperbolic tangent function upto degree 7 using Taylor series to obtain

$$\begin{aligned}\tanh(m_i) &= m_i - \frac{1}{3}m_i^3 + \frac{2}{15}m_i^5 + \mathcal{O}(|m_i|^7) \\ &= m - \frac{1}{3}m^3 + \frac{2}{15}m^5 - \frac{\sigma_i}{n}\left(1 - m^2 + \frac{2}{3}m^3\right) + \varepsilon_i\end{aligned}$$

where $\mathbb{E}\,|\varepsilon_i| \leq Cn^{-9/4}$ for all $i = 1, 2, \ldots, n$ and C is a universal constant. Substituting in (1.3) it follows that

$$\begin{aligned}n^{3/2}\,\mathbb{E}[X_n' - X_n|X_n] &= n^{3/4}\left[-\frac{1}{3}m^3 + \frac{2}{15}m^5 - \frac{1}{n}\left(m - m^3 + \frac{2}{3}m^4\right)\right] + n^{-1/4}\sum_{i=1}^{n}\varepsilon_i \\ &= -\frac{1}{3}X_n^3 - \frac{1}{\sqrt{n}}\left(X_n - \frac{2}{15}X_n^5\right) + R_1\end{aligned}$$

where $\mathbb{E}\,|R_1| \leq Cn^{-1}$. Similarly we have

$$\begin{aligned}\frac{1}{2}n^{3/2}\,\mathbb{E}[(X_n' - X_n)^2|X_n] &= \mathbb{E}[1 - \sigma_I\sigma_I'|X_n] \\ &= 1 - \frac{1}{n}\sum_{i=1}^{n}\sigma_i\tanh(m_i(\boldsymbol{\sigma})) = 1 - \frac{X_n^2}{\sqrt{n}} - R_2\end{aligned}$$

where $\mathbb{E}\,|R_2| \leq Cn^{-1}$. Substituting in equation (1.2) we finally have

$$T_nf(x) = f'(x)\left(1 - \frac{x^2}{\sqrt{n}}\right) - f(x)\left(\frac{1}{3}x^3 + \frac{1}{\sqrt{n}}\left(x - \frac{2}{15}x^5\right)\right) + R'$$

where $|R'| \leq Cn^{-3/4}$ for some constant $C < \infty$. Now claerly

$$\sqrt{n}\,\mathbb{E}(T_nf(X_n) - T_0f(X_n)) = -\,\mathbb{E}(X_n^2f'(X_n) + \left(X_n - \frac{2}{15}X_n^5\right)f(X_n)) + R''$$

where $|R''| \leq Cn^{-1/4}$. It thus follows that

$$d_{\mathcal{W}}(X_n, Z) \leq cn^{-1/2}$$

for some constant c. Now note that

$$\begin{aligned}&\mathbb{E}\left[X_n^2f'(X_n) + \left(X_n - \frac{2}{15}X_n^5\right)f(X_n)\right] \\ &\longrightarrow \mathbb{E}\left(Z^2f'(Z) + Z\left(1 - \frac{2}{15}Z^4\right)f(Z)\right) = \frac{1}{5}\,\mathbb{E}[Z(Z^4 - 5)f(Z)]\end{aligned}$$

as $n \to \infty$ by uniform integrability. Here we used the fact that $\mathbb{E}[f'(Z)] = \frac{1}{3}\,\mathbb{E}[Z^3 f(Z)]$ for all f, specially for $x^2 f(x)$. Define the function

$$f(x) := \frac{cx}{1+x^4}, x \in \mathbb{R}$$

and $g(x) := f'(x) - \frac{x^3}{3} f(x)$ where $c > 0$ is some constant to be specified later. It is easy to check that g is 1-Lipschitz for appropriate choice of c. Now

$$\mathbb{E}[Z(Z^4-5)f(Z)] = \mathbb{E}\,\frac{cZ^2(Z^4-5)}{1+Z^4} \neq 0.$$

Hence $d_{\mathcal{W}}(X_n, Z) = \Theta(n^{-1/2})$. Moreover we have

$$\lim_{n\to\infty} n^{1/2} d_{\mathcal{W}}(X_n, Z) = \frac{1}{5} \sup_{\substack{f: f = Ug \\ g \text{ 1-Lipschitz}}} \left|\mathbb{E}[Z(Z^4-5)f(Z)]\right|.$$

□

Chapter 2

Concentration inequalities using exchangeable pairs

2.1 Introduction

Stein's method was introduced by Charles Stein in the early seventies to prove central limit theorem for dependent random variables and more importantly to find explicit estimates for the accuracy of the approximation. The technique is primarily used for proving distributional limit theorems (both Gaussian and non-Gaussian). Stein's attempts [104] at devising a version of the method for large deviations did not prove fruitful. Some progress for sums of dependent random variables was made by Raič [93]. The problem was finally solved in full generality in [24]. A selection of results and examples from [24] appeared in the later papers [28, 27]. In this chapter we extend the theory and work out some further examples.

The sections are organized as follows. In Section 2.2 we state the main results. In Section 2.3 we state the examples and some proof sketches. The complete proofs are in Section 2.4.

2.2 Results

The following abstract theorem is quoted from [28]. It summarizes a collection of results from [24]. This is a generalization of Stein's method of exchangeable pairs to the realm of concentration inequalities and large deviations.

Theorem 2.2.1 ([28], Theorem 1.5)**.** *Let $\mathcal{X}$ be a separable metric space and suppose (X, X') is an exchangeable pair of $\mathcal{X}$-valued random variables. Suppose $f : \mathcal{X} \to \mathbb{R}$ and $F : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ are square-integrable functions such that F is antisymmetric (i.e. $F(X, X') = -F(X', X)$ a.s.), and $\mathbb{E}(F(X, X') \mid X) = f(X)$ a.s. Let*

$$\Delta(X) := \frac{1}{2}\,\mathbb{E}\big(|(f(X) - f(X'))F(X, X')|\,\big|\,X\big).$$

Then $\mathbb{E}(f(X)) = 0$, *and the following concentration results hold for* $f(X)$*:*

(i) *If* $\mathbb{E}(\Delta(X)) < \infty$, *then* $\operatorname{Var}(f(X)) = \frac{1}{2}\mathbb{E}((f(X) - f(X'))F(X,X'))$.

(ii) *Assume that* $\mathbb{E}(e^{\theta f(X)}|F(X,X')|) < \infty$ *for all* θ. *If there exists nonnegative constants* B *and* C *such that* $\Delta(X) \le Bf(X) + C$ *almost surely, then for any* $t \ge 0$,

$$\mathbb{P}\{f(X) \ge t\} \le \exp\left(-\frac{t^2}{2C + 2Bt}\right) \quad \textit{and} \quad \mathbb{P}\{f(X) \le -t\} \le \exp\left(-\frac{t^2}{2C}\right).$$

(iii) *For any positive integer* k, *we have the following exchangeable pairs version of the Burkholder-Davis-Gundy inequality:*

$$\mathbb{E}(f(X)^{2k}) \le (2k-1)^k\, \mathbb{E}(\Delta(X)^k).$$

Note that the finiteness of the exponential moment for all θ ensures that the tail bounds hold for all t. If it is finite only in a neighborhood of zero, the tail bounds will hold for t less than a threshold.

One of the contributions of the present thesis is the following generalization of the above result for non-Gaussian tail behavior. We apply it to obtain a concentration inequality with the correct tail behavior in the Curie-Weiss model at criticality.

Theorem 2.2.2. *Suppose* (X, X') *is an exchangeable pair of random variables. Let* $F(X,X')$, $f(X)$ *and* $\Delta(X)$ *be as in Theorem 2.2.1. Suppose that we have*

$$\Delta(X) \le \psi(f(X)) \textit{ almost surely}$$

for some nonnegative symmetric function ψ *on* $\mathbb{R}$. *Assume that* ψ *is nondecreasing and twice continuously differentiable in* $(0,\infty)$ *with*

$$\alpha := \sup_{x>0} x\psi'(x)/\psi(x) < 2 \tag{2.1}$$

$$\textit{and } \delta := \sup_{x>0} x\psi''(x)/\psi(x) < \infty. \tag{2.2}$$

Assume that $\mathbb{E}(|f(X)|^k) < \infty$ *for all positive integer* $k \ge 1$. *Then for any* $t \ge 0$ *we have*

$$\mathbb{P}(|f(X)| > t) \le c\exp\left(-\frac{t^2}{2\psi(t)}\right)$$

for some constant c *depending only on* α, δ. *Moreover, if* ψ *is only once differentiable with* $\alpha < 2$ *as in* (2.1), *then the tail inequality holds with exponent* $t^2/4\psi(t)$.

An immediate corollary of Theorem 2.2.2 is the following.

Corollary 2.2.3. *Suppose (X, X') is an exchangeable pair of random variables. Let $F(X, X')$, $f(X)$ and $\Delta(X)$ be as in Theorem 2.2.1. Suppose that for some real number $\alpha \in (0, 2)$ we have*

$$\Delta(X) \leq B\,|f(X)|^{\alpha} + C \text{ almost surely}$$

where $B > 0, C \geq 0$ are constants. Assume that $\mathbb{E}(|f(X)|^k) < \infty$ for all positive integer $k \geq 1$. Then for any $t \geq 0$ we have

$$\mathbb{P}(|f(X)| > t) \leq c_\alpha \exp\left(-\frac{1}{2} \cdot \frac{t^2}{Bt^\alpha + C}\right)$$

for some constant c_α depending only on α.

The result in Theorem 2.2.2 states that the tail behavior of $f(X)$ is essentially given by the behavior of $f(X)^2/\Delta(X)$. Condition (2.1) implies that $\psi(x) < \psi(1)(1+x^2)$ for all $x \in \mathbb{R}$. Moreover, the constant c_α appearing in Theorem 2.2.2 can be written down explicitly but we did not attempt to optimize the constant. The proof of Theorem 2.2.2 is along the same lines as Theorem 2.2.1, but somewhat more involved. Deferring the proof to Section 2.4, let us move on to examples.

2.3 Examples

2.3.1 Curie-Weiss model at criticality

The 'Curie-Weiss model of ferromagnetic interaction' at inverse temperature β and zero external field is given by the following Gibbs measure on $\{+1, -1\}^n$. For a typical configuration $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n) \in \{+1, -1\}^n$ the probability of $\boldsymbol{\sigma}$ is given by

$$\mu_\beta(\{\boldsymbol{\sigma}\}) := Z_\beta^{-1} \exp\left(\frac{\beta}{n} \sum_{i<j} \sigma_i \sigma_j\right)$$

where $Z_\beta = Z_\beta(n)$ is the normalizing constant. It is well known that the Curie-Weiss model shows a phase transition at $\beta_c = 1$. For $\beta < \beta_c$ the magnetization $m(\boldsymbol{\sigma}) := \frac{1}{n}\sum_{i=1}^n \sigma_i$ is concentrated at 0 but for $\beta > \beta_c$ the magnetization is concentrated on the set $\{-x^*, x^*\}$ where $x^* > 0$ is the largest solution of the equation $x = \tanh(\beta x)$. In fact using concentration inequalities for exchangeable pairs it was proved in [24] (Proposition 1.3) that for all $\beta \geq 0, h \in \mathbb{R}, n \geq 1, t \geq 0$ we have

$$\mathbb{P}\left(|m - \tanh(\beta m + h)| \geq \frac{\beta}{n} + \frac{t}{\sqrt{n}}\right) \leq 2\exp\left(-\frac{t^2}{4(1+\beta)}\right),$$

where h is the external field, which is zero in our case. Although a lot is known about this model (see Ellis [40] Section IV.4 for a survey), the above result – to the best of our knowledge – is the first rigorously proven concentration inequality that holds at all temperatures. (See also [33] for some related results.)

Incidentally, the above result shows that when $\beta < 1$, the magnetization is at most of order $n^{-1/2}$. It is known that at the critical temperature the magnetization $m(\boldsymbol{\sigma})$ shows a non Gaussian behavior and is of order $n^{-1/4}$. In fact, at $\beta = 1$ as $n \to \infty$, $n^{1/4}m(\boldsymbol{\sigma})$ converges to the probability distribution on $\mathbb{R}$ having density proportional to $\exp(-t^4/12)$. This limit theorem was first proved by Simon and Griffiths [100] and error bounds were obtained recently [30, 39]. The following concentration inequality, derived using Theorem 2.2.2, fills the gap in the tail bound at the critical point.

Proposition 2.3.1. *Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at the critical temperature $\beta = 1$. Then, for any $n \geq 1$ and $t \geq 0$ the magnetization satisfies*

$$\mathbb{P}(n^{1/4}|m(\boldsymbol{\sigma})| \geq t) \leq 2e^{-ct^4}$$

where $c > 0$ is an absolute constant.

Here we may remark that such a concentration inequality probably cannot be obtained by application of standard off-the-shelf results (e.g. those surveyed in Ledoux [75], the famous results of Talagrand [106] or the recent breakthroughs of Boucheron, Lugosi and Massart [22]), because they generally give Gaussian or exponential tail bounds. There are several recent remarkable results giving tail bounds different from exponential and Gaussian. The papers [74, 45, 33] deal with tails between exponential and Gaussian and [9, 15] deal with sub-exponential tails. Also in [14, 49, 50] the authors deal with tails (possibly) larger than Gaussian. However, it seems that none of the techniques given in these references would lead to the result of Proposition 2.3.1.

It is possible to derive a similar tail bound using the asymptotic results of Martin-Löf [79] about the partition function $Z_\beta(n)$ (see also Bolthausen [19]). An application of their results gives that

$$\sum_{\boldsymbol{\sigma}\in\{-1,+1\}^n} e^{\frac{n}{2}m(\boldsymbol{\sigma})^2+n\theta m(\boldsymbol{\sigma})^4} \simeq \frac{2^{n+1}\Gamma(5/4)}{\sqrt{2\pi}}\left(\frac{12n}{1-12\theta}\right)^{1/4}$$

for $\theta < 1/12$ in the sense that the ratio of the two sides converges to one as n goes to infinity and from here the tail bound follows easily (without an explicit constant). However this approach depends on a precise estimate of the partition function (for example, large deviation estimates or finding the limiting free energy $\lim n^{-1}\log Z_\beta(n)$ are not enough) and this precise estimate is hard to prove. Our method, on the other hand, depends only on simple properties of the Gibbs measure and is not tied specifically to the Curie-Weiss model.

The idea used in the proof of Proposition 2.3.1 can be used to prove a tail inequality that holds for all $0 \leq \beta \leq 1$. We state the result below without proof. Note that the inequality gives the correct tail bound for all $0 \leq \beta \leq 1$.

Proposition 2.3.2. *Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at inverse temperature β where $0 \le \beta \le 1$. Then, for any $n \ge 1$ and $t \ge 0$ the magnetization satisfies*

$$\mathbb{P}(3(1-\beta)m(\boldsymbol{\sigma})^2 + \beta^3 m(\boldsymbol{\sigma})^4 \ge t) \le 2e^{-nt/160}.$$

It is possible to derive similar non-Gaussian tail inequalities for general Curie-Weiss models at the critical temperature. We briefly discuss the general case below. Let ρ be a symmetric probability measure on $\mathbb{R}$ with $\int x^2 \; d\rho(x) = 1$ and $\int \exp(\beta x^2/2) \; d\rho(x) < \infty$ for all $\beta \ge 0$. The general Curie-Weiss model $\mathrm{CW}(\rho)$ at inverse temperature β is defined as the array of spin random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ with joint distribution

$$d\nu_n(\mathbf{x}) = Z_n^{-1} \exp\left(\frac{\beta}{2n}(x_1 + x_2 + \cdots + x_n)^2\right) \prod_{i=1}^n d\rho(x_i) \tag{2.3}$$

for $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ where

$$Z_n = \int \exp\left(\frac{\beta}{2n}(x_1 + x_2 + \cdots + x_n)^2\right) \prod_{i=1}^n d\rho(x_i)$$

is the normalizing constant. The magnetization $m(\mathbf{x})$ is defined as usual by $m(\mathbf{x}) = n^{-1}\sum_{i=1}^n x_i$. Here we will consider the case when ρ satisfies the following two conditions:

(A) ρ has compact support, that is, $\rho([-L, L]) = 1$ for some $L < \infty$.

(B) The equation $h'(s) = 0$ has a unique root at $s = 0$ where

$$h(s) := \frac{s^2}{2} - \log \int \exp(sx) \; d\rho(x) \text{ for } s \in \mathbb{R}.$$

The second condition says that $h(\cdot)$ has a unique global minima at $s = 0$ and $|h'(s)| > 0$ for $|s| > 0$. The behavior of this model is quite similar to the classical Curie-Weiss model and there is a phase transition at $\beta = 1$. For $\beta < 1$, $m(\mathbf{X})$ is concentrated around zero while for $\beta > 1, m(\mathbf{X})$ is bounded away from zero a.s. (see Ellis and Newman [42, 41]). We will prove the following concentration result.

Proposition 2.3.3. *Suppose $\mathbf{X} \sim \nu_n$ at the critical temperature $\beta = 1$ where ρ satisfies condition* (A) *and* (B). *Let k be such that $h^{(i)}(0) = 0$ for $0 \le i < 2k$ and $h^{(2k)}(0) \neq 0$, where*

$$h(s) := \frac{s^2}{2} - \log \int \exp(sx) \; d\rho(x) \textit{ for } s \in \mathbb{R}$$

and $h^{(i)}$ is the i-th derivative of h. Then, $k > 1$ and for any $n \geq 1$ and $t \geq 0$ the magnetization satisfies

$$\mathbb{P}(n^{1/2k}|m(\mathbf{X})| \geq t) \leq 2e^{-ct^{2k}}$$

where $c > 0$ is an absolute constant depending only on ρ.

Here we mention that in Ellis and Newman [42], convergence results were proved for the magnetization in CW(ρ) model under optimal condition on ρ. Under our assumption their result says that $n^{1/2k}m(\mathbf{X})$ converges weakly to a distribution having density proportional to $\exp(-\lambda x^{2k}/(2k)!)$ where $\lambda := h^{(2k)}(0)$. Hence the tail bound gives the correct convergence rate.

Let us now give a brief sketch of the proof of Proposition 2.3.1. Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at the critical temperature. We construct $\boldsymbol{\sigma}'$ by taking one step in the heat-bath Glauber dynamics: A coordinate I is chosen uniformly at random, and σ_I is replace by σ'_I drawn from the conditional distribution of the I-th coordinate given $\{\sigma_j : j \neq I\}$. Let

$$F(\boldsymbol{\sigma}, \boldsymbol{\sigma}') := \sum_{i=1}^{n}(\sigma_i - \sigma'_i) = \sigma_I - \sigma'_I.$$

For each $i = 1, 2, \ldots, n$, define $m_i = m_i(\boldsymbol{\sigma}) = n^{-1}\sum_{j\neq i}\sigma_j$. An easy computation gives that $\mathbb{E}(\sigma_i|\{\sigma_j, j \neq i\}) = \tanh(m_i)$ for all i and so we have

$$f(\boldsymbol{\sigma}) := \mathbb{E}(F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')|\boldsymbol{\sigma}) = m - \frac{1}{n}\sum_{i=1}^{n}\tanh(m_i) = \frac{m}{n} + \frac{1}{n}\sum_{i=1}^{n}g(m_i)$$

where $g(x) := x - \tanh(x)$. Note that $|m_i - m| \leq 1/n$, and hence $f(\boldsymbol{\sigma}) = m - \tanh m + O(1/n)$. A simple analytical argument using the fact that, for $x \approx 0$, $x - \tanh x = x^3/3 + O(x^5)$ then gives

$$\Delta(\boldsymbol{\sigma}) \leq \frac{6}{n}|f(\boldsymbol{\sigma})|^{2/3} + \frac{12}{n^{5/3}}$$

and using Corollary 2.2.3 with $\alpha = 2/3, B = 6/n$ and $C = 12/n^{5/3}$ we have

$$\mathbb{P}(|m - \tanh m| \geq t + n^{-1}) \leq \mathbb{P}(|f(\boldsymbol{\sigma})| \geq t) \leq 2e^{-cnt^{4/3}}$$

for all $t \geq 0$ for some constant $c > 0$. It is easy to see that this implies the result. The critical observation, of course, is that $x - \tanh(\beta x) = O(x^3)$ for $\beta = 1$, which is not true for $\beta \neq 1$.

2.3.2 Triangles in Erdős-Rényi graphs

Consider the Erdős-Rényi random graph model $G(n,p)$ which is defined as follows. The vertex set is $[n] := \{1, 2, \ldots, n\}$ and each edge (i,j), $1 \le i < j \le n$ is present with probability p and not present with probability $1-p$ independently of each other. For any three distinct vertex $i < j < k$ in $[n]$ we say that the triple (i,j,k) forms a triangle in the graph $G(n,p)$ if all the three edges $(i,j),(j,k),(i,k)$ are present in $G(n,p)$ (see figure 2.1). Let T_n be the number of triangles in $G(n,p)$, that is

$$T_n := \sum_{1\le i<j<k\le n} \mathbf{1}\{(i,j,k) \text{ forms a triangle in } G(n,p)\}.$$

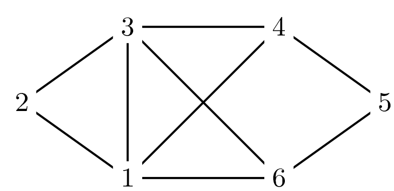

Figure 2.1: A graph with 3 triangles: $(1,2,3),(1,3,4)$ and $(1,3,6)$.

Let us define the function $I(\cdot,\cdot)$ on $(0,1)\times(0,1)$ as

$$I(r,s) := r\log\frac{r}{s} + (1-r)\log\frac{1-r}{1-s}. \tag{2.4}$$

Note that $I(r,s)$ is the Kullback-Leibler divergence of the measure ν_s from ν_r and also the relative entropy of ν_r w.r.t. ν_s where ν_p is the Bernoulli(p) measure. We have the following result about the large deviation rate function for the number of triangles in $G(n,p)$.

Theorem 2.3.4. *Let T_n be the number of triangles in $G(n,p)$, where $p > p_0$ where $p_0 = 2/(2+e^{3/2}) \approx 0.31$. Then for any $r \in (p,1]$,*

$$\mathbb{P}\left(T_n \ge \binom{n}{3} r^3\right) = \exp\left(-\frac{n^2 I(r,p)}{2}(1+O(n^{-1/2}))\right). \tag{2.5}$$

Moreover, even if $p \le p_0$, there exist p', p'' such that $p < p' \le p'' < 1$ and the same result holds for all $r \in (p,p') \cup (p'',1]$. For all p and r in the above domains, we also have the more precise estimate

$$\mathbb{P}\left(\left|T_n - \binom{n}{3} r^3\right| \le C(p,r) n^{5/2}\right) = \exp\left(-\frac{n^2 I(r,p)}{2}(1+O(n^{-1/2}))\right), \tag{2.6}$$

where $C(p,r)$ is a constant depending on p and r.

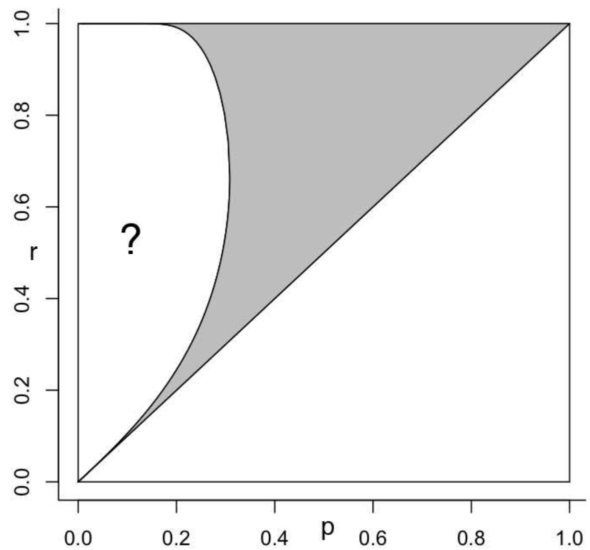

Figure 2.2: The set of $(p, r), r \geq p$ for which our large deviation result holds.

The behavior of the upper tail of subgraph counts in $G(n, p)$ is a problem of great interest in the theory of random graphs (see [17, 60, 62, 110, 70], and references contained therein). The best upper bounds to date were obtained by Kim and Vu [70] (triangles) and Janson, Oleszkiewicz, and Ruciński [61] (general subgraph counts). For triangles, the results of these papers essentially state that for a fixed $\epsilon > 0$,

$$\exp(-\Theta(n^2p^2 \log(1/p))) \leq \mathbb{P}(T_n \geq \mathbb{E}(T_n) + \epsilon n^3p^3) \leq \exp(-\Theta(n^2p^2)).$$

In a very recent development Chatterjee [29] proved that in the case of triangles, in fact, for any fixed $\epsilon > 0$,

$$\mathbb{P}(T_n \geq \mathbb{E}(T_n) + \epsilon n^3p^3) = \exp(-\Theta(n^2p^2 \log(1/p))).$$

Clearly, our result gives a lot more in the situations where it works (see Figure 2.2). The method of proof can be easily extended to prove similar results for general subgraph counts and are discussed in Subsection 2.3.3. However, there is an obvious incompleteness in Theorem 2.3.4 (and also for general subgraphs counts), namely, that it does not work for all (p, r).

In this context, we should mention that another paper on large deviations for subgraph counts by Bolthausen, Comets and Dembo [20] is in preparation. As of now, to the best of our knowledge, the authors of [20] have only looked at subgraphs that do not complete loops, like 2-stars. Another related article is the one by Döring and Eichelsbacher [38], who obtain moderate deviations for a class of graph-related objects, including triangles. Very recently using Szemerédi regularity lemma, Chatterjee and Varadhan [31] obtained the large deviation rate function in the full regime in an article in preparation.

Unlike the previous two examples, Theorem 2.3.4 is far from being a direct consequence of any of our abstract results. Therefore, let us give a sketch of the proof, which involves a new idea.

The first step is standard: consider tilted measures. However, the appropriate tilted measure in this case leads to what is known as an 'exponential random graph', a little studied object in the rigorous literature. Exponential random graphs have become popular in the statistical physics and network communities in recent years (see the survey of Park and Newman [90]). The only rigorous work we are aware of is the recent paper of Bhamidi et. al. [12], who look at convergence rates of Markov chains that generate such graphs.

We will not go into the general definition or properties of exponential random graphs. Let us only define the model we need for our purpose.

Fix two numbers $\beta \geq 0$ and $h \in \mathbb{R}$. Let $\Omega = \{0,1\}^{\binom{n}{2}}$ be the space of all tuples like $\mathbf{x} = (x_{ij})_{1\leq i<j\leq n}$, where $x_{ij} \in \{0,1\}$ for each i,j. Let $\mathbf{X} = (X_{ij})_{1\leq i<j\leq n}$ be a random element of Ω following the probability measure proportional to $e^{H(\mathbf{x})}$, where H is the Hamiltonian

$$H(\mathbf{x}) = \frac{\beta}{n} \sum_{1\leq i<j<k\leq n} x_{ij}x_{jk}x_{ik} \;+\; h \sum_{1\leq i<j\leq n} x_{ij}.$$

Note that any element of Ω naturally defines an undirected graph on a set of n vertices. For each $\mathbf{x} \in \Omega$, let $T(\mathbf{x}) = \sum_{i<j<k} x_{ij}x_{jk}x_{ik}$ denote the number of triangles in the graph defined by $\mathbf{x}$, and let $E(\mathbf{x}) = \sum_{i<j} x_{ij}$ denote the number of edges. Then the above Hamiltonian is nothing but

$$\frac{\beta T(\mathbf{x})}{n} + hE(\mathbf{x}).$$

For notational convenience we will assume that $x_{ij} = x_{ji}$. Let $Z_n(\beta, h)$ be the corresponding partition function, that is

$$Z_n(\beta, h) = \sum_{\mathbf{x}\in\Omega} e^{H(\mathbf{x})}.$$

Note that $\beta = 0$ corresponds to the Erdős-Rényi random graph with $p = e^h/(1+e^h)$. The following theorem 'solves' this model in a 'high temperature region'. Once this solution is known, the computation of the large deviation rate function is just one step away.

Theorem 2.3.5 (Free energy in high temperature regime)**.** *Suppose we have* $\beta \geq 0$*,* $h \in \mathbb{R}$*, and* $Z_n(\beta, h)$ *defined as above. Define a function* $\varphi : [0,1] \to \mathbb{R}$ *as*

$$\varphi(x) = \frac{e^{\beta x + h}}{1 + e^{\beta x + h}}.$$

Suppose β and h are such that the equation $u = \varphi(u)^2$ has a unique solution u^* in $[0,1]$ and $2\varphi(u^*)\varphi'(u^*) < 1$. Then

$$\lim_{n\to\infty} \frac{\log Z_n(\beta,h)}{n^2} = -\frac{1}{2}I(\varphi(u^*),\varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta\varphi(u^*)^3}{6},$$

where $I(\cdot,\cdot)$ is the function defined in (2.4). Moreover, there exists a constant $K(\beta,h)$ that depends only on β and h (and not on n) such that difference between the limit and $n^{-2}\log Z_n(\beta,h)$ is bounded by $K(\beta,h)n^{-1/2}$ for all n.

Incidentally, the above solution was obtained using physical heuristics by Park and Newman [91] in 2005. Here we mention that, in fact, the following result is always true.

Lemma 2.3.6. *For any $\beta \ge 0, h \in \mathbb{R}$ we have*

$$\begin{aligned}\liminf_{n\to\infty} \frac{\log Z_n(\beta,h)}{n^2} &\ge \sup_{r\in(0,1)} \left\{-\frac{1}{2}I(r,\varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta r^3}{6}\right\} \qquad (2.7)\\ &= \sup_{u:\varphi(u)^2=u} \left\{-\frac{1}{2}I(\varphi(u),\varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta\varphi(u)^3}{6}\right\}.\end{aligned}$$

We will characterize the set of β, h for which the conditions in Theorem 2.3.5 hold in Lemma 2.3.9. First of all, note that the appearance of the function $\varphi(u)^2 - u$ is not magical. For each $i < j$, define

$$L_{ij} = \frac{1}{n}\sum_{k\notin\{i,j\}} X_{ik}X_{jk}.$$

This is the number of 'wedges' or 2-stars in the graph that have the edge ij as base. The key idea is to use Theorem 2.2.1 to show that these quantities approximately satisfy the following set of 'mean field equations':

$$L_{ij} \simeq \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \text{ for all } i<j. \qquad (2.8)$$

(The idea of using Theorem 2.2.1 to prove mean field equations was initially developed in Section 3.4 of [24].) The following lemma makes this notion precise. Later, we will show that under the conditions of Theorem 2.3.5, this system has a unique solution.

Lemma 2.3.7 (Mean field equations)**.** *Let φ be defined as in Theorem 2.3.5. Then for any $1 \le i < j \le n$, we have*

$$\mathbb{P}\left(\sqrt{n}\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}}\varphi(L_{ik})\varphi(L_{jk})\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{8(1+\beta)}\right)$$

for all $t \geq 8\beta/n$. In particular we have

$$\mathbb{E}\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk})\right| \leq \frac{C(1+\beta)^{1/2}}{n^{1/2}} \tag{2.9}$$

where C is a universal constant.

In fact, one would expect that $L_{ij} \simeq u^*$ for all $i < j$, if the equation

$$\psi(u) := \varphi(u)^2 - u = 0 \tag{2.10}$$

has a unique solution u^* in $[0, 1]$. The intuition behind is as follows. Define $L_{\max} = \max_{i,j} L_{ij}$ and $L_{\min} = \min_{i,j} L_{ij}$. It is easy to see that φ is an increasing function. Hence from the mean-field equations (2.8) we have $L_{\max} \leq \varphi(L_{\max})^2 + o(1)$ or $\psi(L_{\max}) \geq o(1)$. But $\psi(u) \geq 0$ iff $u \leq u^*$. Hence $L_{\max} \leq u^* + o(1)$. Similarly we have $L_{\min} \geq u^* - o(1)$ and thus all $L_{ij} \simeq u^*$. Lemma 2.3.8 formalizes this idea. Here we mention that one can easily check that equation (2.10) has at most three solutions. Moreover, $\psi(0) > 0 > \psi(1)$ implies that $\psi'(u^*) \leq 0$ or $2\varphi(u^*)\varphi'(u^*) \leq 1$ if u^* is the unique solution to (2.10).

Lemma 2.3.8. *Let u^* be the unique solution of the equation $u = \varphi(u)^2$. Assume that $2\varphi(u^*)\varphi'(u^*) < 1$. Then for each $1 \leq i < j \leq n$, we have*

$$\mathbb{E}\,|L_{ij} - u^*| \leq \frac{K(\beta, h)}{n^{1/2}}$$

where $K(\beta, h)$ is a constant depending only on β, h. Moreover, if $2\varphi(u^*)\varphi'(u^*) = 1$ then we have

$$\mathbb{E}\,|L_{ij} - u^*| \leq \frac{K(\beta, h)}{n^{1/6}} \text{ for all } 1 \leq i < j \leq n.$$

Now observe that the Hamiltonian $H(\mathbf{X})$ can be written as

$$H(\mathbf{X}) = \frac{\beta}{6}\sum_{1\leq i<j\leq n} X_{ij}L_{ij} + h\sum_{1\leq i<j\leq n} X_{ij}.$$

The idea then is the following: once we know that the conclusion of Lemma 2.3.8 holds, each L_{ij} in the above Hamiltonian can be replaced by u^*, which results in a model where the coordinates are independent. The resulting probability measure is presumably quite different from the original measure, but somehow the partition functions remain comparable.

The following lemma (Lemma 2.3.9) characterizes the region $S \in \mathbb{R} \times [0, \infty)$ such that the equation $u = \varphi(u)^2$ has a unique solution u^* in $[0, 1]$ and $2\varphi(u^*)\varphi'(u^*) < 1$ for $(h, \beta) \in S$ (see figure 2.3).

Let $h_0 = \log 2 - \frac{3}{2} < 0$. For $h < h_0$ there exist exactly two solutions $0 < a_* = a_*(h) < 1/2 < a^* = a^*(h) < \infty$ to the equation

$$\log x + \frac{1+x}{2x} + h = 0.$$

Define $a_*(h) = a^*(h) = 1/2$ for $h = h_0$ and

$$\beta_*(h) = \frac{(1+a_*)^3}{2a_*} \text{ and } \beta^*(h) = \frac{(1+a^*)^3}{2a^*} \tag{2.11}$$

for $h \le h_0$.

Lemma 2.3.9 (Characterization of high temperature regime)**.** *Let S be the set of pairs (h, β) for which the function $\psi(u) := \varphi(u)^2 - u$ has a unique root u^* in $[0, 1]$ and $2\varphi(u^*)\varphi'(u^*) < 1$ where $\varphi(u) := e^{\beta u+h}/(1 + e^{\beta u+h})$. Then we have*

$$S^c = \{(h, \beta) : h \le h_0 \text{ and } \beta_*(h) \le \beta \le \beta^*(h)\}$$

where β^*, β_* are as given in equation (2.11). In particular, $(h, \beta) \in S$ if $\beta \le (3/2)^3$ or $h > h_0$.

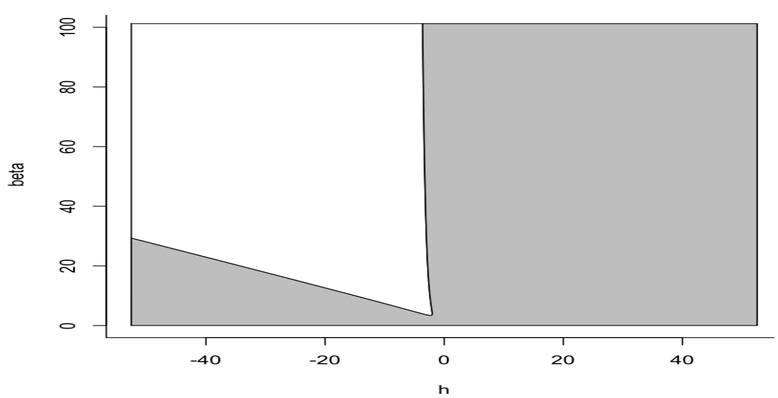

Figure 2.3: The set S of (h, β) for which the conditions of Theorem 2.3.5 hold.

Remark. The point $h = h_0, \beta = \beta_0 := (3/2)^3$ is the critical point and the curve

$$\gamma(t) = \left(-\log t - \frac{1+t}{2t}, \frac{(1+t)^3}{2t}\right) \tag{2.12}$$

for $t > 0$ is the phase transition curve. It corresponds to $\psi(u^*) = 0$ and $2\psi(u^*)\psi'(u^*) = 1$. In fact, at the critical point (h_0, β_0) the function $\psi(u) = \varphi(u)^2 - u$ has a unique root of order three at $u^* = 4/9$, *i.e.*, $\psi(u^*) = \psi'(u^*) = \psi''(u^*) = 0$ and $\psi'''(u^*) < 0$. The second part of lemma 2.3.8 shows that all the above conclusions (including the limiting free energy result) are true for the critical point but with an error rate of $n^{-1/6}$. Define the "energy" function

$$e(r) = \frac{1}{2}I(r, \varphi(0)) + \frac{1}{2}\log(1-\varphi(0)) - \frac{\beta r^3}{6}$$

appearing in of the r.h.s. of equation (2.7). The "high temperature" regime corresponds to the case when $e(\cdot)$ has a unique minima and no local maxima or saddle point. The critical point corresponds to the case when $e(\cdot)$ has a non-quadratic global minima. The boundary corresponds to the case when $e(\cdot)$ has a unique minima and a saddle point. In the "low temperature" regime $e(\cdot)$ has two local minima. In fact, one can easily check that there is a one dimensional curve inside the set S^c, starting from the critical point, on which $e(\cdot)$ has two global minima and outside one global minima. Below we provide the solution on the boundary curve. Unfortunately, as of now, we don't have a rigorous solution in the "low temperature" regime.

For (h, β) on the phase transition boundary curve (excluding the critical point) the function $\psi(\cdot)$ has two roots and one of them, say v^*, is an inflection point. Let u^* be the other root. Here we mention that u^* is a minima of $e(\cdot)$ while v^* is a saddle point of $e(\cdot)$. On the lower part of the boundary, which corresponds to $\{\gamma(t) : t < 1/2\}$, the inflection point $v^* = (1+t)^{-2}$ is larger than u^*, while on the upper part of the boundary corresponding to $\{\gamma(t) : t > 1/2\}$, the inflection point $v^* = (1+t)^{-2}$ is smaller than u^*. The following lemma "solves" the model at the boundary point $\gamma(t)$ (see eqn. 2.12).

Lemma 2.3.10. *Let $\gamma(\cdot), u^*, v^*$ be as above and $(h, \beta) = \gamma(t)$ for some $t \neq 1/2$. Then, for each $1 \leq i < j \leq n$, we have*

$$\mathbb{E}(|L_{ij} - u^*|) \leq \frac{K(\beta, h)}{n^{1/2}} \tag{2.13}$$

for some constant $K(\beta, h)$ depending on β, h. Moreover, we have

$$\frac{\log Z_n(\beta, h)}{n^2} = -\frac{1}{2}I(\varphi(u^*), \varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta\varphi(u^*)^3}{6} + O(n^{-1/2})$$

and

$$\begin{aligned} \mathbb{P}\left(\left|T_n(\mathbf{Y}) - \binom{n}{3}\varphi(u^*)^3\right| \leq C(\beta, h)n^{5/2}\right) \\ = \exp\left(-\frac{n^2 I(\varphi(u^*), \varphi(0))}{2}(1 + O(n^{-1/2}))\right), \end{aligned} \tag{2.14}$$

where $\mathbf{Y} = ((Y_{ij}))_{i<j}$ follows $G(n, \varphi(0))$ and the constant appearing in $O(\cdot)$ and $C(\beta, h)$ depend only on β, h.

In the next subsection we will briefly discuss about the results for general subgraph counts that can be proved using similar ideas.

2.3.3 General subgraph counts

Let $F = (V(F), E(F))$ be a fixed finite graph on $\boldsymbol{v}_F := |V(F)|$ many vertices with $\boldsymbol{e}_F := |E(F)|$ many edges. Without loss of generality we will assume that $V(F) = [\boldsymbol{v}_F] := \{1, 2, \dots, \boldsymbol{v}_F\}$. Let $\alpha_F = |\mathrm{Aut}(F)|$ be the number of graph automorphism of the graph F. Let N_n be the number of copies of F, not necessarily induced, in the Erdős-Rényi random graph $G(n,p)$ (so the number of 2-stars in a triangle will be three). We have the following result about the large deviation rate function for the random variable N_n.

Theorem 2.3.11. *Let N_n be the number of copies of F in $G(n,p)$, where*

$$p > p_0 := \frac{\boldsymbol{e}_F - 1}{\boldsymbol{e}_F - 1 + \exp\left(\frac{\boldsymbol{e}_F}{\boldsymbol{e}_F - 1}\right)}.$$

Then for any $r \in (p, 1]$,

$$\mathbb{P}\left(N_n \geq \frac{\boldsymbol{v}_F!}{\alpha_F}\binom{n}{\boldsymbol{v}_F} r^{\boldsymbol{e}_F}\right) = \exp\left(-\frac{n^2 I(r,p)}{2}(1 + O(n^{-1/2}))\right). \tag{2.15}$$

Moreover, even if $p \leq p_0$, there exist p', p'' such that $p < p' \leq p'' < 1$ and the same result holds for all $r \in (p, p') \cup (p'', 1]$. For all p and r in the above domains, we also have the more precise estimate

$$\begin{aligned}\mathbb{P}\left(\left|N_n - \frac{\boldsymbol{v}_F!}{\alpha_F}\binom{n}{\boldsymbol{v}_F} r^{\boldsymbol{e}_F}\right| \leq C(p,r) n^{\boldsymbol{v}_F - 1/2}\right)\\ = \exp\left(-\frac{n^2 I(r,p)}{2}(1 + O(n^{-1/2}))\right),\end{aligned}$$

where $C(p,r)$ is a constant depending on p and r.

Note that p_0 as a function of $\boldsymbol{e}_F$ is increasing and converges to 1 as number of edges goes to infinity (see Figure 2.4). So there is an obvious gap in the large deviation result, namely the proof does not work when $r \geq p,\ p \leq p_0$ and the gap becomes larger as the number of edges in F increases. Note that $p_0 \to 1$ as $\boldsymbol{e}_F \to \infty$.

The proof of Theorem 2.3.11 uses the same arguments that were used in the triangle case. Here the tilted measure leads to an exponential random graph model where the Hamiltonian depends on number of copies of F in the random graph. Let $\beta \geq 0, h \in \mathbb{R}$ be two fixed numbers. As before we will identify elements of $\Omega := \{0,1\}^{\binom{n}{2}}$ with undirected graphs on a set of n vertices. For each $\mathbf{x} \in \Omega$, let $N(\mathbf{x})$ denote the number of copies of F in the graph defined by $\mathbf{x}$, and let $E(\mathbf{x}) = \sum_{i<j} x_{ij}$ denote the number of edges. Let $\mathbf{X} = (X_{ij})_{1 \leq i < j \leq n}$ be a random element of Ω following the probability measure proportional to $e^{H(\mathbf{x})}$, where H is the Hamiltonian

$$H(\mathbf{x}) = \frac{\beta}{(n-2)_{\boldsymbol{v}_F - 2}} N(\mathbf{x}) + hE(\mathbf{x}).$$

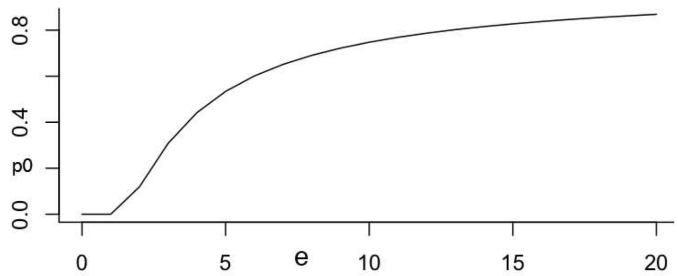

Figure 2.4: The curve $p_0(\boldsymbol{e}_F)$ above which our large deviation result holds.

where $(n)_m = \frac{n!}{(n-m)!}$. Recall that $\boldsymbol{v}_F$ is the number of vertices in the graph F. The scaling was done to make the two summands comparable. Also we used $(n-2)_{\boldsymbol{v}_F-2}$ instead of $n^{\boldsymbol{v}_F}$ to make calculations simpler. Let $Z_n(\beta, h)$ be the partition function. Note that $N(\mathbf{x})$ can be written as

$$N(\mathbf{x}) = \frac{1}{\alpha_F} \sum_{\substack{1 \le t_1, t_2, \ldots, t_{\boldsymbol{v}_F} \le n, \\ t_i \neq t_j \text{ for } i \neq j}} \prod_{(i,j) \in E(F)} x_{t_i t_j}. \tag{2.16}$$

For $\mathbf{x} \in \Omega, 1 \le i < j \le n$, define $\mathbf{x}^1_{(i,j)}$ as the element of Ω which is same as $\mathbf{x}$ in every coordinate except for the (i,j)-th coordinate where the value is 1. Similarly define $\mathbf{x}^0_{(i,j)}$. For $i < j$, define the random variable

$$L_{ij} := \frac{N(\mathbf{X}^1_{(i,j)}) - N(\mathbf{X}^0_{(i,j)})}{(n-2)_{\boldsymbol{v}_F-2}}.$$

The main idea is as in the triangle case. We show that L_{ij}'s satisfy a system of "mean-field equations" similar to (2.8) which has a unique solution under the condition of Theorem 2.3.12. In fact, we will show that L_{ij}" $\approx$ "u^* for all $i < j$ and $E(\mathbf{X})$" $\approx$ "$\binom{n}{2}\varphi(u^*)$ under the condition of Theorem 2.3.12. Now note that we can write the hamiltonian as

$$H(\mathbf{X}) = \frac{\beta}{\boldsymbol{e}_F} \sum_{i<j} X_{ij} L_{ij} + h \sum_{i<j} X_{ij}$$

which is approximately equal to $h^* E(\mathbf{X})$ where $h^* = h + \beta u^* / \boldsymbol{e}_F$. Now the remaining is a calculus exercise.

So the first step in proving the large deviation bound is the following theorem, which gives the limiting free energy in the "high temperature" regime. Note the similarity with the triangle case.

Theorem 2.3.12. *Suppose we have $\beta \ge 0$, $h \in \mathbb{R}$, and $Z_n(\beta, h)$ defined as above. Define a function $\varphi : [0,1] \to \mathbb{R}$ as*

$$\varphi(x) = \frac{e^{\beta x + h}}{1 + e^{\beta x + h}}.$$

Suppose β *and* h *are such that the equation* $\alpha_F u = 2\boldsymbol{e}_F \varphi(u)^{\boldsymbol{e}_F - 1}$ *has a unique solution* u^* *in* $[0,1]$ *and* $2\boldsymbol{e}_F(\boldsymbol{e}_F - 1)\varphi(u^*)^{\boldsymbol{e}_F-2}\varphi'(u^*) < \alpha_F$*. Then*

$$\lim_{n\to\infty} \frac{\log Z_n(\beta,h)}{n^2} = -\frac{1}{2} I(\varphi(u^*), \varphi(0)) - \frac{1}{2}\log(1-\varphi(0)) + \frac{\beta\varphi(u^*)^{\boldsymbol{e}_F}}{\alpha_F},$$

where $I(\cdot,\cdot)$ *is the function defined in* (2.4)*. Moreover, there exists a constant* $K(\beta,h)$ *that depends only on* β *and* h *(and not on* n*) such that difference between* $n^{-2}\log Z_n(\beta,h)$ *and the limit is bounded by* $K(\beta,h)n^{-1/2}$ *for all* n*.*

Here also we can identify the region where the conditions in Theorem 2.3.12 hold. Let

$$h_0 = \log(\boldsymbol{e}_F - 1) - \frac{\boldsymbol{e}_F}{\boldsymbol{e}_F - 1}. \tag{2.17}$$

For $h < h_0$ there exist exactly two solutions $0 < a_* = a_*(h) < 1/2 < a^* = a^*(h) < \infty$ of the equation

$$\log x + \frac{1+x}{(\boldsymbol{e}_F - 1)x} + h = 0$$

Define $a_*(h) = a^*(h) = 1/(\boldsymbol{e}_F - 1)$ for $h = h_0$ and

$$\beta_*(h) = \frac{\alpha_F(1+a_*)^{\boldsymbol{e}_F}}{2\boldsymbol{e}_F(\boldsymbol{e}_F-1)a_*} \text{ and } \beta^*(h) = \frac{\alpha_F(1+a^*)^{\boldsymbol{e}_F}}{2\boldsymbol{e}_F(\boldsymbol{e}_F-1)a^*} \tag{2.18}$$

for $h \le h_0$.

Lemma 2.3.13. *Let* S *be the set of pairs* (h,β) *for which the function*

$$\psi(u) := 2\boldsymbol{e}_F\varphi(u)^{\boldsymbol{e}_F-1} - \alpha_F u$$

has a unique root u^* *in* $[0,1]$ *and* $2\boldsymbol{e}_F(\boldsymbol{e}_F-1)\varphi(u^*)^{\boldsymbol{e}_F-2}\varphi'(u^*) < \alpha_F$ *where* $\varphi(u) := e^{\beta u+h}/(1+e^{\beta u+h})$*. Then we have*

$$S^c = \{(h,\beta) : h \le h_0 \text{ and } \beta_*(h) \le \beta \le \beta^*(h)\}$$

where h_0, β^*, β_* *are as given in equations* (2.17), (2.18)*. In particular,* $(h,\beta) \in S$ *if*

$$\beta \le \frac{\alpha_F \boldsymbol{e}_F^{\boldsymbol{e}_F-1}}{2(\boldsymbol{e}_F-1)^{\boldsymbol{e}_F}} \text{ or } h > h_0.$$

In fact Lemma 2.3.13 identifies the critical point and the phase transition curve where the model goes from ordered phase to a disordered phase. But the results above does not say what happens at the boundary or in the low temperature regime. However note that the mean-field equations hold for all values of β and h.

2.3.4 Ising model on $\mathbb{Z}^d$

Fix any $\beta \geq 0, h \in \mathbb{R}$ and an integer $d \geq 1$. Also fix $n \geq 2$. Let $\mathbb{B} = \{1, 2, \ldots, n+1\}^d$ be a hypercube with $(n+1)^d$ many points in the d-dimensional hypercube lattice $\mathbb{Z}^d$. Let Ω be the graph obtained from $\mathbb{B}$ by identifying the opposite boundary points, *i.e.,* for $x = (x_1, x_2, \ldots, x_d), y = (y_1, y_2, \ldots, y_d) \in \mathbb{B}$ we have x is identified with y if $x_i - y_i \in \{-n, 0, n\}$ for all i. This identification is known in the literature as periodic boundary condition. Note that Ω is the d-dimensional lattice torus with linear size n. We will write $x \sim y$ for $x, y \in \Omega$ if x, y are nearest neighbors in Ω. Also let us denote by N_x the set of nearest neighbors of x in Ω, *i.e.,* $N_x = \{y \in \Omega : y \sim x\}$.

Now consider the Gibbs measure on $\{+1, -1\}^\Omega$ given by the following Hamiltonian

$$H(\boldsymbol{\sigma}) := \beta \sum_{x \sim y, x, y \in \Omega} \sigma_x \sigma_y + h \sum_{x \in \Omega} \sigma_x$$

where $\boldsymbol{\sigma} = (\sigma_x)_{x \in \Omega}$ is a typical element of $\{+1, -1\}^\Omega$. So the probability of a configuration $\boldsymbol{\sigma} \in \{+1, -1\}^\Omega$ is

$$\mu_{\beta,h}(\{\boldsymbol{\sigma}\}) := Z_{\beta,h}^{-1} \exp\left(H(\boldsymbol{\sigma})\right) = Z_{\beta,h}^{-1} \exp\left(\beta \sum_{x \sim y, x, y \in \Omega} \sigma_x \sigma_y + h \sum_{x \in \Omega} \sigma_x\right) \tag{2.19}$$

where $Z_{\beta,h} = \sum_{\boldsymbol{\sigma} \in \{+1,-1\}^\Omega} e^{H(\boldsymbol{\sigma})}$ is the normalizing constant. Here σ_x is the spin of the magnetic particle at position x in the discrete torus Ω. This is the famous Ising model of ferromagnetism on the box $\mathbb{B}$ with periodic boundary condition at inverse temperature β and external field h.

The one-dimensional Ising model is probably the first statistical model of ferromagnetism to be proposed or analyzed [58]. The model exhibits no phase transition in one dimension. But for dimensions two and above the Ising ferromagnet undergoes a transition from an ordered to a disordered phase as β crosses a critical value. The two dimensional Ising model with no external field was first solved by Lars Onsager in a ground breaking paper [89], who also calculated the critical β as $\beta_c = \sinh^{-1}(1)$. For dimensions three and above the model is yet to be solved, and indeed, very few rigorous results are known.

In this subsection, we present some concentration inequalities for the Ising model that hold for all values of β. These 'temperature-free' relations are analogous to the mean field equations that we obtained for subgraph counts earlier.

The magnetization of the system, as a function of the configuration $\boldsymbol{\sigma}$, is defined as $m(\boldsymbol{\sigma}) := \frac{1}{|\Omega|} \sum_{x \in \Omega} \sigma_x$. For each integer $k \in \{1, 2, \ldots, 2d\}$, define a degree k polynomial function $r_k(\boldsymbol{\sigma})$ of a spin configuration $\boldsymbol{\sigma}$ as follows:

$$r_k(\boldsymbol{\sigma}) := \left(\binom{2d}{k} |\Omega|\right)^{-1} \sum_{x \in \Omega} \sum_{S \subseteq N_x, |S| = k} \sigma_S \tag{2.20}$$

where $\sigma_S = \prod_{x\in S}\sigma_x$ for any $S \subseteq \Omega$. In particular $r_k(\boldsymbol{\sigma})$ is the average of the product of spins of all possible k out of $2d$ neighbors. Note that $r_1(\boldsymbol{\sigma}) \equiv m(\boldsymbol{\sigma})$. We will show that when $h = 0$ and n is large, $m(\boldsymbol{\sigma})$ and $r_k(\boldsymbol{\sigma})$'s satisfy the following "mean-field relation" with high probability under the Gibbs measure:

$$(1-\theta_0(\beta))m(\boldsymbol{\sigma}) \approx \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma}). \tag{2.21}$$

These relations hold for all values of $\beta \geq 0$. Here θ_k's are explicit rational functions of $\tanh(2\beta)$ for $k = 0, 1, \ldots, d-1$, defined in equation (2.22) below. (Later we will prove in Proposition 2.3.16 that an external magnetic field h will add an extra linear term in the above relation (2.21).) The following Proposition makes this notion precise in terms of finite sample tail bound. It is a simple consequence of Theorem 2.2.1.

Theorem 2.3.14. *Suppose $\boldsymbol{\sigma}$ is drawn from the Gibbs measure $\mu_{\beta,0}$. Then, for any $\beta \geq 0, n \geq 1$ and $t \geq 0$ we have*

$$\mathbb{P}\left(\sqrt{|\Omega|}\left|(1-\theta_0(\beta))m(\boldsymbol{\sigma}) - \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma})\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{4b(\beta)}\right)$$

where $m(\boldsymbol{\sigma}) := \frac{1}{|\Omega|}\sum_{x\in\Omega}\sigma_x$ is the magnetization, $r_k(\boldsymbol{\sigma})$ is as given in (2.20) and for $k = 0, 1, \ldots, d-1$

$$\begin{aligned}\theta_k(\beta) &= \frac{1}{4^d}\binom{2d}{2k+1}\sum_{\boldsymbol{\sigma}\in\{-1,+1\}^{2d}}\tanh\left(\beta\sum_{i=1}^{2d}\sigma_i\right)\prod_{j=1}^{2k+1}\sigma_j \\ \text{and } b(\beta) &= |1-\theta_0(\beta)| + \sum_{k=1}^{d-1}(2k+1)|\theta_k(\beta)|.\end{aligned} \tag{2.22}$$

Moreover, we can explicitly write down $\theta_0(\beta)$ as

$$\theta_0(\beta) = \frac{1}{4^{d-1}}\sum_{k=1}^{d}k\binom{2d}{d+k}\tanh(2k\beta)$$

and for $d \geq 2$ there exists $\beta_1 \in (0,\infty)$, depending on d, such that $1-\theta_0(\beta) > 0$ for $\beta < \beta_1$ and $1-\theta_0(\beta) < 0$ for $\beta > \beta_1$.

Here we may remark that for any fixed k, $\theta_k(\beta/2d)$ converges to the coefficient of x^{2k+1} in the power series expansion of $\tanh(\beta x)$ and $2d\beta_1(d) \downarrow 1$ as $d \to \infty$. For small values of d we can explicitly calculate the θ_k's. For instance, in $d = 2$,

$$\theta_0(\beta) = \frac{1}{2}\left(\tanh(4\beta) + 2\tanh(2\beta)\right),\ \theta_1(\beta) = \frac{1}{2}\left(\tanh(4\beta) - 2\tanh(2\beta)\right).$$

For $d = 3$,

$$\begin{aligned}
\theta_0(\beta) &= \frac{3}{16}\left(\tanh(6\beta) + 4\tanh(4\beta) + 5\tanh(2\beta)\right), \\
\theta_1(\beta) &= \frac{10}{16}\left(\tanh(6\beta) - 3\tanh(2\beta)\right), \\
\theta_2(\beta) &= \frac{3}{16}\left(\tanh(6\beta) - 4\tanh(4\beta) + 5\tanh(2\beta)\right).
\end{aligned}$$

For $d = 4$,

$$\begin{aligned}
\theta_0(\beta) &= \frac{1}{16}\left(\tanh(8\beta) + 6\tanh(6\beta) + 14\tanh(4\beta) + 14\tanh(2\beta)\right), \\
\theta_1(\beta) &= \frac{7}{16}\left(\tanh(8\beta) + 2\tanh(6\beta) - 2\tanh(4\beta) - 6\tanh(2\beta)\right), \\
\theta_2(\beta) &= \frac{7}{16}\left(\tanh(8\beta) - 2\tanh(6\beta) - 2\tanh(4\beta) + 6\tanh(2\beta)\right), \\
\theta_3(\beta) &= \frac{1}{16}\left(\tanh(8\beta) - 6\tanh(6\beta) + 14\tanh(4\beta) - 14\tanh(2\beta)\right).
\end{aligned}$$

Corollary 2.3.15. *For the Ising model on Ω at inverse temperature β with no external magnetic field for all $t \geq 0$ we have,*

(i) *if $d = 1$,*

$$\mathbb{P}(|m(\boldsymbol{\sigma})| \geq t) \leq 2\exp\left(-\frac{1}{4}|\Omega|(1 - \tanh(2\beta))t^2\right)$$

(ii) *if $d = 2$,*

$$\mathbb{P}(|[(1-u)^2 - u^3]m(\boldsymbol{\sigma}) + u^3 r_3(\boldsymbol{\sigma})| \geq t) \leq 2\exp\left(-\frac{|\Omega|t^2}{32}\right)$$

where $u = \tanh(2\beta)$ and $r_3(\boldsymbol{\sigma}) = \frac{1}{4|\Omega|}\sum^* \sigma_x\sigma_y\sigma_z$ where the sum $\sum^*$ is over all $x, y, z \in \Omega$ such that $|x - y| = 2, |z - y| = 2, |x - z| = 2$.

(iii) *if $d = 3$,*

$$\mathbb{P}(|g(u)m(\boldsymbol{\sigma}) + 5u^3(1 + u^2)r_3(\boldsymbol{\sigma}) - 3u^5 r_5(\boldsymbol{\sigma})| \geq t) \leq 2\exp\left(-c|\Omega|t^2\right)$$

where c is an absolute constant, $g(u) = 1 - 3u + 4u^2 - 9u^3 + 3u^4 - 3u^5$, $u = \tanh(2\beta)$ and r_3, r_5 are as defined in (2.20).

Although we do not yet know the significance of the above relations, it seems somewhat striking that they are not affected by phase transitions. The exponential tail bounds show that many such relations can hold simultaneously. For completeness, we state below the corresponding result for nonzero external field.

Proposition 2.3.16. *Suppose $\boldsymbol{\sigma}$ is drawn from the Gibbs measure $\mu_{\beta,h}$. Let $r_k(\boldsymbol{\sigma})$, $\theta_k(\beta)$, $b(\beta)$ be as in proposition* (2.3.14). *Then, for any $\beta \geq 0, h \in \mathbb{R}, n \geq 1$ and $t \geq 0$ we have*

$$\mathbb{P}\left(|(1-\theta_0(\beta))m(\boldsymbol{\sigma}) - g(\boldsymbol{\sigma})| \geq t\right) \leq 2\exp\left(-\frac{|\Omega|t^2}{4b(\beta)(1+\tanh|h|)}\right) \tag{2.23}$$

where

$$g(\boldsymbol{\sigma}) := \sum_{k=1}^{d-1} \theta_k(\beta) r_{2k+1}(\boldsymbol{\sigma}) + \tanh(h)\left(1 - \sum_{k=0}^{d-1}\theta_k(\beta)s_{2k+1}(\boldsymbol{\sigma})\right)$$

and

$$s_k(\boldsymbol{\sigma}) := \left(\binom{2d}{k}|\Omega|\right)^{-1} \sum_{x\in\Omega}\sum_{S\subseteq N_x, |S|=k} \sigma_{S\cup\{x\}}$$

is the average of products of spins over all k-stars for $k = 1, 2, \ldots, 2d$ and Ω is the discrete torus in $\mathbb{Z}^d$ with n^d many points.

2.4 Proofs

Instead of proving Theorem 2.2.2 first, let us see how it is applied to prove the result for the Curie-Weiss model at critical temperature. The proof is simply an elaboration of the sketch given at the end of Subsection 2.3.1.

Proof of Proposition 2.3.1. Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at critical temperature. We construct $\boldsymbol{\sigma}'$ by taking one step in the heat-bath Glauber dynamics: A coordinate I is chosen uniformly at random, and σ_I is replace by σ'_I drawn from the conditional distribution of the I-th coordinate given $\{\sigma_j : j \neq I\}$. Let

$$F(\boldsymbol{\sigma}, \boldsymbol{\sigma}') := \sum_{i=1}^{n}(\sigma_i - \sigma'_i) = \sigma_I - \sigma'_I.$$

For each $i = 1, 2, \ldots, n$, define $m_i = m_i(\boldsymbol{\sigma}) = n^{-1}\sum_{j\neq i}\sigma_j$. An easy computaion gives that $\mathbb{E}(\sigma_i|\{\sigma_j, j \neq i\}) = \tanh(m_i)$ for all i and so we have

$$f(\boldsymbol{\sigma}) := \mathbb{E}(F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')|\boldsymbol{\sigma}) = m - \frac{1}{n}\sum_{i=1}^{n}\tanh(m_i) = \frac{m}{n} + \frac{1}{n}\sum_{i=1}^{n} g(m_i)$$

where $g(x) := x - \tanh(x)$. By definition $m_i(\boldsymbol{\sigma}) - m(\boldsymbol{\sigma}) = \sigma_i/n$ and $m_i(\boldsymbol{\sigma}') - m(\boldsymbol{\sigma}) = (\sigma_i + \sigma_I - \sigma'_I)/n$ for all i. Hence using Taylor expansion upto first degree and noting that $|g'(x)| = \tanh^2(x) \le x^2$ we have

$$\begin{aligned}|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| &\le \frac{2}{n}|g'(m(\boldsymbol{\sigma}))| + \frac{2 + 5\max_{|x|\le 1}|g''(x)|}{n^2} \\ &\le \frac{2}{n}m(\boldsymbol{\sigma})^2 + \frac{6}{n^2}.\end{aligned}$$

Clearly $|F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')| \le 2$. Thus we have

$$\Delta(\boldsymbol{\sigma}) := \frac{1}{2}\,\mathbb{E}[|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| \cdot |F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')| \mid \boldsymbol{\sigma}] \le \frac{2}{n}m(\boldsymbol{\sigma})^2 + \frac{6}{n^2}.$$

Now it is easy to verify that $|x|^3 \le 5|x - \tanh x|$ for all $|x| \le 1$. Note that this is the place where we need $\beta = 1$. For $\beta \ne 1$, the linear term dominates in $m - \tanh(\beta m)$. Hence it follows that

$$m(\boldsymbol{\sigma})^2 \le 5^{2/3}|m(\boldsymbol{\sigma}) - \tanh m(\boldsymbol{\sigma})|^{2/3} \le 3|f(\boldsymbol{\sigma})|^{2/3} + 3n^{-2/3}$$

where in the last line we used the fact that $|f(\boldsymbol{\sigma}) - (m - \tanh m)| \le 1/n$ and $5^{2/3} < 3$. Thus

$$\Delta(\boldsymbol{\sigma}) \le \frac{6}{n}|f(\boldsymbol{\sigma})|^{2/3} + \frac{12}{n^{5/3}}$$

and using Corollary 2.2.3 with $\alpha = 2/3, B = 6/n$ and $C = 12/n^{5/3}$ we have

$$\mathbb{P}(|m - \tanh m| \ge t + n^{-1}) \le \mathbb{P}(|f(\boldsymbol{\sigma})| \ge t) \le 2e^{-cnt^{4/3}}$$

for all $t \ge 0$ for some constant $c > 0$. This clearly implies that

$$\mathbb{P}(|m| \ge t) \le \mathbb{P}(|m - \tanh m| \ge t^3/5) \le 2e^{-cnt^4}$$

for all $t \ge 0$ and for some absolute constant $c > 0$. Thus we are done. □

Proof of Proposition 2.3.3. The proof is along the lines of proof of proposition 2.3.1. Suppose $\mathbf{X}$ is drawn from the distribution ν_n. We construct $\mathbf{X}'$ as follows: a coordinate I is chosen uniformly at random, and X_I is replace by X_I' drawn from the conditional distribution of the I-th coordinate given $\{X_j : j \ne I\}$. Let

$$F(\mathbf{X}, \mathbf{X}') := \sum_{i=1}^{n}(X_i - X_i') = X_I - X_I'.$$

For each $i = 1, 2, \ldots, n$, define $m_i(\mathbf{X}) = n^{-1}\sum_{j\ne i} X_j$. An easy computaion gives that $\mathbb{E}(X_i|\{X_j, j \ne i\}) = g(m_i)$ for all $i = 1, 2, \ldots, n$ where $g(s) = \frac{d}{ds}(\log\int\exp(x^2/2n + sx)\,d\rho(x))$ for $s \in \mathbb{R}$. So we have

$$f(\mathbf{X}) := \mathbb{E}(F(\mathbf{X}, \mathbf{X}')|\mathbf{X}) = m(\mathbf{X}) - \frac{1}{n}\sum_{i=1}^{n} g(m_i(\mathbf{X})).$$

Define the function

$$h(s) = \frac{s^2}{2} - \log \int \exp(sx)\, d\rho(x) \text{ for } s \in \mathbb{R}. \tag{2.24}$$

Clearly h is an even function. Recall that k is an integer such that $h^{(i)}(0) = 0$ for $0 \le i < 2k$ and $h^{(2k)}(0) \neq 0$. We have $k \ge 2$ since $h''(0) = 1 - \int x^2\, d\rho(x) = 0$.

Now using the fact that $\rho([-L, L]) = 1$ it is easy to see that $|f(\mathbf{X}) - h'(m(\mathbf{X}))| \le c/n$ for some constant c depending on L only. In the subsequent calculations c will always denote a constant depending only on L that may vary from line to line. Similarly we have

$$\begin{aligned} |f(\mathbf{X}) - f(\mathbf{X}')| &\le \frac{|X_I - X_I'|}{n}\left(|1 - g'(m(\mathbf{X}))| + \frac{c(1 + \sup_{|x|\le L} |g''(x)|)}{n}\right) \\ &\le \frac{2L}{n}|h''(m(\mathbf{X}))| + \frac{c}{n^2}. \end{aligned}$$

Note that $|h''(s)| \le cs^{2k-2}$ for some constant c for all $s \ge 0$. This follows since $\lim_{s\to 0} h''(s)/s^{2k-2}$ exists and $h''(\cdot)$ is a bounded function. Also $\lim_{s\to 0} |h'(s)|/|s|^{2k-1} = |h^{(2k)}(0)| \neq 0$ and $|h'(s)| > 0$ for $s > 0$. So we have $|h'(s)| \ge c|s|^{2k-1}$ for some constant $c > 0$ and all $|s| \le L$. From the above results we deduce that

$$\begin{aligned} |f(\mathbf{X}) - f(\mathbf{X}')| \le \frac{c}{n}|(m(\mathbf{X}))|^{2k-2} + \frac{c}{n^2} &\le \frac{c}{n}|h'(m(\mathbf{X}))|^{\frac{2k-2}{2k-1}} + \frac{c}{n^2} \\ &\le \frac{c}{n}|f(\mathbf{X})|^{\frac{2k-2}{2k-1}} + \frac{c}{n^{2-1/(2k-1)}}. \end{aligned}$$

Now the rest of the proof follows exactly as for the classical Curie-Weiss model. □

2.4.1 Proof of the large deviation result for triangles

First, let us state and prove a simple technical lemma.

Lemma 2.4.1. *Let $x_1, \ldots, x_k, y_1, \ldots, y_k$ be real numbers. Then*

$$\max_{1\le i\le n}\left|\frac{e^{x_i}}{\sum_{j=1}^k e^{x_j}} - \frac{e^{y_i}}{\sum_{j=1}^k e^{y_j}}\right| \le 2 \max_{1\le i\le n} |x_i - y_i|.$$

and

$$\left|\log \sum_{i=1}^k e^{x_i} - \log \sum_{i=1}^k e^{y_i}\right| \le \max_{1\le i\le k} |x_i - y_i|.$$

Proof. Fix $1 \le i \le k$. For $t \in [0, 1]$, let

$$h(t) = \frac{e^{tx_i + (1-t)y_i}}{\sum_{j=1}^k e^{tx_j + (1-t)y_j}}.$$

Then

$$h'(t) = \left[(x_i - y_i) - \frac{\sum_{j=1}^k (x_j - y_j)e^{tx_j+(1-t)y_j}}{\sum_{j=1}^k e^{tx_j+(1-t)y_j}}\right]h(t).$$

This shows that $|h'(t)| \le 2\max_i |x_i - y_i|$ for all $t \in [0,1]$ and completes the proof of the first assertion. The second inequality is proved similarly. □

Proof of Lemma 2.3.7. Fix two numbers $1 \le i < j \le n$. Given a configuration $\mathbf{X}$, construct another configuration $\mathbf{X}'$ as follows. Choose a point $k \in \{1,\ldots,n\}\backslash\{i,j\}$ uniformly at random, and replace the pair (X_{ik}, X_{jk}) with (X'_{ik}, X'_{jk}) drawn from the conditional distribution given the rest of the edges. Let L'_{ij} be the revised value of L_{ij}. From the form of the Hamiltonian it is now easy to read off that for $x, y \in \{0,1\}$,

$$\begin{aligned}&\mathbb{P}(X'_{ik} = x, X'_{jk} = y \mid \mathbf{X})\\ &\propto \exp\left(\beta x L_{ik} + \beta y L_{jk} + hx + hy - \frac{\beta}{n}xX_{ij}X_{jk} - \frac{\beta}{n}yX_{ij}X_{ik} + \frac{\beta}{n}xyX_{ij}\right).\end{aligned}$$

An application of Lemma 2.4.1 shows that the terms having β/n as coefficient can be ‘ignored’ in the sense that for each $x, y \in \{0,1\}$,

$$\left|\mathbb{P}(X'_{ik} = x, X'_{jk} = y \mid \mathbf{X}) - \frac{e^{\beta x L_{ik} + \beta y L_{jk} + hx + hy}}{(1+e^{\beta L_{ik}+h})(1+e^{\beta L_{jk}+h})}\right| \le \frac{2\beta}{n}$$

In particular,

$$|\,\mathbb{E}(X'_{ik}X'_{jk} \mid \mathbf{X}) - \varphi(L_{ik})\varphi(L_{jk})| \le \frac{2\beta}{n}. \tag{2.25}$$

Now,

$$\begin{aligned}\mathbb{E}(L_{ij} - L'_{ij} \mid \mathbf{X}) &= \frac{1}{n(n-2)}\sum_{k\notin\{i,j\}}(X_{ik}X_{jk} - \mathbb{E}(X'_{ik}X'_{jk} \mid \mathbf{X}))\\ &= \frac{1}{n-2}L_{ij} - \frac{1}{n(n-2)}\sum_{k\notin\{i,j\}}\mathbb{E}(X'_{ik}X'_{jk} \mid \mathbf{X}).\end{aligned} \tag{2.26}$$

Let $F(\mathbf{X},\mathbf{X}') = (n-2)(L_{ij} - L'_{ij})$ and $f(\mathbf{X}) = \mathbb{E}(F(\mathbf{X},\mathbf{X}') \mid \mathbf{X})$. Let

$$g(\mathbf{X}) = L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}}\varphi(L_{ik})\varphi(L_{jk}).$$

From (2.25) and (2.26) it follows that

$$|f(\mathbf{X}) - g(\mathbf{X})| \le \frac{2\beta}{n}. \tag{2.27}$$

Since X' has the same distribution as X, the same bound holds for $|f(X') - g(X')|$ as well. Now clearly, $|F(X, X')| \leq 1$. Again, $|g(X) - g(X')| \leq 2/n$, and therefore

$$|f(X) - f(X')| \leq \frac{4(1+\beta)}{n}.$$

Combining everything, and applying Theorem 2.2.1 with $B = 0$ and $C = 2(1+\beta)/n$, we get

$$\mathbb{P}(|f(\mathbf{X})| \geq t) \leq 2\exp\left(-\frac{nt^2}{4(1+\beta)}\right)$$

for all $t \geq 0$. From (2.27) it follows that

$$\mathbb{P}(|g(\mathbf{X})| \geq t) \leq \mathbb{P}(|f(\mathbf{X})| \geq t - 2\beta/n) \leq 2\exp\left(-\frac{nt^2}{8(1+\beta)}\right)$$

for all $t \geq 8\beta/n$. This completes the proof of the tail bound. The bound on the mean absolute value is an easy consequence of the tail bound. □

Proof of Lemma 2.3.8. The proof is in two steps. In the first step we will get an error bound of order $n^{-1/2}\sqrt{\log n}$. In the second step we will improve it to $n^{-1/2}$. Define

$$\Delta = \max_{1\leq i<j\leq n}\left|L_{ij} - \frac{1}{n}\sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk})\right|.$$

By Lemma 2.3.7 and union bound we have

$$\mathbb{P}\left(\Delta \geq t\right) \leq n^2\exp\left(-\frac{nt^2}{8(1+\beta)}\right)$$

for all $t \geq 8\beta/n$. Intuitively the above equation says that Δ is of the order of $\sqrt{\log n/n}$, in fact we have $\mathbb{E}(\Delta^2) = O(\log n/n)$. Clearly φ is an increasing function. Hence we have

$$\varphi(L_{\min})^2 - \Delta \leq L_{\min} \leq L_{\max} \leq \varphi(L_{\max})^2 + \Delta$$

where $L_{\max} = \max_{1\leq i<j\leq n} L_{ij}$ and $L_{\min} = \min_{1\leq i<j\leq n} L_{ij}$.

Now assume that there exists a unique solution u^* of the equation $\varphi(u)^2 = u$ with $2\varphi(u^*)\varphi'(u^*) < 1$. For ease of notations, define the function $\psi(u) = \varphi(u)^2 - u$. We have $\psi(0) > 0 > \psi(1)$, u^* is the unique solution to $\psi(u) = 0$ and $\psi'(u^*) < 0$. It is easy to see that $\psi'(u) = 0$ has at most three solution ($\psi'(u) = 2\beta\varphi(u)^2(1-\varphi(u)) - 1$ is a third degree polynomial in $\varphi(u)$ and φ is a strictly increasing function).

Hence there exist positive real numbers ε, δ such that $|\psi(u)| > \varepsilon$ if $|u - u^*| > \delta$. Note that $\psi(u) > 0$ if $u < u^*$ and $\psi(u) < 0$ is $u > u^*$. Decreasing ε, δ without loss of generality we can assume that

$$\inf_{0<|u-u^*|\leq\delta}\left[\frac{u-u^*}{-\psi(u)}\right] = c > 0. \tag{2.28}$$

This is possible because $\psi'(u^*) < 0$. Note that $\psi(L_{\max}) \geq -\Delta$ and $\psi(L_{\min}) \leq \Delta$. Thus we have

$$u^* - \delta \leq L_{\min} \leq L_{\max} \leq u^* + \delta$$

when $\Delta < \varepsilon$. Using (2.28), $u^* \leq L_{\max} \leq u^* + \delta$ implies that $|L_{\max} - u^*| \leq c\Delta$ and $u^* - \delta \leq L_{\min} \leq u^*$ implies that $|L_{\min} - u^*| \leq c\Delta$. Thus, when $\Delta < \varepsilon$, we have $|L_{\max} - u^*| \leq c\Delta$ and $|L_{\min} - u^*| \leq c\Delta$ and in particular, $|L_{ij} - u^*| \leq c\Delta$ for all $i < j$. So we can bound the L^2 distance of L_{ij} from u^* by

$$\mathbb{E}(L_{ij} - u^*)^2 \leq c^2\, \mathbb{E}(\Delta^2) + \mathbb{P}(\Delta \geq \varepsilon) \leq K(\beta, h)\frac{\log n}{n}$$

for all $i < j$.

Now let us move to the second step. Recall from (2.9) that

$$\mathbb{E}\left| L_{ij} - \frac{1}{n}\sum_{k \notin \{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \right| \leq \frac{C(1+\beta)^{1/2}}{n^{1/2}} \tag{2.29}$$

for all $i < j$. Let $D_{ij} = L_{ij} - u^*$. Using Taylor expansion around u^* upto degree one we have

$$\begin{aligned}\varphi(L_{ik})\varphi(L_{jk}) - \varphi(u^*)^2 &= \varphi(u^*)(\varphi(L_{ik}) - \varphi(u^*)) + \varphi(u^*)(\varphi(L_{jk}) - \varphi(u^*)) \\ &\quad + (\varphi(L_{ik}) - \varphi(u^*))(\varphi(L_{jk}) - \varphi(u^*)) \\ &= \varphi(u^*)\varphi'(u^*)(D_{ik} + D_{jk}) + R_{ijk}\end{aligned}$$

where $\mathbb{E}(|R_{ijk}|) \leq C\, \mathbb{E}(D_{ij}^2) \leq Cn^{-1}\log n$ for some constant C depending only on β, h. Thus

$$\begin{aligned}\mathbb{E}\Bigg| L_{ij} - \frac{1}{n}\sum_{k \notin \{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) - D_{ij} + \frac{\varphi(u^*)\varphi'(u^*)}{n}\sum_{k \notin \{i,j\}} (D_{ik} + D_{jk}) \Bigg| \\ \leq \frac{2u^*}{n} + \frac{1}{n}\sum_{k \notin \{i,j\}} \mathbb{E}\,|R_{ijk}| \leq \frac{C\log n}{n}.\end{aligned} \tag{2.30}$$

Here we used the fact that $u^* = \varphi(u^*)^2$. Combining (2.29) and (2.30) we have

$$\mathbb{E}\left| D_{ij} - \frac{\varphi(u^*)\varphi'(u^*)}{n}\sum_{k \notin \{i,j\}} (D_{ik} + D_{jk}) \right| \leq \frac{C}{\sqrt{n}}$$

for all $i < j$. By symmetry, $\mathbb{E}\,|D_{ij}|$ is the same for all i, j. Thus finally we have

$$\mathbb{E}\,|L_{ij} - u^*| = \mathbb{E}\,|D_{ij}| \leq \frac{1}{1 - 2\varphi(u^*)\varphi'(u^*)} \cdot \frac{C}{\sqrt{n}} = \frac{K(\beta, h)}{\sqrt{n}}$$

where $K(\beta, h)$ is a constant depending on β, h.

When $\psi(u) = 0$ has a unique solution at $u = u^*$ with $2\psi(u^*)\psi'(u^*) = 1$, which happens at the critical point $\beta = (3/2)^3, h = \log 2 - 3/2$, instead of equation (2.28) we have

$$\inf_{0<|u-u^*|\leq\delta} \left[\frac{(u-u^*)^3}{-\psi(u)}\right] = c > 0$$

since $\psi(u^*) = \psi'(u^*) = \psi''(u^*) = 0$ and $\psi'''(u^*) < 0$. Then using a similar idea as above one can easily show that

$$\mathbb{E}\,|L_{ij} - u^*| \leq K(\beta, h) n^{-1/6}$$

for some constant K depending on β, h. This completes the proof of the Lemma. □

Remark. The proof becomes lot easier if we have

$$c := \varphi(1) \cdot \sup_{0\leq x\leq 1} \frac{|\varphi(x) - \varphi(u^*)|}{|x - u^*|} < \frac{1}{2}. \tag{2.31}$$

This is because, by the triangle inequality we have

$$\begin{aligned} \sum_{i<j} |L_{ij} - u^*| \leq \sum_{i<j} \left| L_{ij} - \frac{1}{n} \sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \right| \\ + \sum_{i<j} \left(\frac{1}{n} \sum_{k\notin\{i,j\}} \left|\varphi(L_{ik})\varphi(L_{jk}) - u^*\right| + \frac{2u^*}{n} \right). \end{aligned} \tag{2.32}$$

Now recall that condition (2.31) says that $\varphi(1)|\varphi(x) - \varphi(u^*)| \leq c|x - u^*|$ for all $x \in [0, 1]$. Moreover $L_{ij} \in [0, 1]$ for all i, j, and $u^* = \varphi(u^*)^2$. Thus,

$$\left|\varphi(L_{ik})\varphi(L_{jk}) - u^*\right| \leq c|L_{ik} - u^*| + c|L_{jk} - u^*|.$$

Combining everything we get

$$\sum_{i<j} |L_{ij} - u^*| \leq \frac{\sum_{i<j} \left| L_{ij} - \frac{1}{n} \sum_{k\notin\{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \right| + nu^*}{1 - 2c}.$$

Taking expectation on both sides, and applying Lemma 2.3.7, we get

$$\sum_{i<j} \mathbb{E}\,|L_{ij} - u^*| \leq \frac{C(1+\beta)n^{3/2}}{1 - 2c}.$$

And this gives the required result. In fact using basic calculus results one can easily check that condition (2.31) is satisfied when $h \geq 0$ or $\beta \leq 2$.

Now we will prove that in the exponential random graph model, the number of edges and number of triangles also satisfy certain 'mean-field' relations.

Lemma 2.4.2. *Recall that $E(\mathbf{x})$ and $T(\mathbf{x})$ denote the number of edges and number of triangles in the graph defined by the edge configuration $\mathbf{x} \in \Omega$. If $\mathbf{X}$ is drawn from the Gibbs' measure in Theorem 2.3.5, we have the bound*

$$\mathbb{E}\left|E(\mathbf{X}) - \sum_{i<j} \varphi(L_{ij})\right| \le C(1+\beta)^{1/2} n$$

$$\mathbb{E}\left|\frac{T(\mathbf{X})}{n} - \frac{1}{3}\sum_{i<j} L_{ij}\varphi(L_{ij})\right| \le C(1+\beta)^{1/2} n$$

where and C is a universal constant.

Proof. It is not difficult to see that

$$\mathbb{E}(X_{ij} \mid (X_{kl})_{(k,l)\neq(i,j)}) = \varphi(L_{ij}).$$

Let us create $\mathbf{X}'$ by choosing $1 \le i < j \le n$ uniformly at random and replacing X_{ij} with X'_{ij} drawn from the conditional distribution of X_{ij} given $(X_{kl})_{(k,l)\neq(i,j)}$. Let $F(\mathbf{X}, \mathbf{X}') = \binom{n}{2}(X_{ij} - X'_{ij})$. Then

$$f(\mathbf{X}) = \mathbb{E}(F(\mathbf{X}, \mathbf{X}')|\mathbf{X}) = \sum_{k<l}(X_{kl} - \varphi(L_{kl})) = E(\mathbf{X}) - \sum_{k<l}\varphi(L_{kl}).$$

Now $|F(\mathbf{X}, \mathbf{X}')| \le \binom{n}{2}$ and $|f(\mathbf{X}) - f(\mathbf{X}')| \le 1 + \beta$. Here we used the fact that $|\varphi'(x)| \le \beta/4$. Combining the above result and Theorem 2.2.1 with $B = 0, C = \frac{1}{2}(1+\beta)\binom{n}{2}$, we get the required bound.

Similarly, if we define $F(\mathbf{X}, \mathbf{X}') = \binom{n}{2}(X_{ij}L_{ij} - X'_{ij}L_{ij})$. Then

$$\begin{aligned} f(\mathbf{X}) = \mathbb{E}(F(\mathbf{X}, \mathbf{X}')|\mathbf{X}) &= \sum_{k<l}(X_{kl}L_{kl} - \varphi(L_{kl})L_{kl}) \\ &= \frac{3}{n}T(\mathbf{X}) - \sum_{k<l}\varphi(L_{kl})L_{kl}. \end{aligned}$$

Again, $|F(\mathbf{X}, \mathbf{X}')| \le \binom{n}{2}$ and $|f(\mathbf{X}) - f(\mathbf{X}')| \le C(1+\beta)$. The bound follows easily as before. □

The following result is an easy corollary of Lemma 2.3.8 and Lemma 2.4.2.

Corollary 2.4.3. *Suppose the conditions of Theorem 2.3.5 are satisfied. Then we have*

$$\mathbb{E}\left|E(\mathbf{X}) - \frac{n^2\varphi(u^*)}{2}\right| \le Cn^{3/2} \text{ and } \mathbb{E}\left|\frac{T(\mathbf{X})}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le Cn^{3/2}$$

where C is a constant depending only on β, h.

Lemma 2.4.4. *Suppose the conditions of Theorem 2.3.5 are satisfied. Let T_n be the number of triangles in the Erdős-Rényi graph $G(n, \varphi(0))$. Then there is a constant $K(\beta, h)$ depending only on β and h such that for all n*

$$\left|\frac{\log \mathbb{P}(|T_n - \binom{n}{3}\varphi(u^*)^3| \le K(\beta,h)n^{5/2})}{n^2} - \frac{-I(\varphi(u^*), \varphi(0))}{2}\right| \le \frac{K(\beta,h)}{\sqrt{n}}.$$

Proof. Let X be drawn from the Gibbs' measure in Theorem 2.3.5 with parameters β, h. From corollary 2.4.3 we see that there exists a constant $K(\beta, h)$ such that (for all n)

$$\mathbb{P}\left(\left|E(X) - \frac{n^2\varphi(u^*)}{2}\right| \le K(\beta,h)n^{3/2}\right) \ge \frac{3}{4}$$

and

$$\mathbb{P}\left(\left|\frac{T(X)}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le K(\beta,h)n^{3/2}\right) \ge \frac{3}{4}.$$

Now let

$$A = \left\{x \in \{0,1\}^n : \left|\frac{T(x)}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le K(\beta,h)n^{3/2}\right\}$$

and

$$B = A \cap \left\{x \in \{0,1\}^n : \left|E(x) - \frac{n^2\varphi(u^*)}{2}\right| \le K(\beta,h)n^{3/2}\right\}.$$

Now suppose $Y = (Y_{ij})_{1 \le i < j \le n}$ is a collection of i.i.d. random variables satisfying $\mathbb{P}(Y_{ij} = 1) = 1 - \mathbb{P}(Y_{ij} = 0) = \varphi(0)$ and $Z = (Z_{ij})_{1 \le i < j \le n}$ is another collection of i.i.d. random variables with $\mathbb{P}(Z_{ij} = 1) = 1 - \mathbb{P}(Z_{ij} = 0) = \varphi(u^*)$. Without loss of generality we can assume that $K(\beta, h)$ was chosen large enough to ensure that (again, for all n) $\mathbb{P}(Z \in A) \ge 1/2$ and $\mathbb{P}(Z \in B) \ge 1/2$. Now, it follows directly from the definition of A and Lemma 2.4.1 that

$$\begin{aligned}
&\left|\log \sum_{x \in A} e^{hE(x)} - \log \sum_{x \in A} e^{\frac{\beta T(x)}{n} + hE(x)} + \frac{\beta n^2 \varphi(u^*)^3}{6}\right| \\
&= \left|\log \sum_{x \in A} e^{hE(x) + \frac{\beta n^2\varphi(u^*)^3}{6}} - \log \sum_{x \in A} e^{\frac{\beta T(x)}{n} + hE(x)}\right| \qquad (2.33) \\
&\le \beta \max_{x \in A}\left|\frac{T(x)}{n} - \frac{n^2\varphi(u^*)^3}{6}\right| \le \beta K(\beta,h)n^{3/2}.
\end{aligned}$$

Next, observe that

$$\begin{aligned}
&\left|\log \sum_{x \in A} e^{\frac{\beta T(x)}{n} + hE(x)} - \log \sum_{x \in \Omega} e^{\frac{\beta T(x)}{n} + hE(x)}\right| \qquad (2.34) \\
&= |\log \mathbb{P}(X \in A)| \le |\log(3/4)|.
\end{aligned}$$

Similarly we have

$$\begin{aligned}\left|\log \sum_{x\in B} e^{\frac{\beta T(x)}{n}+hE(x)} - \log \sum_{x\in\Omega} e^{\frac{\beta T(x)}{n}+hE(x)}\right| \\ = |\log \mathbb{P}(X\in B)| \le |\log(1/2)|\end{aligned} \tag{2.35}$$

where we used the fact that $\mathbb{P}(X\in A\cap C)\ge \mathbb{P}(X\in A)+\mathbb{P}(X\in C)-1$. Combining the last two inequalities, we get

$$\left|\log \sum_{x\in A} e^{\frac{\beta T(x)}{n}+hE(x)} - \log \sum_{x\in B} e^{\frac{\beta T(x)}{n}+hE(x)}\right| \le \log(8/3). \tag{2.36}$$

Next, note that by the definition of B and Lemma 2.4.1, we have that for any h',

$$\begin{aligned}&\left|\log \sum_{x\in B} e^{\frac{\beta T(x)}{n}+hE(x)} - \frac{n^2(h-h')\varphi(u^*)}{2} - \frac{\beta n^2 \varphi(u^*)^3}{6} - \log\sum_{x\in B} e^{h'E(x)}\right| \\ &\le \sup_{x\in B}\left|\frac{\beta T(x)}{n} + hE(x) - \frac{n^2(h-h')\varphi(u^*)}{2} - \frac{\beta n^2\varphi(u^*)^3}{6} - h'E(x)\right| \\ &\le (\beta+|h-h'|)K(\beta,h)n^{3/2}.\end{aligned} \tag{2.37}$$

Now choose $h' = \log\frac{\varphi(u^*)}{1-\varphi(u^*)}$. Then

$$\left|\log\sum_{x\in B} e^{h'E(x)} - \log\sum_{x\in\Omega} e^{h'E(x)}\right| = |\log \mathbb{P}(Z\in B)| \le \log 2. \tag{2.38}$$

Adding up (2.33), (2.36), (2.37), and (2.38), and using the triangle inequality, we get

$$\left|\log\sum_{x\in A} e^{hE(x)} - \frac{n^2(h-h')\varphi(u^*)}{2} - \log\sum_{x\in\Omega} e^{h'E(x)}\right| \le K'(\beta,h)n^{3/2} \tag{2.39}$$

where $K'(\beta,h)$ is a constant depending only on β, h. For any $s\in\mathbb{R}$, a trivial verification shows that

$$\log\sum_{x\in\Omega} e^{sE(x)} = \binom{n}{2}\log(1+e^s).$$

Again, note that $\log\mathbb{P}(Y\in A) = \log\sum_{x\in A} e^{hE(x)} - \log\sum_{x\in\Omega} e^{hE(x)}$. Therefore it follows from inequality (2.39) that

$$\left|\frac{\log\mathbb{P}(Y\in A)}{n^2} - \frac{(h-h')\varphi(u^*)+\log(1+e^{h'})-\log(1+e^h)}{2}\right| \le \frac{K'(\beta,h)}{\sqrt{n}}.$$

Now $h = \log\frac{\varphi(0)}{1-\varphi(0)}$ and $h' = \log\frac{\varphi(u^*)}{1-\varphi(u^*)}$. Also, $\log(1+e^h) = -\log(1-\varphi(0))$ and $\log(1+e^{h'}) = -\log(1-\varphi(u^*))$. Substituting these in the above expression, we get

$$\left|\frac{\log\mathbb{P}(Y\in A)}{n^2} - \frac{-I(\varphi(u^*),\varphi(0))}{2}\right| \le \frac{K'(\beta,h)}{\sqrt{n}}.$$

This completes the proof of the Lemma. ∎

We are now ready to finish the proof of Theorem 2.3.5.

Proof of Theorem 2.3.5. Note that by adding the terms in (2.35), (2.37), and (2.38) from the proof of Lemma 2.4.4, and applying the triangle inequality, we get

$$\left| \frac{\log Z_n(\beta, h)}{n^2} - \frac{(h-h')\varphi(u)}{2} - \frac{\beta\varphi(u)^3}{6} - \frac{1}{2}\log(1+e^{h'}) \right| \le \frac{K(\beta, h)}{\sqrt{n}}.$$

This can be rewritten as

$$\left| \frac{\log Z_n(\beta, h)}{n^2} + \frac{I(\varphi(u), \varphi(0)) + \log(1-\varphi(0))}{2} - \frac{\beta\varphi(u)^3}{6} \right| \le \frac{K(\beta, h)}{\sqrt{n}}.$$

This completes the proof of Theorem 2.3.5. □

Note that the proof of Theorem 2.3.5 contains a proof for the lower bound in the general case. We provide the proof below for completeness.

Proof of Lemma 2.3.6. Fix any $r \in (0, 1)$. Define the set B_r as

$$B_r = \left\{ x \in \{0,1\}^n : \left| \frac{T(x)}{n} - \frac{n^2 r^3}{6} \right| \le K(r) n^{3/2}, \left| E(x) - \frac{n^2 r}{2} \right| \le K(r) n^{3/2} \right\}$$

where $K(r)$ is chosen in such a way that $\mathbb{P}(Z \in B_r) \ge 1/2$ where $Z = ((Z_{ij}))_{i<j}$ and Z_{ij}'s are i.i.d. Bernoulli(r). From the proof of Lemma 2.4.4 it is easy to see that

$$\left| \log \sum_{x \in B_r} e^{\frac{\beta T(x)}{n} + hE(x)} - \frac{n^2}{2}\left((h-h')r + \frac{\beta r^3}{3} + \log(1+e^{h'}) \right) \right| \le K' n^{3/2}$$

where $h' = \log \frac{r}{1-r}$ and K' is a constant depending on β, h, r. Simplifying we have

$$\begin{aligned} \frac{2}{n^2}\log Z_n(\beta, h) &\ge \frac{2}{n^2}\log \sum_{x\in B_r} e^{\frac{\beta T(x)}{n} + hE(x)} \\ &\ge \frac{\beta r^3}{3} + \log(1-p) - I(r,p) - \frac{K'}{\sqrt{n}} \end{aligned} \tag{2.40}$$

for all r where $p = e^h/(1+e^h)$. Now taking limit as $n \to \infty$ and maximizing over r we have the first inequality (2.7). Given β, h, define the function

$$f(r) = \frac{\beta r^3}{3} + \log(1-p) - I(r,p)$$

where $p = e^h/(1+e^h)$. One can easily check that $f'(r) \gtreqless 0$ iff $\varphi(u)^2 - u \gtreqless 0$ for $u = r^2$. From this fact the second equality follows. □

Lemma 2.4.5. *Let T_n be the number of triangles in the Erdős-Rényi graph $G(n, \varphi(0))$. Then there is a constant $K(\beta, h)$ depending only on β and h such that for all n*

$$\frac{\log \mathbb{P}(T_n \geq \binom{n}{3}\varphi(u^*)^3)}{n^2} \leq \frac{-I(\varphi(u^*), \varphi(0))}{2} + \frac{K(\beta, h)}{\sqrt{n}}.$$

Proof. By Markov's inequality, we have

$$\frac{\log \mathbb{P}(T_n \geq \binom{n}{3}\varphi(u^*)^3)}{n^2} \leq -\frac{\beta}{n^3}\binom{n}{3}\varphi(u^*)^3 + \frac{\mathbb{E}(e^{\beta T_n/n})}{n^2}.$$

From the last part of Theorem 2.3.5, it is easy to obtain an optimal upper bound of the second term on the right hand side, which finishes the proof of the Lemma. ☐

Proof of Theorem 2.3.4. Given p and r, if for all r' belonging to a small neighborhood of r there exist β and h satisfying the conditions of Theorem 2.3.5 such that $\varphi(0) = p$ and $\varphi(u^*) = r'$, then a combination of Lemma 2.4.4 and Lemma 2.4.5 implies the conclusion of Theorem 2.3.4. If $p \geq p_0 = 2/(2 + e^{3/2})$, we can just choose $h \geq h_0 = -\log 2 - 3/2$ such that $p = e^h/(1 + e^h)$ and conclude, from Theorem 2.3.5, Lemma 2.4.4 and Lemma 2.3.9, that the large deviations limit holds for any $\beta \geq 0$. Varying β between 0 and ∞, it is possible to get for any $r \geq p$ a β such that $\varphi(u^*) = r$.

For $p \leq p_0$, we again choose h such that $\varphi(0) = p$. Note that $h \leq h_0$. The large deviations limit should hold for any $r \geq p$ for which there exists $\beta > 0$ such that $r = \varphi(u^*) = \sqrt{u^*}$ and $(h, \beta) \in S$. It is not difficult to verify that given h, u^* is a continuously increasing function of β in the regime for which $(h, \beta) \in S$. Recall the settings of Lemma 2.3.9. Thus, the values of r that is allowed is in the set $(p, p_*) \cup (p^*, 1]$, where p^*, p_* are the unique non-touching solutions to the equations

$$\sqrt{p^*} = \frac{e^{\beta_*(h)p^*+h}}{1 + e^{\beta_*(h)p^*+h}}, \quad \sqrt{p_*} = \frac{e^{\beta^*(h)p_*+h}}{1 + e^{\beta^*(h)p_*+h}}.$$

This completes the proof of Theorem 2.3.4. ☐

Finally, let us round up by proving Lemma 2.3.9.

Proof of Lemma 2.3.9. Fix $h \in \mathbb{R}$. Define the function

$$\psi(x; h, \beta) := \varphi(x; h, \beta)^2 - x$$

where

$$\varphi(x; h, \beta) = \frac{e^{\beta x+h}}{1 + e^{\beta x+h}} \text{ for } x \in [0, 1].$$

For simplicity, we will omit β, h in $\varphi(x; \beta, h)$ and $\psi(x; \beta, h)$ when there is no chance of confusion. Note that $\psi(0) > 0 > \psi(1)$. Hence the equation $\varphi(x; \beta, h) = 0$ has at least one solution. Also we have $\psi'(x) = 2\beta\varphi(x)^2(1-\varphi(x)) - 1$ and φ is strictly increasing. Hence the equation $\psi'(x) = 0$ has at most three solutions. So either the function ψ is strictly decreasing or there exist two numbers $0 < a < b < 1$ such that ψ is strictly decreasing in $[0, a] \cup [b, 1]$ and strictly increasing in $[a, b]$. From the above observations it is easy to see that the equation $\psi(x) = 0$ has at most three solutions for any β, h. If $\psi(x) = 0$ has exactly two solutions then $\psi' = 0$ at one of the solution.

Let $u_* = u_*(h, \beta)$ and $u^* = u^*(h, \beta)$ be the smallest and largest solutions of $\psi(x; h, \beta) = 0$ respectively. If $u_* = u^*$ we have a unique solution of $\psi(x) = 0$. From the fact that $\frac{\partial}{\partial\beta}\psi(x; h, \beta) > 0$ for all $x \in [0, 1], \beta \geq 0, h \in \mathbb{R}$ we can deduce that given h, $u_*(h, \beta)$ and $u^*(h, \beta)$ are increasing functions of β. Note that u_* is left continuous and u^* is right continuous in β given h. Also note that given $h \in \mathbb{R}$, $u^* = u_*$ if $\beta > 0$ is very small or very large. So we can define $\beta_*(h)$ and $\beta^*(h)$ such that for $\beta < \beta_*(h)$ and for $\beta > \beta^*(h)$ we have $u_*(h, \beta) = u^*(h, \beta)$. β_* is the largest and β^* is the smallest such number.

Therefore, we can deduce that at $\beta = \beta_*(h), \beta^*(h)$ the equation $\psi(x; h, \beta) = 0$ has exactly two solutions. Thus we have two real numbers $x_*, x^* \in [0, 1]$ such that

$$\varphi(x)^2 = x \text{ and } 2\beta\varphi(x)^2(1 - \varphi(x)) = 1$$

for $(x, \beta) = (x_*, \beta_*)$ or (x^*, β^*). Thus we have $2\beta x(1 - \sqrt{x}) = 1$ and

$$h = \log\frac{\sqrt{x}}{1-\sqrt{x}} - \frac{1}{2(1-\sqrt{x})}$$

for $x = x_*, x^*$. Define $a_* = x_*^{-1/2} - 1$ and $a^* = (x^*)^{-1/2} - 1$. Note that $x = (1+a)^{-2}, \beta = (1+a)^3/2a^2$ for $(x, a, \beta) = (x_*, a_*, \beta_*)$ or (x^*, a^*, β^*) and we have

$$h = -\log a - \frac{1+a}{2a} \tag{2.41}$$

for $a = a_*, a^*$. Now the function $g(x) = -\log x - (1+x)/2x$ is strictly increasing for $x \in (0, 1/2]$ and strictly decreasing for $x \geq 1/2$. So equation (2.41) has no solution for $h \geq g(1/2) = \log 2 - 3/2 =: h_0$. For $h < h_0$ equation (2.41) has exactly two solutions and for $h = h_0$ equation (2.41) has one solution. One can easily check that $\beta_* \leq \beta^*$ implies that $a_* \leq a^*$. Also from the fact that (2.41) has at most two solutions, we have that for $\beta \in (\beta_*, \beta^*)$ the equation $\psi(u) = 0$ has exactly three solutions. □

Proof of Lemma 2.3.10. For simplicity we will prove the result only for the lower boundary part, that is, for $(h, \beta) = \gamma(t)$ with $t < 1/2$. The proof for the upper boundary is similar. Fix $t < 1/2$. Let us briefly recall the setup. The function $\psi(u) = \varphi(u)^2 - u$ has two roots at $0 < u^* < v^* < 1$ and $\psi'(u_*) < 0$ while $\psi'(v^*) = 0, \psi''(v^*) < 0$.

Define the function

$$f(r) = \frac{\beta r^3}{3} + \log(1-p) - I(r, p) \text{ for } r \in (0, 1).$$

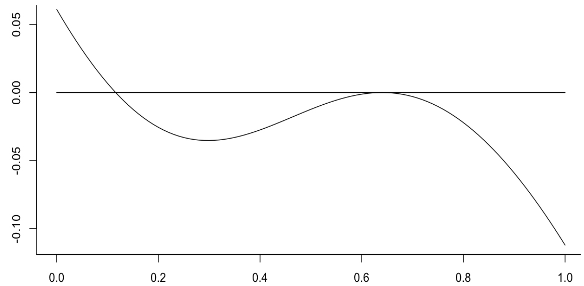

Figure 2.5: The function $\psi(\cdot)$ for $(h,\beta)=\gamma(1/4)$.

From the proof of Lemma 2.3.6 and the fact that $\psi'(u)<0$ for $u\in(u^*,v^*)$ it is easy to see that $f(\varphi(u^*))>f(\varphi(v^*))$ and

$$\frac{2}{n^2}\log Z_n(\beta,h)\geq f(\varphi(u^*))-\frac{K}{\sqrt{n}} \tag{2.42}$$

where K depends on β,h. Now, using the same idea used in the proof of Lemma 2.3.8, we have

$$\mathbb{P}(\Delta\geq t)\leq n^2\exp\left(-\frac{nt^2}{8(1+\beta)}\right)$$

for all $t\geq 8\beta/n$ and $\psi(L_{\max})\geq-\Delta,\psi(L_{\min})\leq\Delta$ where

$$\Delta=\max_{1\leq i<j\leq n}\left|L_{ij}-\frac{1}{n}\sum_{k\notin\{i,j\}}\varphi(L_{ik})\varphi(L_{jk})\right|.$$

Hence there exists $\varepsilon_0>0,c>0$ such that whenever $\Delta<\varepsilon_0$ we have $L_{\min}\geq u^*-c\Delta$ and either $L_{\max}\leq u^*+c\Delta$ or $|L_{\max}-v^*|\leq c\sqrt{\Delta}$. Define

$$U=\{L_{\max}<(u^*+v^*)/2\}. \tag{2.43}$$

Then again using the idea used in Lemma 2.3.8 one can easily show that

$$\mathbb{E}(\mathbb{1}_U\cdot|L_{ij}-u^*|)\leq\frac{K(\beta,h)}{n^{1/2}}\text{ for all }i<j.$$

We will show that $\mathbb{P}(U^c)\leq(\log n)^2/n$ and it will imply that

$$\mathbb{E}(|L_{ij}-u^*|)\leq\mathbb{E}(\mathbb{1}_U\cdot|L_{ij}-u^*|)+\mathbb{P}(U^c)\leq\frac{K(\beta,h)}{n^{1/2}}\text{ for all }i<j.$$

Then the rest of the assertions follow using the steps in the proof of Theorem 2.4.4.

Hence let us concentrate on the event U^c. It is enough to restrict to the event $U^c\cap\{|L_{\max}-v^*|\leq c\sqrt{\Delta}\}\cap\{L_{\min}\geq u^*-c\Delta\}$. Here the rough idea is that, a large fraction of L_{ij}'s has to be near v^* in order to make $L_{\max} \simeq v^*$. Suppose $L_{\max} = L_{i_0 j_0}$. Define the set

$$A = \{k : L_{i_0 k} < L_{\max} - \delta_1\}$$

where δ_1 will be chosen later such that $\delta_1 + c\sqrt{\Delta} < v^* - u^*$. Note that $\varphi(u)^2 \le \max\{u, u^*\}$ for all u and by assumption $|L_{\max} - v^*| \le c\sqrt{\Delta}$. Thus $\varphi(L_{ij}) \le \sqrt{L_{\max}}$ for all i, j and $\varphi(L_{i_0 k}) \le \sqrt{L_{\max} - \delta_1} \le \sqrt{L_{\max}}(1 - \delta_1/2)$ for $k \in A$. Thus we have

$$L_{\max} = L_{i_0 j_0} \le \Delta + \frac{1}{n} \sum_{k \ne i_0, j_0} \varphi(L_{i_0 k})\varphi(L_{j_0 k}) \le \Delta + L_{\max} - \frac{|A|\delta_1}{2n}$$

which clearly implies that $\frac{|A|}{n} \le \frac{2\Delta}{\delta_1}$. Similarly define the set $A_j = \{k : L_{jk} < L_{\max} - \delta_2\}$ where δ_2 will be chosen later such that $\delta_2 + c\sqrt{\Delta} < v^* - u^*$. Using same idea as before, for $j \notin A$ we have

$$L_{\max} - \delta_1 \le L_{i_0 j} \le \Delta + L_{\max} - \frac{|A_j|\delta_2}{2n} \text{ or } \frac{|A_j|}{n} \le \frac{2(\Delta + \delta_1)}{\delta_2} := M(\text{say}).$$

Choose $\delta_2 = \Delta^{1/5}, \delta_1 = \Delta^{3/5}$. Then we have

$$\begin{aligned} \sum_{i<j} |L_{ij} - L_{\max}|^2 &\le \frac{n|A| + nM + n^2\delta_2^2}{2} \\ &\le \frac{n^2\Delta}{\delta_1} + \frac{n^2(\Delta + \delta_1)}{\delta_2} + \frac{n^2\delta_2^2}{2} \le 4n^2\Delta^{2/5}. \end{aligned}$$

Thus, by symmetry and Hölders' inequality, we have

$$\begin{aligned} \mathbb{E}(\mathbb{1}_{U^c} \cdot |L_{ij} - v^*|^2) \le K\,\mathbb{E}(\mathbb{1}_{U^c} \cdot \Delta^{2/5}) &\le K\,\mathbb{P}(U^c)^{9/10} \cdot \mathbb{E}(\Delta^4)^{1/10} \\ &\le \frac{K(\log n)^{1/5}}{n^{1/5}}\,\mathbb{P}(U^c)^{9/10}. \end{aligned} \tag{2.44}$$

for some constant K. Now using lemma 2.4.2 and equation (2.44) we have,

$$\begin{aligned} \mathbb{E}\left[\left|E(\mathbf{X}) - \frac{n^2\varphi(v^*)}{2}\right| \,\middle|\, U^c\right] &\le \frac{Cn^{9/5}(\log n)^{1/5}}{\mathbb{P}(U^c)^{1/10}} \\ \text{and } \mathbb{E}\left[\left|\frac{T(\mathbf{X})}{n} - \frac{n^2\varphi(v^*)^3}{6}\right| \,\middle|\, U^c\right] &\le \frac{Cn^{9/5}(\log n)^{1/5}}{\mathbb{P}(U^c)^{1/10}}. \end{aligned} \tag{2.45}$$

If $\mathbb{P}(U^c) > (\log n)^2/n$, from inequality (2.45) we have

$$\begin{aligned} \mathbb{P}\left(\left|E(\mathbf{X}) - \frac{n^2\varphi(v^*)}{2}\right| \ge Kn^{19/10} \,\middle|\, U^c\right) &\le \frac{1}{4} \\ \text{and } \mathbb{P}\left(\left|\frac{T(\mathbf{X})}{n} - \frac{n^2\varphi(v^*)^3}{6}\right| \ge Kn^{19/10} \,\middle|\, U^c\right) &\le \frac{1}{4} \end{aligned}$$

for some large constant K depending on β, h. Now define the set

$$B = \left\{ x \in \{0,1\}^n : \left| \frac{T(x)}{n} - \frac{n^2 \varphi(v^*)^3}{6} \right| \leq K n^{19/10}, \left| E(x) - \frac{n^2 \varphi(v^*)}{2} \right| \leq K n^{19/10} \right\}.$$

Using the same idea used in the proof of lemma 2.4.4 one can again show that

$$\left| \frac{2}{n^2} \log(Z_n \, \mathbb{P}(U^c)) - f(\varphi(v^*)) \right| \leq \frac{K}{n^{1/10}}$$

for some constant K depending on β, h. The crucial fact is that $\mathbb{P}(\{L_{\max}(\mathbf{Z}) > (u^* + v^*)/2\} \cap \{\mathbf{Z} \in B\})$ is bounded away from zero when $\mathbf{Z} = ((Z_{ij}))_{i<j} \sim \mathrm{G}(n, \varphi(v^*))$. Thus we have

$$\left| \frac{2}{n^2} \log Z_n - f(\varphi(v^*)) \right| \leq \frac{K}{n^{1/10}}.$$

But this leads to a contradiction, since by equation (2.42) we have

$$\frac{2}{n^2} \log Z_n(\beta, h) \geq f(\varphi(u^*)) - \frac{K}{\sqrt{n}}$$

and $f(\varphi(u^*)) > f(\varphi(v^*))$. Thus we have $\mathbb{P}(U^c) \leq (\log n)^2/n$ and we are done.

□

2.4.2 Proof of the large deviation result for general subgraph count

We will prove Theorem 2.3.12 first. The proof follows the same line as the proof of Theorem 2.3.5.

Proof of Theorem 2.3.12. Recall the definition of L_{ij},

$$L_{ij} := \frac{N(\mathbf{X}^1_{(i,j)}) - N(\mathbf{X}^0_{(i,j)})}{(n-2)_{\boldsymbol{v}_F - 2}} \text{ for } i < j. \tag{2.46}$$

In fact we can write L_{ij} explicitly as a horrible sum

$$L_{ij} = \frac{1}{\alpha_F (n-2)_{\boldsymbol{v}_F - 2}} \sum_{\substack{t_1 < t_2 < \cdots < t_{\boldsymbol{v}_F - 2} \\ t_l \in [n] \setminus \{i,j\} \text{ for all } l}} \sum_{(a,b) \in E(F)} \sideset{}{'}\sum_{\pi} \prod_{\substack{(k,l) \in E(F) \\ (k,l) \neq (a,b)}} X_{\pi_k \pi_l}$$

where the sum $\sum'$ is over all one-one onto map π from $V(F) = [\boldsymbol{v}_F]$ to $\{a, b, t_1, \ldots, t_{\boldsymbol{v}_F - 2}\}$ where $\{\pi(a), \pi(b)\} = \{i, j\}$. Now we briefly state the main steps. First we have $\mathbb{E}(X_{ij} \mid \text{rest}) = \varphi(L_{ij})$. Moreover, using lemma 2.4.1 it is easy to see that

$$\left| \mathbb{E}\left(\prod_{j=1}^{k} X_{i_{2j-1} i_{2j}} \mid \text{rest}\right) - \prod_{j=1}^{k} \varphi(L_{i_{2j-1} i_{2j}}) \right| \leq C\beta/n$$

for every distinct pairs $(i_1, i_2), \ldots, (i_{2k-1}, i_{2k})$ where C is an universal constant.

Now, fix $1 \le i < j \le n$. Given a configuration $\mathbf{X}$, construct another one $\mathbf{X}'$ in the following way. Choose $\boldsymbol{v}_F - 2$ distinct points uniformly at random without replacement from $[n] \setminus \{i, j\}$. Replace the coordinates in $\mathbf{X}$ corresponding to the edges in the complete subgraph formed by the chosen points including i, j (except that we do not change X_{ij}) by values drawn from the conditional distribution given the rest of the edges. Call the new configuration $\mathbf{X}'$. Define the antisymmetric function $F(\mathbf{X}, \mathbf{X}') := (n-2)_{\boldsymbol{v}_F - 2}(L_{ij} - L'_{ij})$. and $f(\mathbf{X}) := \mathbb{E}(F(\mathbf{X}, \mathbf{X}') \mid \mathbf{X})$. Using the same idea as before and Theorem 2.2.1 we have

$$\mathbb{P}\left(|L_{ij} - g_{ij}| \ge t\right) \le \exp(-cnt^2/(1+\beta)) \tag{2.47}$$

where c is an absolute constant and g_{ij} is obtained from L_{ij} by replacing X_{kl} by $\varphi(L_{kl})$ for all $k < l$. Note that there is a slight difference with the calculation in the triangle case, since we have to consider collections of edges where some are modified and some are not. But their contribution will be of the order of n^{-1}. Also the conditions on φ arises in the following way, if all the L_{ij}'s are constant, say equal to u, then from the "mean-field equations" for L_{ij}'s we must have

$$\begin{aligned} u &\approx \frac{1}{\alpha_F (n-2)_{\boldsymbol{v}_F - 2}} \sum_{\substack{t_1 < t_2 < \cdots < t_{\boldsymbol{v}_F - 2} \\ t_l \in [n]\setminus\{i,j\} \text{ for all } l}} \sum_{(a,b)\in E(F)} \sum_{\pi}{}' \varphi(u)^{\boldsymbol{e}_F - 1} \\ &= \frac{2\boldsymbol{e}_F}{\alpha_F} \varphi(u)^{\boldsymbol{e}_F - 1}. \end{aligned}$$

The next step is to show that under the conditions on φ, we have $\mathbb{E}\,|L_{ij} - u^*| \le Kn^{-1/2}$ for all $i < j$ where $K = K(\beta, h)$ is a constant depending only on β, h. The crucial fact is that the behavior of the function $\varphi(u)^k - au$ where $a > 0$ is a positive constant and $k \ge 2$ is a fixed integer, is same as the behavior of the function $\varphi(u)^2 - u$.

Now it will follow (using the same proof used for lemma 2.4.2) that

$$\begin{aligned} \mathbb{E}\left| E(\mathbf{X}) - \frac{n^2 \varphi(u^*)}{2} \right| &\le Cn^{3/2} \\ \text{and } \ \mathbb{E}\left| N(\mathbf{X}) - \frac{(n)_{\boldsymbol{v}_F} \varphi(u^*)^{\boldsymbol{e}_F}}{\alpha_F} \right| &\le Cn^{\boldsymbol{v}_F - 1/2} \end{aligned}$$

where C is a constant depending only on β, h. The rest of the proof follows using the arguments used in the proof of Theorem 2.3.5. □

Proof of Theorem 2.3.11. Using the method of proof for the triangle case and the result from Theorem 2.3.12 the proof follows easily. □

Proof of Lemma 2.3.13. The proof is same as the proof of lemma 2.3.9 except for the constants. □

2.4.3 Proof for Ising model on $\mathbb{Z}^d$: Theorem 2.3.14

Suppose $\boldsymbol{\sigma}$ is drawn from the Gibbs distribution $\mu_{\beta,h}$. We construct $\boldsymbol{\sigma}'$ by taking one step in the heat-bath Glauber dynamics as follows: Choose a position I uniformly at random from Ω, and replace the I-th coordinate of $\boldsymbol{\sigma}$ by an element drawn from the conditional distribution of the σ_I given the rest. It is easy to see that $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ is an exchangeable pair. Let

$$F(\boldsymbol{\sigma}, \boldsymbol{\sigma}') := |\Omega|(m(\boldsymbol{\sigma}) - m(\boldsymbol{\sigma}')) = \sigma_I - \sigma'_I$$

be an antisymmetric function in $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$. Since the Hamiltonian is a simple explicit function, one can easily calculate the conditional distribution of the spin of the particle at position x given the spins of the rest. In fact we have $\mathbb{E}(\sigma_x|\{\sigma_y, y \neq x\}]) = \tanh(2\beta d m_x(\boldsymbol{\sigma}))$ where $m_x(\boldsymbol{\sigma}) := \frac{1}{2d}\sum_{y\in N_x} \sigma_y$ is the average spin of the neighbors of x for $x \in \Omega$. Now using Fourier-Walsh expansion we can write the function $\tanh(2\beta d m_x(\boldsymbol{\sigma}))$ as sums of products of spins in the following way. We have

$$\tanh(2d\beta m_x(\boldsymbol{\sigma})) = \sum_{k=0}^{2d} a_k(\beta) \sum_{|S|=k, S\subseteq N_x} \sigma_S \tag{2.48}$$

where

$$a_k(\beta) := \frac{1}{2^{2d}} \sum_{\boldsymbol{\sigma}\in\{-1,+1\}^{2d}} \tanh\left(\beta \sum_{i=1}^{2d} \sigma_i\right) \prod_{j=1}^{k} \sigma_j \tag{2.49}$$

for $k = 0, 1, \ldots, 2d$. It is easy to see that $a_k(\beta) = 0$ if k is even and $a_k(\beta)$ is a rational function of $\tanh(2\beta)$ if k is odd. Note that the dependence of a_k on d is not stated explicitly. Thus using equation (2.48) and the definitions in (2.20) we have

$$\begin{aligned}
f(\boldsymbol{\sigma}) = \mathbb{E}[F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')|\boldsymbol{\sigma}] &= \frac{1}{|\Omega|} \sum_{x\in\Omega} E[\sigma_x - \sigma'_x|\boldsymbol{\sigma}] \\
&= m(\boldsymbol{\sigma}) - \frac{1}{|\Omega|} \sum_{x\in\Omega} \tanh(2\beta d m_x(\boldsymbol{\sigma})) \\
&= (1 - 2d a_1(\beta)) m(\boldsymbol{\sigma}) - \sum_{k=1}^{d-1} \binom{2d}{2k+1} a_{2k+1}(\beta) r_{2k+1}(\boldsymbol{\sigma}).
\end{aligned}$$

Define $\theta_k(\beta) := \binom{2d}{2k+1} a_{2k+1}(\beta)$ for $k = 0, 1, \ldots, d-1$. Note that we can explicitly calculate the value of $\theta_0(\beta)$ as follows,

$$\theta_0(\beta) = \frac{1}{4^d} \sum_{\boldsymbol{\sigma}\in\{-1,+1\}^{2d}} \tanh\left(\beta \sum_{i=1}^{2d} \sigma_i\right) \sum_{i=1}^{2d} \sigma_i = \frac{2}{4^d} \sum_{k=1}^{d} 2k \binom{2d}{d+k} \tanh(2k\beta).$$

Now we have $|F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')| \leq 2$ and

$$|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| \leq \frac{2}{|\Omega|}\left(|1 - \theta_0(\beta)| + \sum_{k=1}^{d-1}(2k+1)\theta_k(\beta)\right) = \frac{2}{|\Omega|}b(\beta)$$

for all values of $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$. Hence the condition of Theorem 2.2.1 is satisfied with $B = 0, C = 2|\Omega|^{-1}b(\beta)$. So by part (ii) of Theorem 2.2.1 we have

$$\mathbb{P}\left(\sqrt{|\Omega|}\left|(1-\theta_0(\beta))m(\boldsymbol{\sigma}) - \sum_{k=1}^{d-1}\theta_k(\beta)r_{2k+1}(\boldsymbol{\sigma})\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{4b(\beta)}\right)$$

for all $t > 0$. Obviously $\theta_0(\cdot)$ is a strictly increasing function of β. Also we have $\theta_0(0) = 0$ and

$$\theta_0(\infty) := \lim_{\beta\to\infty}\theta_0(\beta) = \frac{1}{4^{d-1}}\sum_{k=1}^{d} k\binom{2d}{d+k}.$$

For $d = 1$ we have $\theta_0(\infty) = 1$ and for $d \geq 2$ we have

$$\begin{aligned}\theta_0(\infty) &\geq \frac{1}{4^{d-1}}\left[2\sum_{k=1}^{d}\binom{2d}{d+k} - \binom{2d}{d+1}\right]\\ &= \frac{1}{4^{d-1}}\left[2^{2d} - \binom{2d}{d} - \binom{2d}{d+1}\right] = 4 - \frac{8}{2^{2d+1}}\binom{2d+1}{d+1}\end{aligned}$$

and from the fact that $\sum_{k=d-1}^{d+2}\binom{2d+1}{k} \leq 2^{2d+1}$ we have

$$\frac{1}{2^{2d+1}}\binom{2d+1}{d+1} \leq \frac{d+2}{4(d+1)} \leq \frac{1}{3} \text{ for } d \geq 2.$$

Hence for $d \geq 2$ we have $\theta_0(\infty) > 1$ and there exists $\beta_1 \in (0, \infty)$, depending on d, such that $1 - \theta_0(\beta) > 0$ for $\beta < \beta_1$ and $1 - \theta_0(\beta) < 0$ for $\beta > \beta_1$. This completes the proof.

Proof of Proposition 2.3.16. The proof is almost same as the proof of proposition 2.3.14. Define $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$ as before. Define the antisymmetric function $F(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ as follows

$$\begin{aligned}F(\boldsymbol{\sigma}, \boldsymbol{\sigma}') &:= |\Omega|(1 + \tanh(h)\tanh(2\beta d m_I(\boldsymbol{\sigma})))(m(\boldsymbol{\sigma}) - m(\boldsymbol{\sigma}'))\\ &= (1 + \tanh(h)\tanh(2\beta d m_I(\boldsymbol{\sigma})))(\sigma_I - \sigma_I').\end{aligned}$$

Recall that $m_x(\boldsymbol{\sigma}) := \frac{1}{2d}\sum_{y\in N_x}\sigma_y$ is the average spin of the neighbors of x for $x \in \Omega$. Now under $\mu_{\beta,h}$ we have

$$\begin{aligned}\mathbb{E}(\sigma_x|\{\sigma_y, y \neq x\}) &= \tanh(2\beta d m_x(\boldsymbol{\sigma}) + h)\\ &= \frac{\tanh(h) + \tanh(2\beta d m_x(\boldsymbol{\sigma}))}{1 + \tanh(h)\tanh(2\beta d m_x(\boldsymbol{\sigma}))}.\end{aligned}$$

Thus we have

$$\begin{aligned} f(\boldsymbol{\sigma}) &= \mathbb{E}(F(\boldsymbol{\sigma}, \boldsymbol{\sigma'})|\boldsymbol{\sigma}) \\ &= \frac{1}{|\Omega|}\sum_{x\in\Omega}(1+\tanh(h)\tanh(2\beta d m_x(\boldsymbol{\sigma})))\,\mathbb{E}(\sigma_x - \sigma'_x|\boldsymbol{\sigma}) \\ &= m(\boldsymbol{\sigma}) - \tanh(h) + \frac{1}{|\Omega|}\sum_{x\in\Omega}(\tanh(h)\sigma_x - 1)\tanh(2\beta d m_x(\boldsymbol{\sigma})). \end{aligned}$$

After some simplifications and using the definitions of the functions r, s we have

$$\begin{aligned} f(\boldsymbol{\sigma}) = (1-\theta_0(\beta))m(\boldsymbol{\sigma}) &- \sum_{k=1}^{d-1}\theta_k(\beta) r_{2k+1}(\boldsymbol{\sigma}) \\ &- \tanh(h)\left(1 - \sum_{k=0}^{d-1}\theta_k(\beta)s_{2k+1}(\boldsymbol{\sigma})\right). \end{aligned}$$

Now for all values of $\boldsymbol{\sigma}, \boldsymbol{\sigma'}$ we have

$$|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma'})| \le \frac{2}{|\Omega|} b(\beta)(1+\tanh|h|)$$

and the proof onwards is exactly as in the proof of proposition 2.3.14. □

2.4.4 Proof of the main theorem: Theorem 2.2.2

Assume that $\psi(0) > 0$. We will handle the case $\psi(0) = 0$ later. Note that condition (2.1) implies that $x^\alpha/\psi(x)$ is a nondecreasing function for $x > 0$. Define the function

$$\varphi(x) := \frac{x^2}{\psi(x)} \text{ and } \gamma(x) := 2 - \frac{x\psi'(x)}{\psi(x)} \text{ for } x \neq 0$$

and $\varphi(0) = 0, \gamma(0) = 2$. Clearly we have $2 - \alpha \le \gamma(x) \le 2$ for all $x \in \mathbb{R}$. Now, $\limsup_{x\to 0}\varphi(x) \le \lim_{x\to 0+} x^{2-\alpha}/\psi(1) = 0 = \varphi(0)$ as $\alpha < 2$. Also $\varphi(x)$ is differentiable in $\mathbb{R}\setminus\{0\}$ with

$$\varphi'(x) = \frac{x\gamma(x)}{\psi(x)} > 0 \text{ for } x \neq 0. \tag{2.50}$$

Hence φ is absolutely continuous in $\mathbb{R}$ and is increasing for $x \ge 0$.

Define $Y = f(X)$. First we will prove that all moments of $\varphi(Y)$ are finite. Next we will estimate the moments which will in turn show that $\varphi(Y)^{1/2}$ has finite exponential moment in $\mathbb{R}$. Finally using Chebyshev's inequality we will prove the tail probability.

By monotonicity of ψ in $[0,\infty)$ and definition of α, we have

$$0 \le \frac{x\psi'(x)}{\psi(x)} \le \alpha \text{ for all } x \ge 0. \tag{2.51}$$

It also follows from (2.50) that $0 \leq (\log \varphi(x))' \leq 2/x$ for $x > 0$ and integrating we have $\varphi(x) \leq \varphi(1)x^2$ for all $x \geq 1$. Hence $\varphi(x) = \varphi(|x|) \leq \varphi(1)(1+x^2)$ for all $x \in \mathbb{R}$ and this, combined with our assumption that $\mathbb{E}(|f(X)|^k) < \infty$ for all $k \geq 1$, implies that $\mathbb{E}(\varphi(Y)^k) < \infty$ for all $k \geq 1$.

Define

$$\beta := \left\lceil \frac{5(2-\alpha) + \delta + 1/4}{(2-\alpha)^2} \right\rceil \geq 3.$$

Fix an integer $k \geq \beta$ and define

$$g(x) = \frac{x^{2k-1}}{\psi^k(x)} \text{ and } h(x) = \frac{x^{2k-2}}{\psi^k(x)} \text{ for } x \in \mathbb{R}.$$

Clearly $\mathbb{E}(|Yg(Y)|) < \infty$. Note that g, h are continuously differentiable in $\mathbb{R}$ as $k \geq 3$. Moreover, for $x \in \mathbb{R}$ we have, $|g'(x)| = h(x)\,|k\gamma(x) - 1| \leq (2k-1)h(x)$, $h'(x) = (k\gamma(x) - 2)\, x^{2k-3}/\psi^k(x)$ and

$$h''(x) = \left[(k\gamma(x) - 2)\,(k\gamma(x) - 3) + kx\gamma'(x)\right] \frac{x^{2k-4}}{\psi^k(x)}.$$

We also have

$$x\gamma'(x) = -\frac{x\psi'(x)}{\psi(x)}\left(1 - \frac{x\psi'(x)}{\psi(x)}\right) - \frac{x\psi''(x)}{\psi(x)} \geq -1/4 - \delta$$

for $x \in \mathbb{R}$. Now $k \geq \beta$ implies that

$$\begin{aligned}&(k\gamma(x) - 2)\,(k\gamma(x) - 3) + kx\gamma'(x)\\ &\qquad \geq (k(2-\alpha) - 2)\,(k(2-\alpha) - 3) - k(\delta + 1/4) \geq 0\end{aligned}$$

for all x. Thus $h''(x) \geq 0$ for all x and h is convex in $\mathbb{R}$.

Let $X', F(X, X')$ be as given in the hypothesis. Define $Y' = f(X')$. Recall that (X, X') is an exchangeable pair and so is (Y, Y'). Using the fact that $f(X) = \mathbb{E}(F(X, X')|X)$ almost surely, exchangeability of (X, X') and antisymmetry of F, we have

$$\begin{aligned}\mathbb{E}(Yg(Y)) = \mathbb{E}(f(X)g(Y)) &= \mathbb{E}(F(X, X')g(Y))\\ &= \frac{1}{2}\,\mathbb{E}(F(X, X')(g(Y) - g(Y'))). \end{aligned} \tag{2.52}$$

Now, for any $x < y$ we have

$$\left|\frac{g(x) - g(y)}{x - y}\right| = \left|\int_0^1 g'(tx + (1-t)y)\; dt\right| \leq (2k-1)\int_0^1 h(tx + (1-t)y)dt$$

and convexity of h implies that

$$\int_0^1 h(tx+(1-t)y)dt \le \int_0^1 (th(x)+(1-t)h(y))dt = (h(x)+h(y))/2.$$

Hence, from equation (2.52) we have

$$\begin{aligned}\mathbb{E}(Yg(Y)) &\le \frac{2k-1}{4}\,\mathbb{E}(|(Y-Y')F(X,X')|(h(Y)+h(Y'))) \\ &= (2k-1)\,\mathbb{E}(\Delta(X)h(Y)) \le (2k-1)\,\mathbb{E}(\psi(Y)h(Y)) \end{aligned} \tag{2.53}$$

where the equality follows by definition of $\Delta(X)$ and exchangeability of (Y,Y'). Thus for any $k \ge \beta$ we have, from (2.53),

$$\mathbb{E}(\varphi(Y)^k) \le (2k-1)\,\mathbb{E}(\varphi(Y)^{k-1}). \tag{2.54}$$

Using induction for $k \ge \beta$ we have

$$\mathbb{E}(\varphi(Y)^k) \le \frac{(2k)!2^\beta \beta!}{2^k k!(2\beta)!}\,\mathbb{E}(\varphi(Y)^\beta) \text{ for } k \ge \beta.$$

Also Hölder's inequality applied to (2.54) for $k=\beta$ implies that $\mathbb{E}(\varphi(Y)^\beta) \le (2\beta-1)^\beta$. Thus we have

$$\mathbb{E}(\varphi(Y)^k) \le \begin{cases} \frac{(2k)!2^\beta\beta!}{k!2^k(2\beta)!}\,\mathbb{E}(\varphi(Y)^\beta) & \text{if } k>\beta \\ (2\beta-1)^k & \text{if } 0 \le k \le \beta. \end{cases} \tag{2.55}$$

Note that we have $e^x \le e^x + e^{-x} = 2\sum_{k\ge 0} x^{2k}/(2k)!$ for all $x \in \mathbb{R}$. Combining everything we finally have

$$\begin{aligned}\mathbb{E}(\exp(\theta\varphi(Y)^{1/2})) &\le 2\sum_{k=0}^{\infty} \frac{\theta^{2k}}{(2k)!}\,\mathbb{E}(\varphi(Y)^k) \\ &\le \frac{2^{\beta+1}\beta!}{(2\beta)!}\,\mathbb{E}(\varphi(Y)^\beta)\sum_{k=\beta}^{\infty}\frac{\theta^{2k}}{2^k k!} + \sum_{k=0}^{\beta-1}\frac{2(2\beta-1)^k\theta^{2k}}{(2k)!} \\ &\le C_\beta \exp(\theta^2/2)\end{aligned}$$

for all $\theta \ge 0$ where the constant C_β is given by

$$C_\beta := \max\left\{\frac{2(2\beta-1)^k 2^k k!}{(2k)!}\,\middle|\, 0 \le k \le \beta\right\}.$$

Here we used the fact that $(2k)! \ge 2^{2k-1}k!^2/k$. Now recall that φ is an increasing function in $[0,\infty)$. Thus using Chebyshev's inequality for $\exp(\theta\varphi(x)^{1/2})$ with $\theta = \varphi(t)^{1/2}$ we have

$$\mathbb{P}(|f(X)| \ge t) \le C_\beta e^{-\theta\varphi(t)^{1/2}+\theta^2/2} = C_\beta e^{-\varphi(t)/2}.$$

Now suppose that $\psi(0)=0$. For $\varepsilon>0$ fixed, define $\psi_\varepsilon(x)=\psi(x)+\varepsilon$. Clearly we have $\Delta(X)\le\psi_\varepsilon(f(X))$ a.s. and ψ_ε satisfies all the other properties of ψ including

$$\begin{aligned} x\psi_\varepsilon'(x)/\psi_\varepsilon(x) &= x\psi'(x)/\psi(x)\cdot\psi(x)/(\psi(x)+\varepsilon)\le\alpha \\ \text{and } x\psi_\varepsilon''(x)/\psi_\varepsilon(x) &= x\psi''(x)/\psi(x)\cdot\psi(x)/(\psi(x)+\varepsilon)\le\delta \end{aligned}$$

for all $x>0$. Hence all the above results hold for ψ_ε and $\varphi_\varepsilon(x)=x^2/\psi_\varepsilon(x)$. Now $\varphi_\varepsilon\uparrow\varphi$ as $\varepsilon\downarrow 0$. Letting $\varepsilon\downarrow 0$ we have the result.

When ψ is once differentiable with $\alpha<2$, it is easy to see that the function h is nondecreasing (need not be convex) in $[0,\infty)$ for $k\ge\beta:=\lceil 2/(2-\alpha)\rceil$. In that case we have

$$\int_0^1 h(tx+(1-t)y)dy\le\max_{z\in[x,y]}h(z)\le h(x)+h(y)$$

for $x\le y$. Hence we have the recursion

$$\mathbb{E}(\varphi(Y)^k)\le 2(2k-1)\,\mathbb{E}(\varphi(Y)^{k-1}) \tag{2.56}$$

for $k\ge\beta$. Using the same proof as before it then follows that

$$\mathbb{P}(|f(X)|\ge t)\le Ce^{-\varphi(t)/4}$$

where C depends only on α. □

Chapter 3

First-passage percolation across thin cylinders

3.1 Introduction

Before stating the results, let us begin with a short review of the first-passage percolation model and some of the known results.

3.1.1 The model

More than forty years ago, Hammersley and Welsh [54] introduced first-passage percolation to model the spread of fluid through a randomly porous media. The standard first-passage percolation model on the d-dimensional square lattice $\mathbb{Z}^d$ is defined as follows. Consider the edge set E consisting of nearest neighbor edges, that is, $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{Z}^d \times \mathbb{Z}^d$ is an edge if and only if $\|\boldsymbol{x} - \boldsymbol{y}\| := \sum_{i=1}^{d} |x_i - y_i| = 1$. With each edge (also called a bond) $e \in E$ is associated an independent nonnegative random variable ω_e distributed according to a fixed distribution F. The random variable ω_e represents the amount of time it takes the fluid to pass through the edge e.

For a path $\mathcal{P}$ (which will always be finite and nearest neighbor) in $\mathbb{Z}^d$ define

$$\omega(\mathcal{P}) := \sum_{e \in \mathcal{P}} \omega_e$$

as the passage time for $\mathcal{P}$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^d$, let $a(\boldsymbol{x}, \boldsymbol{y})$, called the *first-passage time*, be the minimum passage time over all paths from $\boldsymbol{x}$ to $\boldsymbol{y}$. Intuitively $a(\boldsymbol{x}, \boldsymbol{y})$ is the first time the fluid will appear at $\boldsymbol{y}$ if a source of water is introduced at the vertex $\boldsymbol{x}$ at time 0. Formally

$$a(\boldsymbol{x}, \boldsymbol{y}) := \inf\{\omega(\mathcal{P}) \mid \mathcal{P} \text{ is a path connecting } \boldsymbol{x} \text{ to } \boldsymbol{y} \text{ in } \mathbb{Z}^d\}.$$

The principle object of study in first-passage percolation theory is the asymptotic behavior of $a(\mathbf{0}, n\boldsymbol{x})$ for fixed $\boldsymbol{x} \in \mathbb{Z}^d$. We refer the reader to Smythe and Wierman [101] and Kesten [67] for earlier surveys of the subject.

The first limit result proved by Hammersley and Welsh [54] was that the limit

$$\nu(\boldsymbol{x}) := \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[a(0, n\boldsymbol{x})] \tag{3.1}$$

exists and is finite when $\mathbb{E}[\omega] < \infty$ where ω is a generic random variable from the distribution F. Moreover results of Kesten [67] show that $\nu(\boldsymbol{x}) > 0$ if and only if $F(0) < p_c(d)$ where $p_c(d)$ is the critical probability for standard bernoulli bond percolation in $\mathbb{Z}^d$.

First-passage percolation is often regarded as a stochastic growth model by considering the growth of the random set

$$B_t := \{\boldsymbol{x} \in \mathbb{Z}^d \mid a(0, \boldsymbol{x}) \le t\}.$$

When $F(0) = 0$, $a(\cdot, \cdot)$ is a random metric on $\mathbb{Z}^d$ and B_t is the ball of radius t in this metric. Moreover, if $F(0) < p_c(d)$ and $\mathbb{E}[\omega^2] < \infty$ (or under weaker conditions in Cox and Durrett [35]), the growth of B_t is linear in t with a deterministic limit shape, that is, as $t \to \infty$, $B_t \approx tB_0 \cap \mathbb{Z}^d$ for a nonrandom compact set B_0. Precisely, the shape theorem says that (see Richardson [94], Cox and Durrett [35] and Kesten [67]), if $F(0) < p_c(d)$ and $\mathbb{E}[\min\{\omega_1^d, \omega_2^d, \ldots, \omega_{2d}^d\}] < \infty$ where $\omega_1, \ldots, \omega_{2d}$ are i.i.d. from F, there is a nonrandom compact set B_0 such that for all $\varepsilon > 0$

$$(1-\varepsilon)B_0 \subseteq t^{-1}\tilde{B}_t \subseteq (1+\varepsilon)B_0 \text{ eventually with probability one}$$

where $\tilde{B}_t = \{\boldsymbol{y} \in \mathbb{R}^d \mid \exists\, \boldsymbol{x} \in B_t \text{ s.t. } \|\boldsymbol{x} - \boldsymbol{y}\| \le 1\}$ is the "inflated" version of B_t.

3.1.2 Fluctuation exponents and and limit theorems

In the physics literature, there are mainly two fluctuation exponents χ and ξ that describe, respectively, the longitudinal and transversal fluctuations of the growing surface B_t. For example, it is expected under mild conditions that the first-passage time $a(\mathbf{0}, n\boldsymbol{x})$ has standard deviation of order n^χ, and the exponent χ is independent of the direction $\boldsymbol{x} \in \mathbb{Z}^d$. It is also expected that all the paths achieving the minimal time $a(\mathbf{0}, n\boldsymbol{x})$ deviate from the straight line path joining $\mathbf{0}$ to $n\boldsymbol{x}$ by distance at most of the order of n^ξ, that is all the minimal paths are expected to lie entirely inside the cylinder centered on the straight line joining $\mathbf{0}$ to $n\boldsymbol{x}$ whose width is of the order of n^ξ.

In general the exponents χ and ξ are expected to depend only on the dimension d not the distribution F. Moreover they are also conjectured to satisfy the scaling relation $\chi = 2\xi - 1$ for all d (see Krug and Spohn [71]). In fact, the predicted values for $d = 2$ (for models whose exponents are expected to be same in all directions) are

$\chi = 1/3$ and $\xi = 2/3$ (see Kardar, Parisi and Zhang [65]). For higher dimensions there are many conflicting predictions. However it is believed that above some finite critical dimension d_c, the exponents satisfy $\chi = 0$ and $\xi = 1/2$.

We briefly describe the rigorous results known about the exponents χ and ξ. The first nontrivial upper bound on the variance of $a(\mathbf{0}, n\boldsymbol{x})$ was $O(n)$ for all d due to Kesten [68]. The best known upper bound of $n/\log n$ is due to Benjamini, Kalai and Schramm [10]. In $d = 2$ the best known lower bound of $\log n$ is due to Pemantle and Peres [92] for exponential edge weights, Newman and Piza [87] for general edge weights satisfying $F(0) < p_c(2)$ or $F(\lambda) < p_c^{dir}(2)$ for λ being the smallest point in the support of F where $p_c^{dir}(2)$ is the critical probability for directed Bernoulli bond percolation, and Zhang [112] for $\boldsymbol{x} = \boldsymbol{e}_1$ and edge weight distributions having finite exponential moments and satisfying $F(\lambda) \geq p_c^{dir}(2), F(\lambda-) = 0, \lambda > 0$.

Hence the only nontrivial bound known for χ is $\chi \leq 1/2$. Note that the bound $0 \leq \chi \leq 1/2$ along with the scaling relation (which is unproven) would imply that $1/2 \leq \xi \leq 3/4$. In fact using a closely related exponent χ' which satisfies $\chi' \geq 2\xi - 1$ and $\chi' \leq 1/2$ (see Newman and Piza [87], Kesten [68] and Alexander [1]), it was proved in [87] that $\xi \leq 3/4$ in any dimension for paths in the directions of strict convexity of the limit shape. Moreover, Licea, Newman and Piza [76], comparing appropriate variance bounds, proved that $\xi(d) \geq 1/(d+1)$ for all dimensions d. They also proved that $\xi'(d) \geq 1/2$ for all dimensions d for a related exponent ξ' of ξ.

The next natural question is about the tail behavior and distributional convergence of the random variables $a(\mathbf{0}, n\boldsymbol{x})$ as $\boldsymbol{x}$ remains fixed and $n \to \infty$. Kesten [68] used martingale methods to prove that $\mathbb{P}(|a(\mathbf{0}, n\boldsymbol{e}_1) - \mathbb{E}[a(\mathbf{0}, n\boldsymbol{e}_1)]| \geq t\sqrt{n}) \leq c_1 e^{-c_2 t}$ for all $t \leq c_3 n$ for some constants $c_i > 0$, where $\boldsymbol{e}_1$ is the unit vector $(1, 0, \ldots, 0)$. Later, Talagrand [106] used his famous isoperimetric inequality to prove that

$$\mathbb{P}(|a(\mathbf{0}, n\boldsymbol{x}) - M]| \geq t\sqrt{n}\,\|\boldsymbol{x}\|) \leq c_1 e^{-c_2 t^2}$$

for all $t \leq c_3 n$ for some constants $c_i > 0$ where M is a median of $a(\mathbf{0}, n\boldsymbol{x})$ and $\boldsymbol{x} \in \mathbb{Z}^d$. Both these results were proved for distributions F having finite exponential moments and satisfying $F(0) < p_c(d)$.

From these inequalities, one might naïvely expect that a central limit theorem holds for $a(\mathbf{0}, n\boldsymbol{x})$. However, the situation is probably much more complex, and it may not be true that a Gaussian CLT holds. For critical first-passage percolation (assuming $F(0) = 1/2$ and F has bounded support) in two dimensions a Gaussian CLT was proved by Newman and Zhang [69]. However, this is sort of a degenerate case since here $\mathbb{E}[a(\mathbf{0}, n\boldsymbol{x})]$ and $\mathrm{Var}(a(\mathbf{0}, n\boldsymbol{x}))$ are both of order $\log n$ (see Chayes, Chayes and Durrett [32], and Newman and Zhang [69]). When $F(0) < 1/2$, we do not know of any distributional convergence result in any dimension.

Convergence to the Tracy-Widom law is known for *directed* last-passage percolation in $\mathbb{Z}^2$ under very special conditions (see Subsection 3.1.4 for details), but the techniques do not carry over to the undirected case. Naturally, one may expect that convergence to something like the Tracy-Widom distribution may hold for undirected first-passage percolation also, but surprisingly, this does not seem to be the case. In the following subsection, we present our main result: a Gaussian CLT for undirected first-passage percolation when the paths are restricted to lie in thin cylinders. This gives rise to an interesting question: as the cylinders become thicker, when does the CLT break down, if it does?

3.1.3 Our results

We consider first-passage percolation on $\mathbb{Z}^d$ with height restricted by an integer h (that will be allowed to grow with n). We assume that the edge weight distribution F satisfies a standard admissibility criterion, defined below.

Definition 3.1.1. *Given the dimension d, we call a probability distribution function F on the real line* admissible *if F is supported on $[0,\infty)$, is nondegenerate and we have $F(\lambda) < p_c(d)$ where λ is the smallest point in the support of F and $p_c(d)$ is the critical probability for Bernoulli bond percolation in $\mathbb{Z}^d$.*

For simplicity we will consider only first-passage time from $\mathbf{0}$ to $n\boldsymbol{e}_1$ where $\boldsymbol{e}_1$ is the first coordinate vector. The same method can be used to prove similar results for $a(\mathbf{0}, n\boldsymbol{x})$ where $\boldsymbol{x}$ has rational coordinates. Define $a_n(h)$ as the first-passage time to the point $n\boldsymbol{e}_1$ from the origin in the graph $\mathbb{Z} \times [-h,h]^{d-1}$, formally

$$a_n(h) := \inf\{\omega(\mathcal{P}) \mid \mathcal{P} \text{ is a path from } \mathbf{0} \text{ to } n\boldsymbol{e}_1 \text{ in } \mathbb{Z} \times [-h,h]^{d-1}\}.$$

Here, by $[-h,h]$ we mean the subset $[-h,h]\cap\mathbb{Z}$ of $\mathbb{Z}$. Informally, $a_n(h)$ is the minimal passage time over all paths which deviate from the straight line path joining the two end points by a distance at most h. Note that by the definition of the exponent ξ we have $a_n(h) = a(\mathbf{0}, n\boldsymbol{e}_1)$ with high probability when $h \gg n^{\xi}$. We also consider cylinder first-passage time (see Smyth and Wierman [101], Grimmett and Kesten [52]). A path $\mathcal{P}$ from $\mathbf{0}$ to $n\boldsymbol{e}_1$ is called a cylinder path if it is contained within the $x_1 = 0$ and $x_1 = n$ planes. We define

$$\begin{aligned} t_n(h) &:= \inf\{\omega(\mathcal{P}) \mid \mathcal{P} \text{ is a path from } \mathbf{0} \text{ to } n\boldsymbol{e}_1 \text{ in } [0,n] \times [-h,h]^{d-1}\} \text{ and} \\ T_n(h) &:= \inf\{\omega(\mathcal{P}) \mid \mathcal{P} \text{ is a path connecting } \{0\} \times [-h,h]^{d-1} \text{ and} \\ &\qquad\qquad \{n\} \times [-h,h]^{d-1} \text{ in } [0,n] \times [-h,h]^{d-1}\}. \end{aligned}$$

Clearly $a_n(h), t_n(h)$ and $T_n(h)$ are non-increasing in h for any $n \geq 1$. Our main result is that for cylinders that are 'thin' enough, we have Gaussian CLTs for $a_n(h), t_n(h)$ and $T_n(h)$ after proper centering and scaling.

Theorem 3.1.2. *Suppose that the edge-weights ω_e's are i.i.d. random variables from an admissible distribution F. Suppose $\mathbb{E}[\omega^p] < \infty$ for some $p > 2$. Let $\{h_n\}_{n\geq 1}$ be a* *sequence of integers satisfying* $h_n = o(n^\alpha)$ *where*

$$\alpha < \frac{1}{d+1+2(d-1)/(p-2)}$$

Then we have

$$\frac{a_n(h_n) - \mathbb{E}[a_n(h_n)]}{\sqrt{\mathrm{Var}(a_n(h_n))}} \xrightarrow{\mathrm{w}} N(0,1) \text{ as } n \to \infty.$$

In particular, if $\mathbb{E}[\omega^p] < \infty$ *for all* $p \geq 1$ *then the CLT holds when* $h_n = o(n^\alpha)$ *with* $\alpha < 1/(d+1)$*. If* $h_n = O(1)$ *then the* $F(\lambda) < p_c(d)$ *condition is not needed. Moreover, the same result is true for* $t_n(h_n)$ *and* $T_n(h_n)$*.*

In Section 3.2, we will present a generalization of this result (Theorem 3.2.1) to cylinders of the form $\mathbb{Z} \times G_n$ where $\{G_n\}$ is an arbitrary sequence of undirected connected graphs.

Theorem 3.1.2 give rise to a new exponent $\gamma(d)$ defined as

$$\gamma(d) := \sup\left\{\alpha : \frac{a_n(n^\alpha) - \mathbb{E}[a_n(n^\alpha)]}{\sqrt{\mathrm{Var}(a_n(n^\alpha))}} \xrightarrow{\mathrm{w}} N(0,1) \text{ as } n \to \infty\right\}.$$

Clearly we have $\gamma(d) \geq 1/(d+1)$ for F having all moments finite and satisfying the conditions in Theorem 3.1.2. Is $\gamma(d)$ actually equal to $1/(d+1)$? We do not have the answer for that yet. However for $d = 2$, numerical simulations suggest that

Conjecture 3.1.3. *For any* $d = 2$, $\gamma(d) = 2/3$.

An interesting feature of the proof of Theorem 3.1.2 is that while it is relatively easy to get a CLT for cylinders of width n^α for α sufficiently small, to go all the way up to $\alpha = 1/(d+1)$ one needs a somewhat complicated 'renormalization' argument that has to be taken to a certain depth of recursion, where the depth depends on how close α is to $1/(d+1)$. We are not sure whether this renormalization step is fundamental to the problem or just an artifact of our proof.

A deficiency of Theorem 3.1.2 is that we do not have formulas for the mean and the variance of $a_n(h_n)$. Still, we have something: the following result states that under the hypotheses of Theorem 3.1.2 the mean grows linearly with n and the growth rate does not depend on h_n as long as $h_n \to \infty$. It also gives upper and lower bounds for the variance of $a_n(h_n)$.

Proposition 3.1.4. *Let* $\mu_n(h_n)$ *and* $\sigma_n^2(h_n)$ *be the mean and variance of* $a_n(h_n)$*. Assume that* $h_n \to \infty$ *as* $n \to \infty$*. Then*

$$\lim_{n\to\infty} \frac{\mu_n(h_n)}{n} = \nu(\boldsymbol{e}_1),$$

where $\nu(\boldsymbol{e}_1)$ *is defined as in* (3.1)*. Moreover, if* F *is admissible we have*

$$c_1 \frac{n}{h_n^{d-1}} \le \sigma_n^2(h_n) \le c_2 n$$

for some absolute constants $c_1, c_2 > 0$ *depending only on* d *and* F*. If* $h_n = h$ *for all* n *for fixed* $h \in (0,\infty)$*, then both* $\lim_{n\to\infty} \mu_n(h)/n$ *and* $\lim_{n\to\infty} \sigma_n^2(h)/n$ *exist and are positive for any non-degenerate distribution* F *on* $[0,\infty)$*, but their values depend on* h*.*

In fact when $h_n = h$ for all n for fixed $h \in (0,\infty)$, we can say much more. Define $\mu(h) := \lim_{n\to\infty} \mu_n(h)/n$ and $\sigma^2(h) := \lim_{n\to\infty} \sigma_n^2(h)/n$. Existence of the limits follow from Proposition 3.1.4. Now consider the continuous process $X(\cdot)$ defined by $X(n) = t_n(h) - n\mu(h)$ for $n \in \{0,1,\ldots\}$ and extended by linear interpolation. Then we have the following result.

Proposition 3.1.5. *Assume that* $\mathbb{E}[\omega^p] < \infty$ *for some* $p > 2$ *where* $\omega \sim F$*. Then the scaled process* $\{(n\sigma^2(h))^{-1/2} X(nt)\}_{t\ge 0}$ *converges in distribution to the standard Brownian motion as* $n \to \infty$*.*

Here we mention that even though we have lower and upper bounds for the variance of $a_n(h_n)$ in Proposition 3.2.3, none of the bounds seem to be the correct one, at least when $d = 2$ when $h_n \to \infty$. In Section 3.6 we provide some heuristic justification for that. In fact numerical simulation suggests the following.

Conjecture 3.1.6. *For* $d = 2$ *and* $h_n \ll n^{2/3}$*,* $\mathrm{Var}(a_n(h_n)) = \Theta(n h_n^{-1/2})$*.*

Finally let us mention that a variant of Theorem 3.1.2 can be proved for the first-passage *site* percolation model also. Here instead of edge-weights $\{\omega_e \mid e \in E\}$ we have vertex weights $\{\omega_{\boldsymbol{x}} \mid \boldsymbol{x} \in \mathbb{Z}^d\}$ and travel time for a path $\mathcal{P}$ is defined by $\omega(\mathcal{P}) = \sum_{v\in\mathcal{P}} \omega_v$. The same proof technique should work.

3.1.4 Comparison with directed last-passage percolation

In all the previous discussions we used undirected first-passage times. A directed model is obtained when instead of all paths, one considers only directed paths. A directed path is a path that moves only in the positive direction at each step (e.g. in $d = 2$, the path moves only up and right). Let us restrict ourselves to $d = 2$ henceforth. The directed (site/bond) last-passage time to the point (n,h) starting from the origin is defined as

$$L_\uparrow^s(n,h) := \sup\{\omega(\mathcal{P}) \mid \mathcal{P} \in \Pi(n,h)\},$$

where $\Pi(n,h)$ is the set of all directed paths from $(0,0)$ to (n,h). Note that all the paths in $\Pi(n,h)$ are inside the rectangle $[0,n] \times [0,h]$.

The directed last-passage site percolation model in $d = 2$ has received particular attention in recent years, due to its myriad connections with the totally asymmetric simple exclusion process, queuing theory and random matrix theory. An important breakthrough, due to Johansson [63], says that when the vertex weights $\omega_{\boldsymbol{x}}$'s are i.i.d. geometric random variables, $L_{\uparrow}^{s}(n,n)$ has fluctuations of order $n^{1/3}$ and has the same limiting distribution as the largest eigenvalue of a GUE random matrix upon proper centering and scaling. (This is also known as the Tracy-Widom law.) Moreover, this holds if we replace $L_{\uparrow}^{s}(n,n)$ with $L_{\uparrow}^{s}(n,\lfloor\rho n\rfloor)$ for any $\rho \in (0,1]$. This continues to hold if one replaces geometric by exponential or bernoulli random variables [64, 51], but no greater generality has been proved.

Since the above result holds for arbitrary $\rho > 0$, one can speculate whether we can actually take $\rho \to 0$ as $n \to \infty$, i.e. look at directed last-passage percolation in thin rectangles. Indeed, the analog of Johansson's result in this setting was proved by several authors [6, 16, 105] in recent years for quite a general class of vertex weight distributions, provided the rectangles are 'thin' enough. This result contrasts starkly with our result about the Gaussian behavior of first-passage percolation in thin rectangles. A version of the result for last-passage percolation in thin rectangles is stated in Theorem 3.1.7. We recall that the GUE Tracy-Widom distribution has distribution function

$$F_2(x) := \exp\left(-\int_x^\infty (s-x)q^2(s)\,ds\right),$$

where $q(\cdot)$ solves the Painlevé II equation $q'' = 2q^3 + xq$ subject to the condition $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$ and $\mathrm{Ai}(x)$ is the Airy function.

Theorem 3.1.7 (see [6, 16, 105])**.** *Suppose that the vertex weights $\{w_{ij} : (i,j) \in \mathbb{Z}^2\}$ are i.i.d. random variables with mean μ, variance σ^2 and finite p-th moment for some $2 < p < \infty$. Then for the directed first-passage site percolation we have*

$$\frac{L_{\uparrow}^{s}(n,k) - \mu(n+k) - 2\sigma\sqrt{nk}}{\sigma k^{-1/6} n^{1/2}} \xrightarrow{\mathrm{w}} F_2$$

as $n \to \infty$ where $k = o(n^\alpha)$ for some $\alpha < \frac{6}{7}\left(\frac{1}{2} - \frac{1}{p}\right)$ and F_2 is the GUE Tracy-Widom distribution.

In particular, if all moments of the vertex weights are finite, then the result is true for $\alpha < 3/7$. The same result holds if we replace first-passage time $T_{\uparrow}^{s}(n,k)$ by last passage time $L_{\uparrow}^{s}(n,k)$.

Note that in the definition of first-passage time one can restrict the paths to self-avoiding paths (for which all the visited vertices are distinct), as for any path removing a loop decreases the weight of the path. In directed first and last-passage percolation one consider self-avoiding paths of minimal length (which is $n + k$) and number of such paths is $\binom{n+k}{k} = e^{\Theta(k \log n)} = e^{o(n)}$ when $k = o(n^\alpha)$ for some $\alpha < 1$. But in the undirected case number of paths is exponential in n. This follows easily from the fact that, number of self-avoiding paths from $(0,0)$ to $(n,1)$ in the rectangle $\{0,1,\ldots,n\} \times \{0,1\}$ is 2^n. In fact, even if in the previous example one look at the number of paths having length $n+1+2i$ it is $\binom{n+1}{2i+1}$ for $i = 0,1,\ldots,\lfloor n/2 \rfloor$ and the number is exponential when $i = \Theta(n)$.

In [105], Suidan derived universality of oriented last passage percolation for thin rectangles from the result for exponential edge weights using a theorem of Chatterjee [25, 26] which is inspired by Lindeberg's proof of the Central Limit Theorem. In our case that strategy will not work as the number of paths is exponential in n.

3.1.5 Structure of the chapter

The chapter is organized in the following way. In Section 3.2 we state a general result that encompasses Theorem 3.1.2. In Section 3.3 we prove the asymptotic behavior of the mean of $a_n(G_n)$. Sections 3.4 and 3.5 contain, respectively, the lower bound for the variance and upper bounds for general central moments of $a_n(G_n)$. Section 3.6 contains a different proof for the case of exponential edge weights, which also indicates why the variance bounds are not tight in general. In Section 3.7 we prove the generalized version of Theorem 3.1.2 and in Section 3.8 we consider the case of first-passage time across $[0,n] \times G$ when G is a fixed graph. Finally, in Section 3.9 we provide some numerical results in support of our conjectures.

3.2 Generalization

In this section, we generalize the theorems of Section 3.1 to first-passage percolation on graphs on the form $\mathbb{Z} \times G_n$, where $\{G_n\}$ is an arbitrary increasing sequence of undirected graphs.

Before stating the results, let us fix our notations. The set $\{a, a+1, \ldots, b\}$ with the nearest neighbor graph structure will be denoted by $[a,b]$. When $a = 0$, we will simply write $[b]$ instead of $[0,b]$. Throughout the rest of the article we will consider the undirected first-passage bond percolation model with edge weight distribution F, as defined in the previous section. Let μ and σ^2 be the mean and the variance of F. We will use the standard notations $a_n = O(b_n)$ and $a_n = o(b_n)$, respectively, in the case $\sup_{n \geq 1} a_n/b_n < \infty$ and $\lim_{n\to\infty} a_n/b_n = 0$.

For two finite connected graphs H and G, we define the product graph structure on $H \times G$ in the natural way, that is, there is an edge between (u,w) and (v,z) if and only if either (u,v) is an edge in H and $w = z$, or $u = v$ and (w,z) is an edge in G.

We will consider first-passage percolation on a special class of product graphs. Fix an integer n and a connected graph G with a distinguished vertex $o \in G$. Let $a_n(G)$

denote the first-passage time from $(0, o)$ to (n, o) in $\mathbb{Z} \times G$. That is,

$$a_n(G) := \inf\{\omega(\mathcal{P}) \mid \mathcal{P} \text{ is a path from } (0,o) \text{ to } (n,o) \text{ in } \mathbb{Z} \times G\}$$

where $\omega(\mathcal{P}) := \sum_{e\in\mathcal{P}} \omega_e$ is weight of the path $\mathcal{P}$. We define the cylinder first-passage time $t_n(G)$ as

$$t_n(G) := \inf\{\omega(\mathcal{P}) : \mathcal{P} \text{ is a path from } (0,o) \text{ to } (n,o) \text{ in } [0,n] \times G\}.$$

We also define the side-to-side (cylinder) first-passage time as follows:

$$\begin{aligned} T_{a,b}(G) := \min\{\omega(\mathcal{P}) \mid \mathcal{P} &\text{ is a path connecting the two sides} \\ &\{a\} \times G \text{ and } \{b\} \times G \text{ in } [a,b] \times G\}, \end{aligned} \tag{3.2}$$

that is, $T_{a,b}(G)$ is the minimum weight among all paths that join the right boundary of the product graph $[a, b] \times G$ to the left boundary of it. Note that it is enough to consider only those paths that start from some vertex in $\{a\} \times G$ and end at some vertex in $\{b\} \times G$, and lie in the set $[a + 1, b - 1] \times G$ throughout except for the first and last edges. One implication of this fact is that $T_{a,b}(G)$ is independent of the weights of the edges in the left and right boundaries $\{a\} \times G, \{b\} \times G$. We will write $T_{0,n}(G)$ simply as $T_n(G)$.

Now consider a nondecreasing sequence of connected graphs $G_n = (V_n, E_n)$, $n \geq 1$. By 'nondecreasing' we mean that G_n is a subgraph (need not be induced) of G_{n+1} for all n. Let o be a distinguished vertex in G_1, which we will call the *origin* of G_1. Then $o \in G_n$ for all n. Let k_n and d_n be the number of edges and the diameter of G_n, respectively.

Our object of study is first-passage percolation on the product graph $\mathbb{Z} \times G_n$ with i.i.d. edge weights from the distribution F. In particular, we wish to understand the behavior of the first-passage time $a_n(G_n)$ from $(0, o)$ to (n, o).

The main result of this section is the following.

Theorem 3.2.1. *Let G_n be a nondecreasing sequence of connected graphs with a fixed origin o. Let d_n and k_n be the diameter and the number of edges in G_n. Suppose that as $n \to \infty$, $k_n = O(d_n^\theta)$ for some fixed $\theta \geq 1$. Let $a_n(G_n)$ be the first-passage percolation time from $(0, o)$ to (n, o) in the graph $\mathbb{Z} \times G_n$. Suppose that a generic edge weight ω satisfies $\mathbb{E}[\omega^p] < \infty$ for some $p > 2$. Then we have*

$$\frac{a_n(G_n) - \mathbb{E}[a_n(G_n)]}{\sqrt{\mathrm{Var}(a_n(G_n))}} \xrightarrow{\text{w}} N(0,1)$$

as $n \to \infty$ provided one of following holds:

(A) *There is a fixed connected graph G such that $G_n = G$ for all $n \geq 1$, or*

(B) *G_n's are connected subgraphs of $\mathbb{Z}^{d-1}$ for some $d > 1$, the edge weight distribution is admissible and $d_n = o(n^\alpha)$, where*

$$\alpha < \frac{1}{2 + \theta + 2\theta/(p-2)}.$$

Moreover, the same result holds for $t_n(G_n), T_n(G_n)$ in place of $a_n(G_n)$.

Clearly, this theorem implies Theorem 3.1.2 by taking $G_n = [-h_n, h_n]^{d-1}$ with $d_n = 2h_n(d-1)^{1/2}$ and $\theta = d - 1$. Throughout the rest of the paper we will consider the case of general sequence G_n.

As we remarked earlier we do not have explicit formulas for the mean and the variance of $a_n(G_n)$. The following result is the generalization of the 'mean part' of Proposition 3.1.4.

Proposition 3.2.2. *Consider the setup introduced above. Then the limit*

$$\nu := \lim_{n\to\infty} \frac{1}{n} \mathbb{E}[a_n(G_n)]$$

exists and we have

$$\nu n \le \mathbb{E}[a_n(G_n)] \le \mu n \text{ for all } n.$$

Moreover, $\nu > 0$ if $G_n = G$ for all $n \ge 1$ or G_n's are subgraphs of $\mathbb{Z}^{d-1}$ and $F(0) < p_c(d)$. In particular, when $G_n = [-h_n, h_n]^{d-1}$ and $h_n \to \infty$ as $n \to \infty$, we have $\nu = \nu(\boldsymbol{e}_1)$, where $\nu(\boldsymbol{e}_1)$ is defined as in (3.1). We also have

$$\mathbb{E}[a_n(G_n)] \le \mathbb{E}[t_n(G_n)] \le \mathbb{E}[T_n(G_n)] + 2\mu d_n \le \mathbb{E}[a_n(G_n)] + 2\mu d_n$$

for all n.

Now let us state the upper and lower bounds for the variance of $a_n(G_n)$, i.e. the 'variance part' of Proposition 3.1.4.

Proposition 3.2.3. *Under the condition of Theorem 3.2.1 we have*

$$c_1 \frac{n}{k_n} \le \mathrm{Var}(a_n(G_n)) \le c_2 n$$

for some positive constants c_1, c_2 that do not depend on n. Moreover,

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Var}(a_n(G_n))$$

exists for all non-degenerate distribution F on $[0, \infty)$ when $G_n = G$ for all n. The above results hold for $t_n(G_n)$ and $T_n(G_n)$.

In fact when $G_n = G$ for all $n \geq 1$, we can say much more as in Proposition 3.1.5. Define

$$\mu(G) := \lim_{n\to\infty} \frac{\mathbb{E}[a_n(G)]}{n} \text{ and } \sigma^2(G) := \lim_{n\to\infty} \frac{\mathrm{Var}(a_n(G))}{n}. \tag{3.3}$$

Existence and positivity of the limits follow from Propositions 3.2.2 and 3.2.3. Consider the continuous process $X(\cdot)$ defined by $X(n) = t_n(G) - n\mu(G)$ for $n \geq 0$ and extended by linear interpolation. Then we have the following result.

Proposition 3.2.4. *Assume that the generic edge weight ω is non-degenerate and satisfies $\mathbb{E}[\omega^p] < \infty$ for some $p > 2$. Then the scaled process*

$$\{(n\sigma^2(G))^{-1/2} X(nt)\}_{t\geq 0}$$

converges in distribution to the standard Brownian motion as $n \to \infty$.

3.3 Estimates for the mean

In this section we will prove Proposition 3.2.2. We will break the proof into several lemmas. Lemma 3.3.1 shows that the random variables $a_n(G_n), t_n(G_n)$ and $T_n(G_n)$ are close in L^p norm when the diameter d_n of G_n is small.

Corollary 3.3.1. *We have*

$$T_n(G_n) \leq a_n(G_n) \leq t_n(G_n) \text{ for all } n.$$

Moreover we have

$$\mathbb{E}[|t_n(G_n) - T_n(G_n)|^p] \leq 2^p d_n^p\, \mathbb{E}[\omega^p] \text{ for all } n \geq 1$$

when $\mathbb{E}[\omega^p] < \infty$ for some $p \geq 1$ and a typical edge weight $\omega \sim F$.

Proof. Fix any path $\mathcal{P}$ from $(0, o)$ to (n, o) in $\mathbb{Z} \times G_n$. The path $\mathcal{P}$ will hit $\{0\} \times G_n$ and $\{n\} \times G_n$ at some vertices. Let $(0, u)$ be the vertex where $\mathcal{P}$ hits $\{0\} \times G_n$ the last time and (n, v) be the vertex where $\mathcal{P}$ hits $\{n\} \times G_n$ the first time after hitting $(0, u)$. The path segment of $\mathcal{P}$ from $(0, u)$ to (n, v) lies inside $[n] \times G_n$ and by non-negativity of edge weights we have $\omega(\mathcal{P}) \geq T_n(G_n)$. Since this is true for any path $\mathcal{P}$ joining $(0, o)$ to (n, o) in $\mathbb{Z} \times G_n$, we have $T_n(G_n) \leq a_n(G_n)$.

Clearly $a_n(G_n) \leq t_n(G_n)$. Combining the two inequalities, we see that

$$T_n(G_n) \leq a_n(G_n) \leq t_n(G_n) \text{ for all } n.$$

Since the number of paths joining the left side $\{0\} \times G_n$ to the right side $\{n\} \times G_n$ in $[0, n] \times G_n$ is finite there is a path achieving the minimal weight $T_n(G_n)$. Choose such a path $\mathcal{P}^*$ using a deterministic rule. Suppose that the path $\mathcal{P}^*$ starts at $(0,u)$ and ends at (n,w). As we remarked earlier in Section 3.2 the random variables $T_n(G_n), \mathcal{P}^*, u, w$ are independent of the edge weights ω_e where e is an edge in $\{0\} \times G_n$ or $\{n\} \times G_n$.

Let $\mathcal{P}(u), \mathcal{P}(w)$ be some minimal length paths in G_n joining o, u and o, w respectively. We have $t_n(G_n) - T_n(G_n) \leq S_n$ where S_n is the sum of edge weights in the paths $\{0\} \times \mathcal{P}(u)$ and $\{n\} \times \mathcal{P}(w)$ and hence

$$\mathbb{E}[|t_n(G_n) - T_n(G_n)|^p] \leq \mathbb{E}[S_n^p].$$

Moreover by independence of u, w and the edge weights in $\{0, n\} \times G_n$ we have $\mathbb{E}[S_n^p | u, w] \leq (|\mathcal{P}(u)| + |\mathcal{P}(w)|)^p \, \mathbb{E}[\omega^p]$. By definition of diameter we have $|\mathcal{P}(u)| + |\mathcal{P}(w)| \leq 2d_n$ and thus we are done. □

The following lemma combined with Lemma 3.3.1 completes half of the proof of Proposition 3.2.2. Recall that $\{G_n\}$ is a nondecreasing sequence of finite connected graphs.

Corollary 3.3.2. *The limit*

$$\nu = \lim_{n\to\infty} \frac{\mathbb{E}[a_n(G_n)]}{n}$$

exists and we have

$$\nu n \leq \mathbb{E}[a_n(G_n)] \leq \mu n \text{ for all } n.$$

Moreover, we have $\nu < \mu$ if $d_n \geq 1$ and F is non-degenerate.

Proof. Considering the straight line path from $(0, o)$ to (n, o) it is easy to see that $\mathbb{E}[a_n(G_n)] \leq \mu n$. The existence of the limit is easily obtained from subadditivity as follows. Fix n, m. Consider G_n and G_m as subgraphs of G_{n+m}. Let $a_{n,n+m}(G_m)$ denote the first-passage time in $\mathbb{Z} \times G_m$ from (n, o) to $(n+m, o)$. Clearly $a_{n,n+m}(G_m) \stackrel{\mathrm{d}}{=} a_m(G_m)$. Joining the minimal weight paths from $(0, o)$ to (n, o) achieving the weight $a_n(G_n)$ and from (n, o) to $(n+m, o)$ achieving the weight $a_{n,n+m}(G_m)$, we get a path in $\mathbb{Z} \times G_{n+m}$ from $(0, o)$ to $(n+m, o)$. Clearly

$$a_{n+m}(G_{n+m}) \leq a_n(G_n) + a_{n,n+m}(G_m).$$

Now taking expectation in both sides and using the subadditive lemma we have

$$\nu := \lim_{n\to\infty} \frac{\mathbb{E}[a_n(G_n)]}{n}$$

exists and equals $\inf_{n\geq 1} \mathbb{E}[a_n(G_n)]/n$.

To show that $\nu < \mu$ it is enough to consider the one edge graph $G_n = G = \{0, 1\}$ and n even. Consider the following two paths from $(0, 0)$ to $(2n, 0)$. One is the straight line path. The other is the path connecting $(0,0), (0,1), (1,1), (1,0), (2,0)$ and repeating the same pattern. Clearly we have $\mathbb{E}[a_{2n}(G)] \leq \mu n + n\, \mathbb{E}[\min\{\omega_1, \omega_2 + \omega_3 + \omega_4\}]$ where ω_i's are i.i.d. from F. From here it is easy to see that $\nu < \mu$. □

We complete the proof of Proposition 3.2.2 by finding lower bound for ν under appropriate conditions. Recall that $\nu(\boldsymbol{e}_1) > 0$ iff $F(0) < p_c(d)$ where $\boldsymbol{e}_1$ is the first coordinate vector in $\mathbb{Z}^d$ and $\nu(\boldsymbol{x})$ is defined as in (3.1).

Corollary 3.3.3. *Suppose G_n's are subgraphs of $\mathbb{Z}^{d-1}$. Then the limit ν in Lemma 3.3.2 satisfies*

$$\nu \geq \nu(\boldsymbol{e}_1)$$

where $\nu(\boldsymbol{e}_1)$ is as defined in (3.1). *Equality holds when $G_n = [-h_n, h_n]^{d-1}$ with $h_n \to \infty$ as $n \to \infty$. Moreover, the limit ν is positive if $G_n = G$ for all n.*

Proof. First suppose that $G_n = G$ for all n and G has v vertices. It is easy to see that $\mathbb{E}[a_n(G_n)] \geq n\,\mathbb{E}[Y]$ where Y is the minimum of v i.i.d. random variables each having distribution F, because any path from $(0, o)$ to (n, o) must contain at least one edge of the form $((k, u), (k + 1, u))$ for each $k = 0, \ldots, n-1$. Since $\mathbb{E}[Y] > 0$, it follows that $\nu > 0$.

Now consider the case when G_n's are subgraphs of $\mathbb{Z}^{d-1}$ (we will match o with the origin in $\mathbb{Z}^{d-1}$). Then $\mathbb{Z} \times G_n$ is a subgraph of $\mathbb{Z}^d$ with $(0, o) = \mathbf{0}$ and $(n, o) = n\boldsymbol{e}_1$ where $\mathbf{0}$ and $\boldsymbol{e}_1$ denote the origin and the first coordinate vector in $\mathbb{Z}^d$. Clearly we have $a(\mathbf{0}, n\boldsymbol{e}_1) \leq a_n(G_n)$ for all n. Diving both sides by n and taking expectations we have

$$\nu = \lim_{n\to\infty} \frac{1}{n}\,\mathbb{E}[a_n(G_n)] \geq \lim_{n\to\infty} \frac{1}{n}\,\mathbb{E}[a(\mathbf{0}, n\boldsymbol{e}_1)] = \nu(\boldsymbol{e}_1).$$

To prove that $\nu = \nu(\boldsymbol{e}_1)$ when $G_n = [-h_n, h_n]^{d-1}$, break the cylinder graph $[n] \times G_n$ into smaller cylinder graphs of length $\lfloor l_n/C \rfloor$ for some fixed constant $C > 0$ where $l_n = \min\{n^{1/2}, h_n\}$. Note that concatenating paths from $(il_n/C, o)$ to $((i+1)l_n/C, o)$ for $i = 0, 1, \ldots$ we get a path from $(0, o)$ to (n, o). Let $n = m\lceil l_n/C \rceil + r$ with $r < \lceil l_n/C \rceil$. Thus we have

$$\mathbb{E}[a_n(G_n)] \leq m\,\mathbb{E}[X(\lceil l_n/C \rceil, l_n)] + \mathbb{E}[X(r, l_n)] \tag{3.4}$$

where

$$X(n, h) := \inf\{\omega(\mathcal{P}) \mid \mathcal{P} \text{ is a path from } (0, o) \text{ to } (n, o) \text{ that lies in the rectangle } [1, n-1] \times [-h, h]^{d-1} \text{ except for the first and last edge}\}.$$

Dividing both sides of (3.4) by n and taking limits (note $l_n = o(n)$ and $l_n \to \infty$ as $n \to \infty$) we have

$$\nu := \lim_{n\to\infty} \frac{1}{n}\,\mathbb{E}[a_n(G_n)] \leq \liminf_{n\to\infty} \frac{\mathbb{E}[X(\lceil n/C \rceil, n)]}{\lceil n/C \rceil} \leq \lim_{n\to\infty} \frac{\mathbb{E}[X(n, \lfloor Cn \rfloor)]}{n}$$

for any $C > 0$. The last limit exists by subadditivity. Denote the last limit by $\alpha(C)$ which also satisfies $\alpha(C) = \inf_n \mathbb{E}[X(n, \lfloor Cn \rfloor)/n$. Now let us consider the unrestricted cylinder percolation time $t(\mathbf{0}, n\boldsymbol{e}_1)$ defined as the minimum weight among all paths from $\mathbf{0}$ to $n\boldsymbol{e}_1$ lying in the vertical strip $0 < x_1 < n$ except for the first and the last edge. From standard results in first-passage percolation theory (see Section 5.1 in Smythe and Wierman [101] for a proof) we have

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}[t(\mathbf{0}, n\boldsymbol{e}_1)] = \nu(\boldsymbol{e}_1).$$

Now for fixed n, the random variables $X(n, \lfloor Cn \rfloor)$ are decreasing in C and $t(\mathbf{0}, n\boldsymbol{e}_1) = \lim_{C\to\infty} X(n, \lfloor Cn \rfloor)$. By monotone convergence theorem we have

$$\mathbb{E}[t(\mathbf{0}, n\boldsymbol{e}_1)] = \lim_{C\to\infty} \mathbb{E}[X(n, \lfloor Cn \rfloor)] \geq \limsup_{C\to\infty} \alpha(C)n \geq \nu n.$$

Dividing both sides by n and letting $n \to \infty$ we are done. □

3.4 Lower bound for the variance

Here we will prove the lower bound for the variance given in Proposition 3.2.3. First we will prove a uniform lower bound that holds for any n and G. Later we will specialize to the case $G = G_n$ for given n.

Corollary 3.4.1. *Let G be a subgraph of $\mathbb{Z}^{d-1}$ with diameter D and number of edges k. Let F be admissible. Then we have*

$$\mathrm{Var}(t_n(G)) \geq c_1 \frac{n}{k} \text{ and } \mathrm{Var}(T_n(G)) \geq c_1 \frac{n}{k}\left(1 - c_2 \frac{D}{n}\right) \tag{3.5}$$

for some absolute positive constants c_1, c_2 that depend only on d and F. The same result holds for all nondegenerate probability distributions F on $[0, \infty)$ with c_i depending only on G and F. In particular, when $D \leq n/(2c_2)$ we have

$$\mathrm{Var}(T_n(G)) \geq c_3 \frac{n}{k}$$

for all n, k for some absolute constant $c_3 > 0$.

Proof. Fix G and n. Let v be the number of vertices in G. Let $\{e_1, e_2, \ldots, e_N\}$ be a fixed enumeration of the edges in $[n] \times G$ where $N = (n+1)k + nv$ is the number of edges in that graph. For simplicity let us write $t_n(G)$ simply as t. Let $\mathcal{F}_i$ be the sigma-algebra generated by $\{\omega(e_1), \omega(e_2), \ldots, \omega(e_i)\}$ for $i = 0, 1, \ldots, N$. For simplicity we will write ω_i instead of $\omega(e_i)$. Also we will write $t(\boldsymbol{\omega})$ to explicitly write the dependence of t on the sequence of edge-weights $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_N)$.

Using Doob's martingale decomposition we can write the random variable $t - \mathbb{E}[t]$ as a sum of martingale difference sequences $\mathbb{E}[t|\mathcal{F}_i] - \mathbb{E}[t|\mathcal{F}_{i-1}], i = 1, 2, \ldots, N$. Since martingale difference sequences are uncorrelated we have the standard identity

$$\mathrm{Var}(t) = \sum_{i=1}^{N} \mathrm{Var}(\mathbb{E}[t|\mathcal{F}_i] - \mathbb{E}[t|\mathcal{F}_{i-1}]).$$

For $1 \le i \le N$, let $\hat{\boldsymbol{\omega}}^i$ denote the sequence of edge-weights $\boldsymbol{\omega}$ excluding the weight ω_i. Moreover, for $x \in \mathbb{R}^+$, we will write $(\hat{\boldsymbol{\omega}}^i, x)$ to denote the sequence of edge-weights where the weight of the edge e_j is ω_j for $j \neq i$ and x for $j = i$. Clearly we have $\boldsymbol{\omega} = (\hat{\boldsymbol{\omega}}^i, \omega_i)$ for $i = 1, 2, \ldots, N$. If η is a random variable distributed as F and is independent of $\boldsymbol{\omega}$, then we have $\mathbb{E}[t|\mathcal{F}_i] - \mathbb{E}[t|\mathcal{F}_{i-1}] = \mathbb{E}[t(\hat{\boldsymbol{\omega}}^i, \omega_i) - t(\hat{\boldsymbol{\omega}}^i, \eta)|\mathcal{F}_i]$. It is easy to see that (as $\mathrm{Var}(t) \ge \mathrm{Var}(\mathbb{E}[t|\mathcal{F}])$)

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[t(\hat{\boldsymbol{\omega}}^i, \omega_i) - t(\hat{\boldsymbol{\omega}}^i, \eta)|\mathcal{F}_i]) &\ge \mathrm{Var}(\mathbb{E}[\mathbb{E}[t(\hat{\boldsymbol{\omega}}^i, \omega_i) - t(\hat{\boldsymbol{\omega}}^i, \eta)|\mathcal{F}_i]|\omega_i]) \\ &= \mathrm{Var}(\mathbb{E}[t(\boldsymbol{\omega})|\omega_i]).\end{aligned}$$

Now for any random variable X we have $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}(X_1 - X_2)^2$ where X_1, X_2 are i.i.d. copies of X. Thus we have

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[t(\boldsymbol{\omega})|\omega_i]) &= \frac{1}{2}\mathbb{E}[(\mathbb{E}[t(\hat{\boldsymbol{\omega}}^i, \omega_i) - t(\hat{\boldsymbol{\omega}}^i, \eta)|\omega_i, \eta])^2] \\ &= \mathbb{E}[(\mathbb{1}_{\{\omega_i > \eta\}}\mathbb{E}[t(\hat{\boldsymbol{\omega}}^i, \omega_i) - t(\hat{\boldsymbol{\omega}}^i, \eta)|\omega_i, \eta])^2] \end{aligned} \tag{3.6}$$

where in the last line we have used the fact that ω_i and η are i.i.d. . Define

$$\Delta_i := \mathbb{E}[\mathbb{1}_{\{\omega_i > \eta\}}(t(\hat{\boldsymbol{\omega}}^i, \omega_i) - t(\hat{\boldsymbol{\omega}}^i, \eta))|\boldsymbol{\omega}] \tag{3.7}$$

for $i = 1, 2, \ldots, N$. From (3.6) we have $\mathrm{Var}(\mathbb{E}[t(\boldsymbol{\omega})|\omega_i]) \ge (\mathbb{E}[\Delta_i])^2$ for all i. Combining we have

$$\mathrm{Var}(t) \ge \sum_{i=1}^N (\mathbb{E}[\Delta_i])^2 \ge \frac{1}{N}\left(\sum_{i=1}^N \mathbb{E}[\Delta_i]\right)^2 = \frac{1}{N}(\mathbb{E}[g(\boldsymbol{\omega})])^2$$

where

$$g(\boldsymbol{\omega}) := \sum_{i=1}^N \Delta_i = \sum_{i=1}^N \mathbb{E}[\mathbb{1}_{\{\omega_i > \eta\}}(t(\boldsymbol{\omega}) - t(\hat{\boldsymbol{\omega}}^i, \eta))|\boldsymbol{\omega}].$$

Let $\mathcal{P}_*(\boldsymbol{\omega})$ be a minimum weight path for $\boldsymbol{\omega}$ chosen according to a deterministic rule. If the edge e_i is in $\mathcal{P}_*(\boldsymbol{\omega})$, we have

$$\mathbb{1}_{\{\omega_i > \eta\}}(t(\boldsymbol{\omega}) - t(\hat{\boldsymbol{\omega}}^i, \eta)) \ge \mathbb{1}_{\{\omega_i > \eta\}}(\omega_i - \eta) = (\omega_i - \eta)_+$$

as the weight of the path $\mathcal{P}_*(\boldsymbol{\omega})$ for the configuration $(\hat{\boldsymbol{\omega}}^i, \eta)$ is $t(\boldsymbol{\omega}) - \omega_i + \eta$. Thus we have

$$g(\boldsymbol{\omega}) \ge \sum_{i: e_i \in \mathcal{P}_*(\boldsymbol{\omega})} \mathbb{E}[(\omega_i - \eta)_+|\omega_i]. \tag{3.8}$$

Now define the function

$$h(x) = \mathbb{E}[(x - \eta)_+] \text{ where } \eta \sim F.$$

It is easy to see that $h(x) = 0$ iff $x \leq \lambda$ where λ is the smallest point in the support of F and $\mathbb{E}[h(\omega)] < \infty$.

Define a new set of edge weights $\omega_i' = h(\omega_i)$ for $i = 1, 2, \ldots, N$ with distribution function F'. Clearly ω_i''s are i.i.d. with $F'(0) = \mathbb{P}(h(\omega) = 0) = \mathbb{P}(\omega = \lambda)$. Moreover let $t(\boldsymbol{\omega}')$ be the cylinder first-passage time from $(0, o)$ to (n, o) in $[0, n] \times G$ with edge weights $\boldsymbol{\omega}'$. From (3.8) we have $g(\boldsymbol{\omega}) \geq t(\boldsymbol{\omega}')$. Now from Lemma 3.3.2 and 3.3.3 we have $\mathbb{E}[t(\boldsymbol{\omega}')] \geq \nu'(\boldsymbol{e}_1)n$ where $\nu'(\boldsymbol{e}_1)$ is as defined in (3.1) with edge weight distribution F' and $\nu'(\boldsymbol{e}_1) > 0$ as $F'(0) < p_c(d)$. Also note that $N = (n+1)k + nv \leq 3nk$. Thus, finally we have

$$\frac{1}{n}\operatorname{Var}(t) \geq \frac{1}{3k}\left(\frac{\mathbb{E}[t(\boldsymbol{\omega}')]}{n}\right)^2 \geq \frac{\nu'(\boldsymbol{e}_1)^2}{3k}. \tag{3.9}$$

Now assume that F is any non-degenerate distribution supported on $[0, \infty)$. From Lemma 3.3.3 we can see that $\mathbb{E}[t_n(G)] \geq cn$ for all n for some constant $c > 0$ depending on G and F. Thus we are done.

To prove the result for $T_n(G)$ we start with $T_n(G)$ in place of $t_n(G)$ and use $\mathbb{E}[T_n(G)] \geq \mathbb{E}[t_n(G)] - 2\mu D$ from Lemma 3.3.1 in (3.9). ☐

Proof of the lower bound in Proposition 3.2.3. From Lemma 3.3.1 we have

$$\begin{aligned} |\operatorname{Var}(a_n(G_n))^{1/2} - \operatorname{Var}(t_n(G_n))^{1/2}| &\leq (\mathbb{E}[|a_n(G_n) - t_n(G_n)|^2])^{1/2} \\ &\leq 2d_n(\mu^2 + \sigma^2)^{1/2} \end{aligned}$$

for all $n \geq 1$. Now under Theorem 3.2.1 we have $d_n = o(n^{1/(2+\theta)})$ which clearly implies that $d_n^2 = o(n/k_n)$ as $k_n = O(d_n^\theta)$. Thus by Lemma 3.4.1 we are done. Using Lemma 3.5.5 one can drop the condition $d_n = o(n^{1/(2+\theta)})$ when F is admissible. ☐

3.5 Upper bound for Central moments

In this section we will prove upper bounds for central moments of $a_n(G_n)$, $t_n(G_n)$ and $T_n(G_n)$, in particular the upper bound for variance of $a_n(G_n)$ stated in Proposition 3.2.3. Note that by Lemma 3.3.1 we have

$$\mathbb{E}[|t_n(G_n) - a_n(G_n)|^p] \leq \mathbb{E}[|t_n(G_n) - T_n(G_n)|^p] \leq \mathbb{E}[(2d_n\omega)^p]$$

for all n when $\mathbb{E}[\omega^p] < \infty$ for some $p \geq 2$ with $\omega \sim F$. Hence it is enough to prove bounds for $\mathbb{E}[|t_n(G_n) - \mathbb{E}[t_n(G_n)]|^p]$.

Fix $n \geq 1$ and a finite connected graph G. We will prove the following.

Proposition 3.5.1. *Let $\mathbb{E}[\omega^p] < \infty$ for some $p \geq 2$ and $F(0) < p_c(d)$ where $\omega \sim F$. Also suppose that G is a finite subgraph of $\mathbb{Z}^{d-1}$. Then for any $n \geq 1$ we have*

$$\mathbb{E}[|t_n(G) - \mathbb{E}[t_n(G)]|^p] \leq cn^{p/2}$$

where c is a constant depending only on p, d and F. Moreover, the same result holds with c depending on G without any restriction on $F(0)$. The above result holds for $a_n(G)$ and $T_n(G)$ when

$$D \leq Cn^{1/2}$$

for some absolute constant $C > 0$ where D is the diameter of G.

When F has finite exponential moments in some neighborhood of zero, one can use Talagrand's [106] strong concentration inequality along with Kesten's Lemma 3.5.5 to prove a much stronger result $\mathbb{P}(|t_n(G) - \mathbb{E}[t_n(G_n)]| \geq x) \leq 4e^{-c_1x^2/n}$ for $x \leq c_2 n$ for some constants $c_1, c_2 > 0$. Moreover, one can use moment inequalities due to Boucheron, Bousquet, Lugosi and Massart [21] to prove that the p-th moment is bounded by $n^{p/2}k^{p/2-1}$ for $p \geq 2$. But none of that gives what we need for the proof of Theorem 3.2.1, so we have to devise our own proof of Proposition 3.5.2.

The next two technical lemmas will be useful in the proof of Proposition 3.5.1. Proofs of the two technical lemmas and of Proposition 3.5.1 are given at the end of this section.

Corollary 3.5.2. *For any $p > 2$ and $x, y \in \mathbb{R}$ we have*

$$\left|x|x|^{p-2} - y|y|^{p-2}\right| \leq \max\{1, (p-1)/2\}|x-y|(|x|^{p-2} + |y|^{p-2}).$$

Corollary 3.5.3. *Let $\beta > 1, a, b \geq 0$. Let $y \geq 0$ satisfy $y^\beta \leq a + by$. Then*

$$y^{\beta-1} \leq a^{(\beta-1)/\beta} + b.$$

Before proving Proposition 3.5.1 we need to define a new random variable $L_n(G)$. Consider cylinder first-passage time $t_n(G)$ in $[n] \times G$. Call a path $\mathcal{P}$ from $(0, o)$ to (n, o) in $[n] \times G$ a weight minimizing path if its weight $\omega(\mathcal{P})$ equals $t_n(G)$. An edge e of $[n] \times G$ is called an *essential* edge if all weight minimizing paths pass through the edge e. Let $L_n(G)$ denote the number of essential edges given the edge weights $\boldsymbol{\omega}$. Clearly $L_n(G)$ is a random variable. Lemma 3.5.4 gives upper bound for the p-th central moment of $t_n(G)$ in terms of moments of $L_n(G)$. Roughly it says that the fluctuation of $t_n(G)$ around its mean behaves like square root of $L_n(G)$.

Corollary 3.5.4. *Let $\mathbb{E}[\omega^p] < \infty$ for some $p \geq 2$ where $\omega \sim F$. Then we have*

$$\begin{aligned}\mathbb{E}[|t_n(G) - \mathbb{E}[t_n(G)]|^p] \leq (2p)^{p/2}\, \mathbb{E}[L_n(G)^{p/2}]\, \mathbb{E}[\omega^2]^{p/2} \\ + 2^{p/2}(2p)^{p-2}\, \mathbb{E}[L_n(G)]\, \mathbb{E}[\omega^p]\end{aligned}$$

where $L_n(G)$ is the number of essential edges for $t_n(G)$.

Proof. The proof essentially is a general version of the Efron-Stein inequality. Fix n, G and a fixed enumeration $\{e_1, \ldots, e_N\}$ of the edges in $[n] \times G$ where N is the number of edges in that graph. Consider the random variable $t_n(G) - \mathbb{E}[t_n(G)]$ as a function $f(\boldsymbol{\omega})$ of the edge weight configuration $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N) \in \mathbb{R}_+^N$ where ω_i is the weight of the edge e_i.

Let $\omega_1', \ldots, \omega_N'$ be i.i.d. copies of ω_1. For a subset S of $\{1, 2, \ldots, N\}$ define $\boldsymbol{\omega}^S \in \mathbb{R}_+^N$ as the configuration where $(\boldsymbol{\omega}^S)_i = \omega_i$ for $i \notin S$ and $(\boldsymbol{\omega}^S)_i = \omega_i'$ for $i \in S$. Recall that $[i]$ denote the set $\{1, 2, \ldots, i\}$. Clearly $\boldsymbol{\omega}^{[0]} = \boldsymbol{\omega}$.

For illustration we will prove the $p = 2$ case first which is the Efron-Stein inequality. Recall that $\mathbb{E}[f(\boldsymbol{\omega})] = 0$. We have

$$\mathbb{E}[f(\boldsymbol{\omega})^2] = \mathbb{E}[f(\boldsymbol{\omega})(f(\boldsymbol{\omega}) - f(\boldsymbol{\omega}^{[N]}))] = \sum_{i=1}^{N} \mathbb{E}[f(\boldsymbol{\omega})(f(\boldsymbol{\omega}^{[i-1]}) - f(\boldsymbol{\omega}^{[i]}))].$$

Exchanging ω_i, ω_i' one can easily see that $(\boldsymbol{\omega}^{\{i\}}, \boldsymbol{\omega}^{[i]}, \boldsymbol{\omega}^{[i-1]}) \stackrel{\mathrm{d}}{=} (\boldsymbol{\omega}, \boldsymbol{\omega}^{[i-1]}, \boldsymbol{\omega}^{[i]})$ and hence we have

$$\mathbb{E}[f(\boldsymbol{\omega})^2] = \frac{1}{2} \sum_{i=1}^{N} \mathbb{E}[(f(\boldsymbol{\omega}) - f(\boldsymbol{\omega}^{\{i\}}))(f(\boldsymbol{\omega}^{[i-1]}) - f(\boldsymbol{\omega}^{[i]}))].$$

By Cauchy-Schwarz inequality and exchangeability of ω_i, ω_i' we see that

$$\mathbb{E}[f(\boldsymbol{\omega})^2] \leq \sum_{i=1}^{N} \mathbb{E}[(f(\boldsymbol{\omega}) - f(\boldsymbol{\omega}^{\{i\}}))^2 \mathbb{1}\{\omega_i' > \omega_i\}].$$

Now note that $\omega_i' > \omega_i$ and $f(\boldsymbol{\omega}) \neq f(\boldsymbol{\omega}^{\{i\}})$ implies that the i-th edge e_i is essential for the configuration $\boldsymbol{\omega}$ and moreover, $0 < f(\boldsymbol{\omega}^{\{i\}}) - f(\boldsymbol{\omega}) \leq \omega_i' - \omega_i \leq \omega_i'$. Also ω_i' is independent of $\boldsymbol{\omega}$. Thus we have

$$\mathbb{E}[f(\boldsymbol{\omega})^2] \leq \sum_{i=1}^{N} \mathbb{E}[(\omega_i')^2 \mathbb{1}\{e_i \text{ is essential for } \boldsymbol{\omega}\}] = \mathbb{E}[\omega_i^2]\, \mathbb{E}[L_n]$$

where L_n is the number of essential edges for the configuration $\boldsymbol{\omega}$.

Let $g(\cdot)$ be the function $g(x) = x|x|^{p-2}$. Using similar decomposition as was done for $p = 2$ case we have

$$\mathbb{E}[|f(\boldsymbol{\omega})|^p] = \frac{1}{2} \sum_{i=1}^{N} \mathbb{E}[(f(\boldsymbol{\omega}) - f(\boldsymbol{\omega}^{\{i\}}))(g(\boldsymbol{\omega}^{[i-1]}) - g(\boldsymbol{\omega}^{[i]}))].$$

Now Lemma 3.5.2 and symmetry of ω_i and ω_i' imply that

$$\begin{aligned}\mathbb{E}[|f(\boldsymbol{\omega})|^p] \leq a_p \sum_{i=1}^{N} \mathbb{E}\Big[&|f(\boldsymbol{\omega}) - f(\boldsymbol{\omega}^{\{i\}})||f(\boldsymbol{\omega}^{[i-1]}) - f(\boldsymbol{\omega}^{[i]})| \\ &\cdot \left(|f(\boldsymbol{\omega}^{[i-1]})|^{p-2} + |f(\boldsymbol{\omega}^{[i]})|^{p-2}\right) \mathbb{1}\{\omega_i' > \omega_i\}\Big]\end{aligned}$$

where $a_p = \max\{1, (p-1)/2\}$. Note that $\omega_i' > \omega_i$, $f(\boldsymbol{\omega}^{\{i\}}) \neq f(\boldsymbol{\omega})$ and $f(\boldsymbol{\omega}^{[i]}) \neq f(\boldsymbol{\omega}^{[i-1]})$ imply that $0 < f(\boldsymbol{\omega}^{\{i\}}) - f(\boldsymbol{\omega}), f(\boldsymbol{\omega}^{[i]}) - f(\boldsymbol{\omega}^{[i-1]}) \leq \omega_i'$ and the edge e_i is essential for both the configurations $\boldsymbol{\omega}$ and $\boldsymbol{\omega}^{[i-1]}$. Moreover in that case we have

$$\begin{aligned}|f(\boldsymbol{\omega}^{[i]})|^{p-2} &\leq |\,|f(\boldsymbol{\omega}^{[i-1]})| + \omega_i'|^{p-2} \\ &\leq 3|f(\boldsymbol{\omega}^{[i-1]})|^{p-2} + \max\{2, (2(p-3))^{p-3}\}(\omega_i')^{p-2}.\end{aligned}$$

The last line follows easily when $p \leq 3$. For $p > 3$ the last line follows by taking $\varepsilon = e^{-1/(p-3)}$, using Jenson's inequality $(a+b)^{p-2} \leq \varepsilon^{3-p}x^{p-2} + (1-\varepsilon)^{3-p}y^{p-2}$ and $(1-\varepsilon)^{-1} \leq \max\{2, 2(p-3)\}$. Thus

$$\begin{aligned}\mathbb{E}[|f(\boldsymbol{\omega})|^p] \leq \sum_{i=1}^{N} \mathbb{E}\big[&(\omega_i')^2 \mathbb{1}\{e_i \text{ is essential for } \boldsymbol{\omega}^{[i-1]}\} \\ &\cdot \left(4a_p|f(\boldsymbol{\omega}^{[i-1]})|^{p-2} + b_p(\omega_i')^{p-2}\right)\big]\end{aligned}$$

where $b_p = a_p \max\{2, (2(p-3))^{p-3}\}$. Simplifying we have

$$\begin{aligned}\mathbb{E}[|f(\boldsymbol{\omega})|^p] &\leq \sum_{i=1}^{N} \mathbb{E}\left[(\omega_i')^2 \mathbb{1}\{e_i \text{ is essential for } \boldsymbol{\omega}\} \left(4a_p|f(\boldsymbol{\omega})|^{p-2} + b_p(\omega_i')^{p-2}\right)\right] \\ &= 4a_p\, \mathbb{E}[(\omega_i')^2]\, \mathbb{E}[L_n|f(\boldsymbol{\omega})|^{p-2}] + b_p\, \mathbb{E}[(\omega_i')^p]\, \mathbb{E}[L_n]\end{aligned}$$

where L_n is the number of essential edges in the configuration $\boldsymbol{\omega}$. Let

$$y = \mathbb{E}[|f(\boldsymbol{\omega})|^p]^{(p-2)/p}.$$

Using Hölder's inequality we have

$$\begin{aligned}y^{p/(p-2)} &= \mathbb{E}[|f(\boldsymbol{\omega})|^p] \\ &\leq 4a_p\, \mathbb{E}[\omega^2]\, \mathbb{E}[L_n^{p/2}]^{2/p}\, \mathbb{E}[|f(\boldsymbol{\omega})|^p]^{(p-2)/p} + b_p\, \mathbb{E}[\omega^p]\, \mathbb{E}[L_n] \\ &= 4a_p\, \mathbb{E}[L_n^{p/2}]^{2/p}\, \mathbb{E}[\omega^2] y + b_p\, \mathbb{E}[L_n]\, \mathbb{E}[\omega^p].\end{aligned}$$

Now Lemma 3.5.3 with $\beta = p/(p-2)$ gives that

$$\mathbb{E}[|f(\boldsymbol{\omega})|^p]^{2/p} = y^{\beta-1} \leq 4a_p\, \mathbb{E}[L_n^{p/2}]^{2/p}\, \mathbb{E}[\omega^2] + (b_p\, \mathbb{E}[L_n]\, \mathbb{E}[\omega^p])^{2/p}$$

or

$$\mathbb{E}[|f(\boldsymbol{\omega})|^p] \leq 2^{p/2-1}(2a_p)^{p/2}\, \mathbb{E}[L_n^{p/2}]\, \mathbb{E}[\omega^2]^{p/2} + 2^{p/2-1}b_p\, \mathbb{E}[L_n]\, \mathbb{E}[\omega^p].$$

Note that $2a_p \leq p$ and $b_p \leq 2^{p-1}p^{p-2}$. Hence simplifying we finally conclude that

$$\mathbb{E}[|f(\boldsymbol{\omega})|^p] \leq (2p)^{p/2}\, \mathbb{E}[L_n^{p/2}]\, \mathbb{E}[\omega^2]^{p/2} + 2^{p/2}(2p)^{p-2}\, \mathbb{E}[L_n]\, \mathbb{E}[\omega^p].$$

Now we are done. $\square$

It is easy to see that $L_n(G)$ is smaller than the length of any length minimizing path. In fact the random variable $L_n(G)$ grows linearly with n. The following well-known result due to Kesten [67] will be useful to get an upper bound on the length of a weight minimizing path.

Corollary 3.5.5 (Proposition 5.8 in Kesten [67])**.** *If $F(0) < p_c(d)$ then there exist constants $0 < a, b, c < \infty$ depending on d and F only, such that the probability that there exists a selfavoiding path $\mathcal{P}$ from the origin which contains at least n many edges but has $\omega(\mathcal{P}) < cn$ is smaller than ae^{-bn}.*

Combining Lemma 3.5.4 and Lemma 3.5.5 we have the proof of Proposition 3.5.1.

Proof of Proposition 3.5.1. Note that $G_n = G$ for all n clearly implies that $L_n(G) \leq 3nk$ where $k = k(G)$ is the number of edges in G. This completes the proof for the case where the constants depend on G.

Let π_n be the minimum number of edges in a weight minimizing path for $t_n(G_n)$. To complete the proof it is enough to show the following: if G_n's are subgraphs of $\mathbb{Z}^{d-1}$ and $F(0) < p_c(d)$ we have $\mathbb{E}[\pi_n^{p/2}] \leq cn^{p/2}$ for some constant c depending only on d, p and F. We follow the idea from [68]. We have

$$\begin{aligned}\mathbb{P}(\pi_n > tn) \leq \mathbb{P}(t_n(G_n) > ctn) &+ \mathbb{P}(\text{there exists a self avoiding path } \mathcal{P} \\ &\text{starting from 0 of at least } tn \text{ edges but with } \omega(\mathcal{P}) < ctn).\end{aligned}$$

Now using Lemma 3.5.5 we see that the second probability decays like ae^{-btn}. And the first probability is bounded by $\mathbb{P}(S_n > ctn)$ where S_n is the weight of the straight line path joining $(0, o)$ to (n, o). Clearly S_n is sum of n many i.i.d. random variables. Thus we have

$$\begin{aligned}\mathbb{E}[\pi_n^{p/2}] &= \int_0^\infty \frac{n^{p/2}p}{2} t^{p/2-1}\, \mathbb{P}(\pi_n > tn)\, dt \\ &\leq \int_0^\infty \frac{n^{p/2}p}{2} t^{p/2-1}\, \mathbb{P}(S_n > ctn)\, dt + \int_0^\infty \frac{n^{p/2}p}{2} t^{p/2-1} a e^{-btn}\, dt \\ &= c^{-p/2}\, \mathbb{E}[S_n^{p/2}] + \frac{ap}{2b^{p/2}} \mathrm{Gamma}(p/2) \leq c_1 n^{p/2}\end{aligned}$$

where the constant c_1 depends on d, p and F. The result for $a_n(G)$ and $T_n(G)$ follow by Lemma 3.3.1 that

$$\mathbb{E}[|t_n(G) - a_n(G)|^p] \leq \mathbb{E}[|t_n(G) - T_n(G)|^p] \leq \mathbb{E}[(2D\omega)^p]$$

for all n, G when $\mathbb{E}[\omega^p] < \infty$ for some $p \geq 2$ with $\omega \sim F$ and D is the diameter of G.
□

Proof of the first technical Lemma 3.5.2. For $x, y \in \mathbb{R}/\{0\}, x \neq y$, let $z = x/y$. Then we have

$$\frac{x\,|x|^{p-2} - y\,|y|^{p-2}}{(x-y)(|x|^{p-2} + |y|^{p-2})} = \frac{z\,|z|^{p-2} - 1}{(z-1)(|z|^{p-2} + 1)}.$$

Now, the lemma follows from the fact that

$$c_p := \sup_{z\in\mathbb{R}} \left| \frac{z\,|z|^{p-2} - 1}{(z-1)(|z|^{p-2} + 1)} \right| \leq \max\{1, (p-1)/2\}.$$

To prove this note that, by $p > 2$ we have

$$\sup_{z\geq 0} \frac{z^{p-1} + 1}{(z+1)(z^{p-2} + 1)} \leq 1$$

and

$$\sup_{z\geq 0} \frac{z^{p-1} - 1}{(z-1)(z^{p-2} + 1)} = \left(1 - \sup_{x\geq 0} \frac{\sinh \frac{p-3}{p-1} x}{\sinh x}\right)^{-1} = \begin{cases} \left(1 - \frac{p-3}{p-1}\right)^{-1} & \text{if } p > 3, \\ (1-0)^{-1} & \text{if } p \leq 3 \end{cases}$$

and the line can be written succinctly as $\max\{1, (p-1)/2\}$. □

Proof of the second technical Lemma 3.5.3. Define $f(a, b) := (b + a^{1-1/\beta})^{1/(\beta-1)}$ and $g(a, b) := \sup\{y \geq 0 : y^\beta \leq a + by\}$. Without loss of generality assume $b > 0$. Then it is easy to see that

$$g(a, b) = b^{1/(\beta-1)} g(ab^{-\beta/(\beta-1)}, 1) \text{ and } f(a, b) = b^{1/(\beta-1)} f(ab^{-\beta/(\beta-1)}, 1).$$

So again w.l.g. we can assume that $b = 1$. Clearly $f(a, 1) \geq 1, g(a, 1) \geq 1$.

Let $F : [1, \infty) \to \mathbb{R}$ be the strictly increasing function $F(x) := x^\beta - x$. Note that $F(g(a, 1)) = a$. Now $y > f(a, 1)$ implies that $y^\beta - y = F(y) > F(f(a, 1)) = f(a, 1)(f(a, 1)^{\beta-1} - 1) \geq a^{1/\beta}(1 + a^{(\beta-1)/\beta} - 1) = a$. Hence the upper bound is proved. □

3.6 Exponential edge weights

Here we will give a different proof for the variance bounds in Proposition 3.2.3 when the edge weights are exponentially distributed with mean one (without loss of generality). The proof is based on a property of Gaussian distribution. Note that if X, Y are i.i.d. standard normal then $(X^2 + Y^2)/2$ has Exp(1) distribution.

Let $\phi_N(\boldsymbol{z}) := (2\pi)^{-N/2} \exp(-||\boldsymbol{z}||^2/2)$, $\boldsymbol{z} \in \mathbb{R}^N$ be the density of the N-dimensional standard Gaussian vector. For $\boldsymbol{k} = (k_1, k_2, \ldots, k_N) \in \mathbb{Z}_+^N$, define the $\boldsymbol{k}$-th multivariate Hermite polynomial

$$H_{\boldsymbol{k}}(\boldsymbol{x}) := \phi_N(\boldsymbol{x})^{-1} \left(-\frac{\partial}{\partial x_1}\right)^{k_1} \left(-\frac{\partial}{\partial x_2}\right)^{k_2} \cdots \left(-\frac{\partial}{\partial x_N}\right)^{k_N} \phi_N(\boldsymbol{x})$$

and $\boldsymbol{k}! = k_1! k_2! \cdots k_N!$. Then the following is a well known result in Gaussian process.

Theorem 3.6.1. *The collection of polynomials $\{H_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{Z}_+^N\}$ gives an orthogonal basis in the Hilbert space of functions $L^2(\mathbb{R}^N, \phi_N(\boldsymbol{x})d\boldsymbol{x})$ with inner product*

$$\langle f, h\rangle := \int_{\mathbb{R}^N} f(\boldsymbol{x})h(\boldsymbol{x})\phi_N(\boldsymbol{x})\, d\boldsymbol{x} = \mathbb{E}[f(\boldsymbol{Z})h(\boldsymbol{Z})]$$

where $\boldsymbol{Z}$ is N-dimensional standard Gaussian random vector. Moreover we have

$$\langle H_{\boldsymbol{k}}, H_{\boldsymbol{m}}\rangle = \begin{cases} \boldsymbol{k}! & \text{if } \boldsymbol{k} = \boldsymbol{m} \\ 0 & \text{otherwise} \end{cases}$$

for all $\boldsymbol{k}, \mathbf{m} \in \mathbb{Z}_+^N$.

Using Theorem 3.6.1, for any L^2 function f we have

$$\mathbb{E}[f^2(\boldsymbol{Z})] = \sum_{\boldsymbol{k}\in\mathbb{Z}_+^N} \frac{1}{\boldsymbol{k}!}\langle H_{\boldsymbol{k}}, f\rangle^2 \text{ or } \mathrm{Var}(f(\boldsymbol{Z})) == \sum_{\boldsymbol{k}\in\mathbb{Z}_+^N\setminus\{\boldsymbol{0}\}} \frac{1}{\boldsymbol{k}!}\langle H_{\boldsymbol{k}}, f\rangle^2$$

Now if f is once differentiable, using the fact that $H_{2\boldsymbol{e}_i}(\boldsymbol{z}) = z_i^2 - 1$ for $1 \le i \le N$ and $\mathbb{E}[Z_i f(\boldsymbol{Z})] = \mathbb{E}[\frac{\partial f}{\partial z_i}(\boldsymbol{Z})]$ we have

$$\langle H_{2\boldsymbol{e}_i}, f\rangle = \mathbb{E}[(Z_i^2 - 1)f(\boldsymbol{Z})] = \mathbb{E}\left[Z_i \frac{\partial f}{\partial z_i}(\boldsymbol{Z})\right].$$

In particular we have

$$\mathrm{Var}(f(\mathbf{Z})) \ge \frac{1}{2}\sum_{i=1}^N \left(\mathbb{E}\left[Z_i\frac{\partial f}{\partial z_i}(\mathbf{Z})\right]\right)^2. \tag{3.10}$$

Taking limits it is easy to see that the bound (3.10) holds for any absolutely continuous function f. Now in our case N is the number of edges in $[n] \times G$ and the function $T = T_{0,n}(G)$ of the edge weights $\boldsymbol{\omega} = (\omega_i)_{i=1}^N$ is absolutely continuous w.r.t. $(x_i, y_i)_{i=1}^N$ where $\omega_i = (x_i^2 + y_i^2)/2$. Let x_i, y_i be i.i.d. standard Gaussian. Then ω_i's are i.i.d. Exponentially distributed with mean one. Moreover by continuity, the minimum weight path $\mathcal{P}_*(\boldsymbol{\omega})$ is unique a.s. and we have

$$\frac{\partial T(\boldsymbol{\omega})}{\partial x_i} = x_i \mathbb{1}\{e_i \in \mathcal{P}_*(\boldsymbol{\omega})\},\ \frac{\partial T(\boldsymbol{\omega})}{\partial y_i} = y_i \mathbb{1}\{e_i \in \mathcal{P}_*(\boldsymbol{\omega})\} \text{ a.s.}$$

for $i = 1, 2, \ldots, N$. Hence from (3.10) we have

$$\begin{aligned}\mathrm{Var}(T(\boldsymbol{\omega})) &\ge \frac{1}{2}\sum_{i=1}^N \left[\left(\mathbb{E}\left[x_i^2\mathbb{1}\{e_i \in \mathcal{P}_*(\boldsymbol{\omega})\}\right]\right)^2 + \left(\mathbb{E}\left[y_i^2\mathbb{1}\{e_i \in \mathcal{P}_*(\boldsymbol{\omega})\}\right]\right)^2\right] \\ &\ge \frac{1}{4N}\left(\sum_{i=1}^N \mathbb{E}\left[(x_i^2 + y_i^2)\mathbb{1}\{e_i \in \mathcal{P}_*(\boldsymbol{\omega})\}\right]\right)^2 \\ &= \frac{1}{N}\left(\sum_{i=1}^N \mathbb{E}\left[\omega_i\mathbb{1}\{e_i \in \mathcal{P}_*(\boldsymbol{\omega})\}\right]\right)^2 = \frac{1}{N}\left(\mathbb{E}[T(\boldsymbol{\omega})]\right)^2.\end{aligned}$$

Now note that $N \le 3nk$ where k is the number of edges in G and $\mathbb{E}[T(\boldsymbol{\omega})] \ge cn$ for some $c > 0$ by Lemma 3.3.2. Thus we have the required lower bound. The upper bound follows easily from the Poincaré inequality.

In fact, using hypercontractivity and a argument similar to the one used in [10] one can prove that

$$\mathrm{Var}(a_n(h_n)) \le \frac{Cn}{1 + \log h_n}.$$

This implies that for $h_n \to \infty$, $a_n(h_n)$ is noise sensitive and so any constant level Fourier mass is negligible compared to the variance. Since our lower bound is based on the second level Fourier mass, the lower bound is not tight when $h_n \to \infty$.

3.7 Proof of Theorem 3.2.1

The proof of Theorem 3.2.1 will be given in several steps. First we will show that it is enough to prove the CLT for $T_n(G_n)$ after proper centering and scaling. Then we will prove that $T_n(G_n)$ is "approximately" a sum of i.i.d. random variables each having distribution $T_l(G_n)$ and an error term where l depends on n. Finally, using successive breaking of $T_l(G_n)$ into i.i.d. sums (the 'renormalization steps') and controlling the error in each step, we will complete the proof. Recall that the notations $a_n = O(b_n)$ and $a_n = o(b_n)$, respectively, mean that $a_n \le Cb_n$ for all $n \ge 1$ for some constant $C < \infty$ and $a_n/b_n \to 0$ as $n \to \infty$. Throughout the proof c will denote a constant that depends only on q, F and whose value may change from line to line.

3.7.1 Reduction to $T_n(G_n)$

Let us first recall the setting. We have a sequence of nondecreasing graphs G_n with G_n having diameter d_n and k_n edges. We also have $k_n = O(d_n^\theta)$ for some fixed $\theta \ge 1$. Define

$$\mu_n(G) := \mathbb{E}[T_n(G)] \text{ and } \sigma_n^2(G) := \mathrm{Var}(T_n(G))$$

for any integer $n \ge 1$ and any finite connected graph G.

Now from Lemma 3.3.1 we have

$$\mathbb{E}[|a_n(G_n) - T_n(G_n)|^p] \le 2^p d_n^p\, \mathbb{E}[\omega^p]$$

for all n when $\mathbb{E}[\omega^p] < \infty$ for a typical edge weight ω. Moreover, from Proposition 3.2.3 we have $\sigma_n^2(G_n) \ge cnk_n^{-1}$ for all n for some absolute constant $c > 0$ when $d_n = o(n)$. Thus when $d_n^2 = o(nk_n^{-1})$ (which is satisfied if $d_n = o(n^{1/(2+\theta)})$), we have

$$\frac{T_n(G_n) - \mu_n(G_n)}{\sigma_n(G_n)} - \frac{a_n(G_n) - \mathbb{E}[a_n(G_n)]}{\mathrm{Var}(a_n(G_n))^{1/2}} \xrightarrow{L^2} 0.$$

Hence it is enough to prove CLT for $(T_n(G_n)-\mu_n(G_n))/\sigma_n(G_n)$ when $d_n = o(n^{1/(2+\theta)})$. From now on we will assume that

$$d_n = o(n^{\alpha}) \text{ with } \alpha < 1/(2+\theta) \text{ fixed.}$$

3.7.2 Approximation as an i.i.d. sum

In Lemma 3.7.1 we will prove a relation between side-to-side first-passage times in large and small cylinders and this will be crucial to the whole analysis. Fix an integer n and a finite connected graph G. Let $n = ml + r$ with $0 \le r < l$ where $l \ge 1$ is an integer.

We divide the cylinder graph $[n] \times G$ horizontally into m equal-sized smaller cylinder graphs $R_1, \dots, R_m$ with $R_i = [(i-1)l, il] \times G, i = 1, 2, \dots, m$ each having width l and a residual graph $R_{m+1} = [ml, n] \times G$. Let

$$X_i = T_{(i-1)l,il}(G) \tag{3.11}$$

be the side-to-side first-passage time for the product graph R_i for $i = 1, 2, \dots, m$ (see Definition 3.2). We also define $X_{m+1} = T_{ml,n}(G)$ for the residual graph R_{m+1}. Clearly $X_{m+1} = 0$ if $r = 0$. Note that X_i's depend on n and G, but we will suppress n, G for readability. We have the following relation. This is a generalization of Lemma 3.3.1.

Corollary 3.7.1. *Let n, G be fixed. Let X_i be as defined in* (3.11)*. Then the random variable*

$$Y := T_n(G) - (X_1 + X_2 + \cdots + X_{m+1})$$

is nonnegative and is stochastically dominated by S_{mD} where S_{mD} is sum of mD many i.i.d. random variables each having distribution F and D is the diameter of G. Moreover, $X_1, \dots, X_m$ are i.i.d. having the same distribution as $T_l(G)$, X_{m+1} has the distribution of $T_r(G)$ and X_{m+1} is independent of $X_1, \dots, X_m$.

Proof. First of all, it is easy to see that X_i depends only on the weights for the edge set $\{e : e \text{ is an edge in } [(i-1)l, il] \times G\} \setminus \{e \mid e \text{ is an edge in } \{(i-1)l\} \times G \text{ or } \{il\} \times G\}$. Thus, $X_1, \dots, X_m$'s are i.i.d. having the same distribution as $T_l(G)$.

Now choose a minimal weight path $\mathcal{P}^*$ joining the left boundary $\{0\} \times G$ to the right boundary $\{n\} \times G$ (if there are more than one path one can use some deterministic rule to break the tie). The path $\mathcal{P}^*$ hits all the boundaries $\{il\} \times G$ at some vertex for $i = 0, 1, \dots, m$. Let $u_i, v_i, i = 0, 1, \dots, m$ be the vertices in G such that for each i, $\mathcal{P}^*$ hits $\{il\} \times G$ for the last time at the vertex (il, u_i) and after that it hits the boundary $\{(i+1)l\} \times G$ at the vertex $((i+1)l, v_i)$ for the first time (take $(m+1)l$ to be n). Clearly if $\mathcal{P}^*$ hits $\{il\} \times G$ only at a single vertex then $u_i = v_{i-1}$. Now the part of $\mathcal{P}^*$ between the vertices (il, u_i) and $((i+1)l, v_i)$ is a path in $[il, (i+1)l] \times G$ and hence has weight more than X_i. But all these parts are disjoint. Hence we have $T_n(G) = \omega(\mathcal{P}^*) \ge \sum_{i=1}^{m+1} X_i$.

Now to prove upper bound for Y, let $\mathcal{P}_i^*$ be a minimal weight path joining the left boundary $\{il\} \times G$ to the right boundary $\{(i+1)l\} \times G$ and achieving the weight X_i. Suppose $\mathcal{P}_i^*$ hits $\{il\} \times G$ at (il, w_i) and hits $\{(i+1)l\} \times G$ at $((i+1)l, z_i)$ for $i = 0, 1, \dots, m$. Let $\mathcal{P}_i$ be a minimal length path in $\{il\} \times G$ joining (il, z_{i-1}) to (il, w_i) for $i = 1, 2, \dots, m$. Consider the concatenated path $\mathcal{P}_0^*, \mathcal{P}_1, \mathcal{P}_1^*, \mathcal{P}_2, \dots, \mathcal{P}_m^*$ joining $(0, w_0)$ to (n, z_{m+1}). By minimality of weight we have

$$T_n(G) \le \sum_{i=1}^{m} (X_i + \omega(\mathcal{P}_i)) + X_{m+1}.$$

Thus we have $Y = T_n(G) - \sum_{i=1}^{m+1} X_i \le \sum_{i=1}^{m} \omega(\mathcal{P}_i)$. Clearly $\sum_{i=1}^{m} \omega(\mathcal{P}_i)$ is a sum of $\sum_{i=1}^{m} d(z_{i-1}, w_i)$ many i.i.d. random variables each having distribution F where $d(\cdot,\cdot)$ is the graph distance in G_n. But we have $\sum_{i=1}^{m} d(z_{i-1}, w_i) \le mD$ by definition of the diameter. Now F is supported on $\mathbb{R}^+$. Thus we are done. □

An obvious corollary of Lemma 3.7.1 is the following.

Corollary 3.7.2. *For any integer m, l, r and connected graph G we have*

$$|\mu_{ml+r}(G) - (m\mu_l(G) + \mu_r(G))| \le mD\mu$$

and

$$\left|\sigma_{ml+r}(G) - (m\sigma_l^2(G) + \sigma_r^2(G))^{1/2}\right| \le mD(\mu^2 + \sigma^2)^{1/2}$$

where D is the diameter of G.

Proof. Taking expectation of Y in Lemma 3.7.1 with $n = ml + r$ we have $\mathbb{E}[Y] = \mu_n(G) - m\mu_l(G) - \mu_r(G)$ and $0 \le \mathbb{E}[Y] \le mD\mu$.

Moreover, we have

$$\begin{aligned}
&\left|\mathrm{Var}(T_n(G))^{1/2} - \mathrm{Var}(T_n(G) - Y)^{1/2}\right| \\
&= |\|T_n(G) - \mathbb{E}[T_n(G)]\|_2 - \|T_n(G) - Y - \mathbb{E}[T_n(G) - Y]\|_2| \\
&\le \|Y - \mathbb{E}[Y]\|_2 \le (\mathbb{E}[Y^2])^{1/2} \le mD(\mu^2 + \sigma^2)^{1/2}.
\end{aligned}$$

Now the result follows since $T_n(G) - Y = \sum_{i=1}^{m+1} X_i$ and X_i's are independent of each other. □

3.7.3 Lyapounov condition

From here onwards, we return to using n in subscripts and superscripts. From Lemma 3.7.1 and Corollary 3.7.2 clearly we have

$$\begin{aligned}
&\mathbb{E}\,|T_n(G_n) - \mu_n(G_n) - (X_1^{(n)} + X_2^{(n)} + \cdots X_m^{(n)} - m\mu_l(G_n))| \\
&\le \mathbb{E}\,|T_n(G_n) - (X_1^{(n)} + X_2^{(n)} + \cdots X_{m+1}^{(n)})| + md_n\mu + \mathbb{E}\,|X_{m+1}^{(n)} - \mu_r(G_n)| \\
&\le 2md_n\mu + \sigma_r(G_n) \qquad (3.12)
\end{aligned}$$

where $X_i^{(n)}, i = 1, 2, \ldots, m$ are defined as in (3.11) and $n = ml + r$. We will take

$$l = \max\{\lfloor n^\beta \rfloor, 1\} \text{ for some fixed } \beta \in (2/(2+\theta), 1) \text{ and } m = \lfloor n/l \rfloor.$$

Then we have $d_n^2 = o(l)$ and all the lower and upper bounds on moments are valid for $T_l(G_n)$. The dependence of m, l on n is kept implicit. Note that $0 \le r < l$. Moreover, writing $l - r$ in place of l and 1 in place of m, we get from Corollary 3.7.2 that

$$\sigma_r(G_n) \le \sigma_l(G_n) + (\mu^2 + \sigma^2)^{1/2} d_n. \tag{3.13}$$

Thus from (3.12) we have

$$\begin{aligned} &\mathbb{E}\left| \frac{T_n(G_n) - \mu_n(G_n)}{\sqrt{m}\sigma_l(G_n)} - \frac{\sum_{i=1}^m (X_i^{(n)} - \mu_l(G_n))}{\sqrt{m}\sigma_l(G_n)} \right| \\ &\le \frac{2md_n\mu + \sigma_r(G_n)}{\sqrt{m}\sigma_l(G_n)} \le \frac{1}{\sqrt{m}} + 3(\sigma^2 + \mu^2)^{1/2} \frac{\sqrt{m} d_n}{\sigma_l(G_n)}. \end{aligned} \tag{3.14}$$

Recall that we have $l \sim n^\beta$ for some $\beta < 1$ and thus $m \sim n^{1-\beta}$. From the lower bound for the variance in Proposition 3.2.3 (as $d_n = o(l)$) we have

$$\frac{md_n^2}{\sigma_l^2(G_n)} \le \frac{cm^2 d_n^2 k_n}{n},$$

where c is some absolute constant. By our assumption on m, d_n and k_n we have $m^2 d_n^2 k_n = o(n)$ when $\alpha \le (2\beta - 1)/(2+\theta)$ which is true for some $\beta < 1$ as $\alpha < 1/(2+\theta)$. Hence $(T_n(G_n) - \mu_n(G_n))/\sqrt{m}\sigma_l(G_n)$ has the same asymptotic limit as

$$\frac{\sum_{i=1}^m X_i^{(n)} - m\mu_l(G_n)}{\sqrt{m}\sigma_l(G_n)} \tag{3.15}$$

as $n \to \infty$ when

$$\alpha \le \frac{2\beta - 1}{2+\theta} \text{ for some } \beta \in \left(\frac{2}{2+\theta}, 1 \right). \tag{3.16}$$

Now $X_i^{(n)}, i = 1, 2, \ldots, m$ are i.i.d. random variables with finite second moment, hence by the CLT for triangular arrays it is expected that (3.15) has standard Gaussian distribution asymptotically. However we cannot expect CLT for all values of β.

Let $s_n^2 := m\sigma_l^2(G_n)$ be the variance of $\sum_{i=1}^m X_i^{(n)}$. To use Lindeberg condition for triangular arrays of i.i.d. random variables we need to show that

$$\frac{m}{s_n^2} \mathbb{E}[\tilde{T}_l^2 \mathbb{1}\{|\tilde{T}_l| \ge \varepsilon s_n\}] \to 0 \text{ as } n \to \infty$$

for every $\varepsilon > 0$ where $\tilde{T}_l = T_l(G_n) - \mu_l(G_n)$. However, any bound using the relation $T_l(G_n) \le S_l$ where S_l is the weight of the straight line path joining $(0, o)$ and (l, o), gives rise to the condition $\theta\alpha \le 1 - 2\beta$. The last condition is contradictory to (3.16). The difficulty arises from the fact that the lower and upper bounds for the variances are not tight.

Still we can prove a CLT by using estimates for the moments of $\tilde{T}_l(G_n)$ from Proposition 3.5.1 and using a blocking technique which is reminiscent of the renormalization group method. Note that Lindeberg condition follows from the Lyapounov condition

$$\frac{m}{s_n^p}\,\mathbb{E}[|T_l(G_n) - \mu_l(G_n)|^p] \to 0 \text{ as } n \to \infty \text{ for some } p > 2 \tag{3.17}$$

and thus it is enough to prove (3.17) for some $\beta \in (2/(2+\theta), 1)$ where $l = \max\{\lfloor n^\beta \rfloor, 1\}$, $m = \lfloor n/l \rfloor, s_n^2 = m\sigma_l^2(G_n)$. We also need to satisfy (3.16) to complete the proof of Theorem 3.2.1.

3.7.4 A technical estimate

We need the following technical estimate for the next "renormalization" step. The lemma gives an upper bound on the moment of sums of i.i.d. random variables. It is known as Rosenthal's inequality (see [95]) in the literature.

Corollary 3.7.3. *Let $Y_i, i = 1, 2, \ldots, m$ be i.i.d. random variables with mean zero and $\mathbb{E}[Y_i^p] < \infty$ for some $p \ge 2$. Then we have*

$$\mathbb{E}[|Y_1 + Y_2 + \cdots + Y_m|^p] \le A_p(m\,\mathbb{E}[Y^p] + (m\,\mathbb{E}[Y^2])^{p/2}) \tag{3.18}$$

where A_p is a constant depending only on p.

Proof. For simplicity we present the proof when $p = 2q$ is an even integer. Let $Y \stackrel{\mathrm{d}}{=} Y_1$ and $S_m = Y_1 + \cdots + Y_m$. For $\boldsymbol{a} = (a_1, a_2, \ldots, a_{2q}) \in \mathbb{Z}_+^{2q}$, we will denote $\sum_{i=1}^{2q} a_i$ by $|\boldsymbol{a}|$ and $\sum_{i=1}^{2q} i a_i$ by $z(\boldsymbol{a})$. To estimate $\mathbb{E}[S_m^{2q}]$, we will use the following decomposition which is an easy exercise in combinatorics. We have

$$\mathbb{E}[S_m^{2q}] = \sum_{\boldsymbol{a} \in \mathbb{Z}_+^{2q} : z(\boldsymbol{a}) = 2q} \frac{(2q)!}{\prod_{i=1}^{2q} i!^{a_i} a_i!} (m)_{|\boldsymbol{a}|} \prod_{i=1}^{2q} \mathbb{E}[Y^i]^{a_i}$$

where $(m)_k := m!/(m-k)! \le m^k$. Note that here we used the fact that Y_i's are i.i.d.. Since $\mathbb{E}[Y] = 0$ we can and we will assume that $a_1 = 0$. Thus using Hölder's inequality we have

$$\begin{aligned}
\mathbb{E}[S_m^{2q}] &\leq \sum_{z(\boldsymbol{a})=2q} \frac{(2q)!}{\prod_{i=2}^{2q} i!^{a_i} a_i!} (m)_{|\boldsymbol{a}|} \prod_{i=2}^{2q} \mathbb{E}[|Y|^i]^{a_i} \\
&\leq \sum_{z(\boldsymbol{a})=2q} \frac{(2q)!}{\prod_{i=2}^{2q} i!^{a_i} a_i!} m^{|\boldsymbol{a}|} \prod_{i=2}^{2q} \mathbb{E}[Y^2]^{\frac{a_i(q-i/2)}{q-1}} \, \mathbb{E}[Y^{2q}]^{\frac{a_i(i/2-1)}{q-1}} \\
&\leq \sum_{z(\boldsymbol{a})=2q} \frac{(2q)!}{\prod_{i=2}^{2q} i!^{a_i} a_i!} (m^q \, \mathbb{E}[Y^2]^q)^{\frac{|\boldsymbol{a}|-1}{q-1}} (m \, \mathbb{E}[Y^{2q}])^{\frac{q-|\boldsymbol{a}|}{q-1}}.
\end{aligned}$$

Note that $2|\boldsymbol{a}| \leq z(\boldsymbol{a}) = 2q$ as $a_1 = 0$. Now using the fact that $x^\alpha y^{1-\alpha} \leq \alpha x + (1-\alpha) y$ for all $x, y \geq 0, \alpha \in [0, 1]$ we finally have

$$\mathbb{E}[S_m^{2q}] \leq A_q (m \, \mathbb{E}[Y^{2q}] + m^q \, \mathbb{E}[Y^2]^q) \tag{3.19}$$

where

$$A_q := \sum_{z(\boldsymbol{a})=2q} \frac{(2q)!}{\prod_{i=2}^{2q} i!^{a_i} a_i!}$$

is a constant depending only on q. □

3.7.5 Renormalization Step

Now we are ready to start our proof of the Lyapounov condition. For simplicity we will write $T_l(G) - \mu_l(G)$ imply as $\tilde{T}_l(G)$. Recall that

$$\nu = \lim_{n \to \infty} \frac{\mathbb{E}[T_n(G_n)]}{n}.$$

Corollary 3.7.4. *Suppose that $\nu > 0$ and $\mathbb{E}[\omega^p] < \infty$ for some $p > 2$ where ω is a typical edge weight. Suppose either $G_n = G$ for all n or G_n's are subgraphs of $\mathbb{Z}^{d-1}$. Let $l = \max\{\lfloor n^\beta \rfloor, 1\}$, $d_n = o(n^\alpha)$ with $2\alpha < \beta$ and $k_n = O(d_n^\theta)$ for fixed $\theta \geq 1$. Suppose that there exist $t \geq 1$ real numbers $\beta_i, i = 1, 2 \ldots, t$ such that $2\alpha < \beta_t < \beta_{t-1} < \cdots < \beta_1 = \beta$ and we have*

$$\alpha \leq \frac{1 - 2(\beta_i - \beta_{i+1}) - (1 - \beta_i)/q}{2 + \theta} \text{ for all } i = 1, 2, \ldots, t-1,$$

$$\text{and } \alpha \leq \frac{q-1}{q} \cdot \frac{1 - \beta_t}{\theta}$$

where $q = p/2$. Then we have

$$\frac{\sum_{i=1}^m X_i^{(n)} - m\mu_l(G_n)}{\sqrt{m}\sigma_l(G_n)} \xrightarrow{\text{w}} N(0, 1)$$

as $n \to \infty$ where $X_i^{(n)}$'s are i.i.d. with $X_i^{(n)} \stackrel{\text{d}}{=} T_l(G_n)$.

Proof. Since $X_i^{(n)}, i = 1, 2, \ldots, m$ are i.i.d. with mean $\mu_l(G_n)$ and variance $\sigma_l^2(G_n)$ and $\mathbb{E}[\omega^p] < \infty$ for some $p > 2$, we can use the Lyapounov condition to prove the central limit theorem. We need to show that

$$\frac{m}{s_n^p}\,\mathbb{E}[|\tilde{T}_l(G_n)|^p] \to 0 \text{ as } n \to \infty$$

where $s_n^2 = m\sigma_l^2(G_n)$. By the variance lower bound from Proposition 3.2.3 we have

$$s_n^2 \geq c_1 \frac{ml}{k_n} \geq c_2 \frac{n}{k_n} \tag{3.20}$$

for some constants $c_i > 0$ where k_n is the number of edges in G_n. Using the moment bound from Proposition 3.5.1 and lower bound on s_n^2 (note that $d_n^2 = o(l)$) we have

$$\frac{m}{s_n^p}\,\mathbb{E}[|\tilde{T}_l(G_n)|^p] \leq \frac{c_p m l^{p/2}}{(n/k_n)^{p/2}} \leq \frac{c_p m l^{p/2} k_n^{p/2}}{(ml)^{p/2}} = \frac{c_p k_n^{p/2}}{m^{(p-2)/2}}.$$

Thus when $k_n = o(m^{1-2/p})$ or equivalently $\theta\alpha \leq (1 - 2/p)(1 - \beta)$, we see that the right hand side converges to zero and we have a central limit theorem. This proves the assertion of the theorem when $t = 1$.

Let us now look into the bounds more carefully. The random variable $T_l(G_n)$ itself behaves like a sum of i.i.d. random variables each having distribution $T_{l'}(G_n)$ for $l' < l$. We will use this fact to improve the required growth rate of k_n. Let $q = p/2$ and assume that there exist $t \geq 2$ real numbers $\beta_i, i = 1, 2 \ldots, t$ such that $2\alpha < \beta_t < \beta_{t-1} < \cdots < \beta_1 = \beta$ and we have

$$\begin{aligned} \alpha &\leq \frac{1 - 2(\beta_i - \beta_{i+1}) - (1 - \beta_i)/q}{2 + \theta} \text{ for all } i = 1, 2, \ldots, t-1 \\ \text{and } \alpha &\leq \frac{q-1}{q} \cdot \frac{1 - \beta_t}{\theta}. \end{aligned} \tag{3.21}$$

From now on we will write l_1, m_1 and β_1 instead of l, m and β respectively. Recall that we have $l_1 = \max\{\lfloor n^{\beta_1} \rfloor, 1\}$ and $d_n = o(n^\alpha)$. We will take

$$l_i = \max\{\lfloor n^{\beta_i} \rfloor, 1\}, m_i = \lfloor l_{i-1}/l_i \rfloor \text{ for } i = 2, \ldots, t.$$

The idea is as follows. First we will break the cylinder graph $[0, l_1] \times G_n$ into m_2 many equal sized graphs each of which looks like $[0, l_2] \times G_n$. Then we will break each of the new graphs again into m_3 many equal sized graphs each of which looks like $[0, l_3] \times G_n$ and so on. We will stop after t steps. Our goal is to break the error term into smaller and smaller quantities and show that the original quantity is "small" when each of the final quantities are "small". Throughout the proof $q, t, \theta, \alpha, \beta_i, i = 1, 2, \ldots, t$ are fixed.

For simplicity, first we will assume that

$$l_1 = m_2 m_3 \cdots m_t l_t.$$

Under this assumption we have $m_i l_i = l_{i-1}$ for all $i = 2 \ldots, t$. Otherwise one has to look at the error terms which can be easily bounded using essentially the same idea and are considered in (3.29).

First Step. Let us start with the first splitting. We break the rectangular graph $[0, l_1] \times G_n$ into m_2 many equal sized graphs $[(i-1)l_2, il_2] \times G_n$ for $i = 1, 2, \ldots, m_2$. Recall that we have $l_1 = m_2 l_2$.

Let $S_{m_2} = \sum_{i=1}^{m_2} X_i$ where $X_i = T_{(i-1)l_2, il_2}(G_n) - \mu_{l_2}(G_n)$. Recall that X_i's are i.i.d. having the same distribution as $\tilde{T}_{l_2}(G_n)$ where $\tilde{T}_l(G_n) = T_l(G_n) - \mu_l(G_n)$. Let $\varepsilon_1 = \varepsilon_1(n) := m_1/s_n^{2q}$. We need to show the Lyapounov condition:

$$\varepsilon_1 \, \mathbb{E}[\tilde{T}_{l_1}(G_n)^{2q}] = o(1). \tag{3.22}$$

From Lemma 3.7.1 we have

$$\mathbb{E}[|\tilde{T}_{l_1}(G_n) - S_{m_2}|^{2q}] \leq c(m_2 d_n)^{2q} \, \mathbb{E}[\omega^{2q}]$$

for some constant $c > 0$. Moreover, Lemma 3.7.3 implies that

$$\mathbb{E}[S_{m_2}^{2q}] \leq A_q(m_2^q \, \mathbb{E}[\tilde{T}_{l_2}(G_n)^2]^q + m_2 \, \mathbb{E}[\tilde{T}_{l_2}(G_n)^{2q}]).$$

Thus we have

$$\begin{aligned} &\varepsilon_1 \, \mathbb{E}[\tilde{T}_{l_1}(G_n)^{2q}] \\ &\qquad \leq c(\varepsilon_1 (m_2 d_n)^{2q} + \varepsilon_1 m_2^q \, \mathbb{E}[\tilde{T}_{l_2}(G_n)^2]^q + \varepsilon_1 m_2 \, \mathbb{E}[\tilde{T}_{l_2}(G_n)^{2q}]). \end{aligned}$$

Hence we need to show that

$$\varepsilon_1 (m_2 d_n)^{2q} = o(1), \tag{3.23}$$

$$\varepsilon_1 m_2^q \sigma_{l_2}^{2q}(G_n) = o(1) \tag{3.24}$$

$$\text{and } \varepsilon_1 m_2 \, \mathbb{E}[\tilde{T}_{l_2}(G_n)^{2q}] = o(1) \tag{3.25}$$

Using the variance lower bound (3.20) we have

$$\varepsilon_1 (m_2 d_n)^{2q} \leq c \frac{m_1 (m_2)^{2q} (d_n^2 k_n)^q}{n^q} \leq c \left(\frac{d_n^2 k_n}{n^{1-2(\beta_1-\beta_2)-(1-\beta_1)/q}} \right)^q.$$

Now (3.23) follows as $d_n^2 k_n = o(n^{(2+\theta)\alpha})$ and $(2+\theta)\alpha \leq 1 - 2(\beta_1 - \beta_2) - (1-\beta_1)/q$. Moreover, Corollary 3.7.2 with $l_1 = m_2 l_2$ implies that

$$(m_2 \sigma_{l_2}^2(G_n))^{1/2} \leq \sigma_{l_1}(G_n) + c m_2 d_n.$$

Thus using the definition of $\varepsilon_1 = \varepsilon_1(n)$ and the fact that $s_n^2 = m_1\sigma_{l_1}^2(G_n)$ we have

$$\varepsilon_1 m_2^q \sigma_{l_2}^{2q}(G_n) \leq c(\varepsilon_1\sigma_{l_1}^{2q}(G_n) + \varepsilon_1(m_2d_n)^{2q}) \leq c\left(m_1^{1-q} + \varepsilon_1(m_2d_n)^{2q}\right)$$

and the right hand side is $o(1)$ as $q > 1$ and by (3.23). So the only thing that remains to be proved is that

$$\varepsilon_1 m_2\,\mathbb{E}[\tilde{T}_{l_2}(G_n)^{2q}] = o(1).$$

Induction step. From the above calculations in step 1 the induction step is clear. Define

$$\varepsilon_i = \varepsilon_i(n) = \frac{m_1m_2\cdots m_i}{s_n^{2q}} \text{ for } i \geq 1.$$

Claim 1. We have $\varepsilon_i(m_{i+1}d_n)^{2q} = o(1)$ for all $i < t$.
Proof of Claim 1. Fix any i. Using definition of ε_i and the variance lower bound from (3.20) we have

$$\begin{aligned}\varepsilon_i(m_{i+1}d_n)^{2q} = \frac{m_1\cdots m_i(m_{i+1}d_n)^{2q}}{s_n^{2q}} &\leq c\frac{n^{1-\beta_i}m_{i+1}^{2q}(d_n^2k_n)^q}{n^q} \\ &= o\left(\left[\frac{n^{(2+\theta)\alpha}}{n^{1-2(\beta_i-\beta_{i+1})-(1-\beta_i)/q}}\right]^q\right).\end{aligned}$$

Now the claim follows by our assumption (3.21) that $(2+\theta)\alpha \leq 1 - 2(\beta_i - \beta_{i+1}) - (1-\beta_i)/q$ for all $i < t$.

Our next claim is the following.

Claim 2. We have $\varepsilon_i m_{i+1}^q\sigma_{l_{i+1}}^{2q}(G_n) = o(1)$ for all $i \geq 1$.
Proof of Claim 2. We will prove the claim by induction on i. We have already proved the claim for $i = 1$ in (3.24). Now suppose that the claim is true for some $i \geq 1$. Using Corollary 3.7.2 for $l_{i+1} = l_{i+2}m_{i+2}$ we see that

$$\begin{aligned}\varepsilon_{i+1}(m_{i+2}\sigma_{l_{i+2}}^2(G_n))^q &\leq c(\varepsilon_{i+1}\sigma_{l_{i+1}}^{2q}(G_n) + \varepsilon_{i+1}(m_{i+2}d_n)^{2q}) \\ &= c(\varepsilon_i m_{i+1}\sigma_{l_{i+1}}^{2q}(G_n) + \varepsilon_{i+1}(m_{i+2}d_n)^{2q}).\end{aligned}$$

Hence we have $\varepsilon_{i+1}(m_{i+2}\sigma_{l_{i+2}}^2(G_n))^q = o(1)$ by Claim 1 and the induction hypothesis as $q > 1$. This completes the proof.

Claim 3. For any $i \geq 1$, $\varepsilon_i\,\mathbb{E}[\tilde{T}_{l_i}(G_n)^{2q}] = o(1)$ if $\varepsilon_{i+1}\,\mathbb{E}[\tilde{T}_{l_{i+1}}(G_n)^{2q}] = o(1)$.
Proof of Claim 3. Assume that $\varepsilon_{i+1}\,\mathbb{E}[\tilde{T}_{l_{i+1}}(G_n)^{2q}] = o(1)$. We write $\tilde{T}_{l_i}(G_n)$ as a sum of $S_{m_{i+1}}$ and an error term of order $m_{i+1}d_n$ where $S_{m_{i+1}}$ is sum of m_{i+1} many i.i.d. random variables each having distribution $\tilde{T}_{l_{i+1}}(G_n)$. Using Lemma 3.7.3, as was done in the first step, one can easily see that $\varepsilon_i\,\mathbb{E}[\tilde{T}_{l_i}(G_n)^{2q}] = o(1)$ when

$$\varepsilon_i(m_{i+1}d_n)^{2q} = o(1), \tag{3.26}$$

$$\varepsilon_i m_{i+1}^q\sigma_{l_{i+1}}^{2q}(G_n) = o(1) \tag{3.27}$$

$$\text{and } \varepsilon_i m_{i+1}\,\mathbb{E}[\tilde{T}_{l_{i+1}}(G_n)^{2q}] = o(1). \tag{3.28}$$

Now Condition (3.26) holds by Claim 1, Condition (3.27) holds by Claim 2 and Condition (3.28) holds by the hypothesis as $\varepsilon_{i+1} = \varepsilon_i m_{i+1}$.

Hence if we stop at step t, we see that the central limit theorem holds when $\varepsilon_t \, \mathbb{E}[\tilde{T}_{l_t}(G_n)^{2q}] = o(1)$. By the upper bound for the $2q$-th moment from Proposition 3.5.1 (as $d_n^2 = o(l_t)$) we see that $\varepsilon_t \, \mathbb{E}[\tilde{T}_{l_t}(G_n)^{2q}] \leq \varepsilon_t l_t^q$ and by the lower bound for the variance from (3.20) we have

$$\varepsilon_t l_t^q \leq \frac{cm_1 m_2 \cdots m_t l_t^q k_n^q}{n^q} = \frac{ck_n^q}{(m_1 m_2 \cdots m_t)^{q-1}} = o\left(\frac{n^{q\theta\alpha}}{n^{(q-1)(1-\beta_t)}}\right).$$

The last condition also holds by our assumption (3.21) that $q\theta\alpha \leq (q-1)(1-\beta_t)$. Thus we are done when $l_1 = m_2 m_3 \cdots m_t l_t$.

Now, in general we have $l_{i-1} = m_i l_i + r_i$ for $i = 2, \ldots, t$ where $0 \leq r_i < l_i$ for all i. Using the same proof used in the case when all $r_i = 0$, one can easily see from Claim 3, that we need to prove the extra conditions that

$$\varepsilon_i \, \mathbb{E}[\tilde{T}_{r_i}(G_n)^{2q}] = o(1) \text{ for all } i = 2, 3, \ldots, t. \tag{3.29}$$

Fix $i \in \{2, 3, \ldots, t\}$. If $r_i \leq l_t$ then we are done since $\varepsilon_i \leq \varepsilon_t$ and by Proposition 3.5.1 we have $\mathbb{E}[\tilde{T}_{r_i}(G_n)^{2q}] \leq c(d_n^{2q} + l_t^q) \leq c_1 l_t^q$. The last inequality follows since $2\alpha < \beta_t$. Now suppose that $l_{j+1} \leq r_i < l_j$ for some $j \geq i$. Since we have $\varepsilon_i \leq \varepsilon_j$ for $j \geq i$ working with r_i instead of l_j and using the same inductive analysis used before we have the required result (3.29). □

3.7.6 Choosing the sequence

To complete the proof of Theorem 3.2.1 we need to choose an appropriate sequence $(\beta_1, \ldots, \beta_t)$ in (3.21) which will be provided by Lemma 3.7.5. Note that

$$\frac{1 - 2(\beta_0 - \beta_1) - (1 - \beta_0)/q}{2 + \theta} = \frac{2\beta_1 - 1}{2 + \theta}$$

for $\beta_0 = 1$ and we have noted earlier in (3.16) that

$$\frac{a_n(G_n) - \mathbb{E}[a_n(G_n)]}{\mathrm{Var}(a_n(G_n))^{1/2}} \text{ has the same asymptotic limit as } \frac{\sum_{i=1}^m X_i^{(n)} - m\mu_l(G_n)}{\sqrt{m}\sigma_l(G_n)}$$

when $d_n = o(n^\alpha)$ and $\alpha \leq (2\beta_1 - 1)/(2 + \theta)$.

Corollary 3.7.5. *Let $\beta_1, \beta_2, \ldots, \beta_t$ be t real numbers satisfying the system of linear equations*

$$\frac{1 - 2(\beta_i - \beta_{i+1}) - (1 - \beta_i)/q}{2 + \theta} = \frac{q-1}{q} \cdot \frac{1 - \beta_t}{\theta} \tag{3.30}$$

for all $i = 0, 1, 2, \ldots, t-1$ *where* $\beta_0 = 1$. *Then we have*

$$\beta_i := 1 - \frac{q\theta(1-r^i)}{\theta + (q-1)(2+\theta)(1-r^t)} \tag{3.31}$$

for all $i = 1, 2, \ldots, t$ *where* $r = 1 - 1/(2q)$.

Proof. Define $x_i = 1 - \beta_i$ for $i = 0, 1, \ldots, t$. Clearly $x_0 = 0$. Also define the constants

$$c = \frac{q-1}{q} \cdot \frac{2+\theta}{\theta} \text{ and } r = 1 - \frac{1}{2q}.$$

Then the system of equations (3.30) can be written in terms of x_i's as

$$\begin{aligned} 1 - 2x_{i+1} + 2rx_i &= cx_t \text{ for all } i = 0, 1, \ldots, t-1 \\ \text{or } x_{i+1} - rx_i &= (1 - cx_t)/2 \text{ for all } i = 0, 1, \ldots, t-1. \end{aligned} \tag{3.32}$$

Multiplying the i-th equation by r^{-i-1} and summing over $i = 0, 1, \ldots, t-1$ we have

$$r^{-t}x_t = qr^{-t}(cx_t - 1)(r^t - 1) \text{ or } x_t = \frac{q(1-r^t)}{1 + qc(1-r^t)}.$$

Now solving (3.32) recursively starting from $i = t-1, t-2, \ldots, 0$ we have

$$x_i = \frac{q(1-r^i)}{1 + qc(1-r^t)} \text{ for all } i = 1, 2, \ldots, t.$$

Simplifying and reverting back to β_i we finally get

$$x_i = 1 - \frac{q\theta(1-r^i)}{\theta + (q-1)(2+\theta)(1-r^t)}$$

for all $i = 1, 2, \ldots, t$. □

3.7.7 Completing the proof

Now we connect all the loose ends to complete the proof of Theorem 3.2.1.

Recall that the number of edges satisfies $k_n = O(d_n^\theta)$ and moreover we have $d_n = o(n^\alpha)$ for some $\alpha < 1$. We also have $l \sim n^{\beta_1}, m \sim n^{1-\beta_1}$ for some $\beta_1 \in (\alpha, 1)$. We have proved in (3.16) that the CLT will follow if we can find some $\beta_1 \in (\alpha, 1)$ such that $\alpha \le (2\beta_1 - 1)/(2+\theta)$ and

$$\frac{\sum_{i=1}^m X_i - m\mu_l(G_n)}{\sqrt{m}\sigma_l(G_n)} \xrightarrow{\text{w}} N(0,1) \tag{3.33}$$

as $n \to \infty$ where X_i's are i.i.d. having distribution $T_l(G_n)$. Note that $(2\beta-1)/(2+\theta) < \beta/2$ for all $\beta > 0$.

To prove (3.33) we will use the condition in Lemma 3.7.4. Assume that $\mathbb{E}[\omega^p] < \infty$ for some real number $p > 2$. Let $q = p/2$. From Lemma 3.7.4 we see that CLT will hold in (3.33) if there exist $t \geq 1$ real numbers $\beta_i, i = 1, 2 \ldots, t$ such that $2\alpha < \beta_t < \beta_{t-1} < \cdots < \beta_1 < \beta_0 = 1$ and

$$\alpha \leq \frac{q-1}{q} \cdot \frac{1-\beta_t}{\theta} \text{ and } \alpha \leq \frac{1 - 2(\beta_i - \beta_{i+1}) - (1-\beta_i)/q}{2+\theta} \tag{3.34}$$

for all $i = 0, 1, \ldots, t-1$. For $i = 0$ the equation reduces to $\alpha \leq (2\beta_1 - 1)/(2+\theta)$.

Now fix any integer $t \geq 1$. Define $r = 1 - 1/2q$. For $i = 1, \ldots, t$, define

$$\beta_i := 1 - \frac{q\theta(1-r^i)}{\theta + (q-1)(2+\theta)(1-r^t)}. \tag{3.35}$$

As usual we will assume that $\beta_0 = 1$. Clearly $\beta_t < \beta_{t-1} < \cdots < \beta_1 < \beta_0$. The sequence $(\beta_1, \ldots, \beta_t)$ is the unique solution to the system of equations given by equality in the right hand side of (3.34) (see Lemma 3.7.5). In fact we have

$$\frac{q-1}{q} \cdot \frac{1-\beta_t}{\theta} = \frac{(q-1)(1-r^t)}{\theta + (q-1)(2+\theta)(1-r^t)}$$

and

$$\frac{1 - 2(\beta_i - \beta_{i+1}) - (1-\beta_i)/q}{2+\theta} = \frac{(q-1)(1-r^t)}{\theta + (q-1)(2+\theta)(1-r^t)}$$

for any $i = 0, 1, \ldots, t-1$. Now note that

$$\frac{2(q-1)(1-r^t)}{\theta + (q-1)(2+\theta)(1-r^t)} < 1 - \frac{q\theta(1-r^t)}{\theta + (q-1)(2+\theta)(1-r^t)} = \beta_t$$

as $\theta + (q-1)(2+\theta)(1-r^t) - (2(q-1) + q\theta)(1-r^t) = \theta r^t > 0$. Thus combining all the previous results we have

$$\frac{a_n(G_n) - \mathbb{E}[a_n(G_n)]}{\sqrt{m}\sigma_l(G_n)} \xrightarrow{\text{w}} N(0,1) \text{ as } n \to \infty$$

when

$$\alpha \leq \frac{(q-1)(1-r^t)}{\theta + (q-1)(2+\theta)(1-r^t)}$$

for some integer $t \geq 1$. Since $r = 1 - 1/(2q) < 1$, letting $t \to \infty$ we get the CLT when

$$\alpha < \frac{q-1}{\theta + (q-1)(2+\theta)} = \frac{1}{2 + \theta + 2\theta/(p-2)}.$$

Thus we are done. ◻

3.8 The case of fixed graph G

By the arguments given in Section 3.2, we have a Gaussian central limit theorem for $a_n(G)$ and $T_n(G)$ as $n \to \infty$ after proper scaling when G is a fixed graph. Proposition 3.2.2 says that

$$\nu(G) := \lim_{n\to\infty} \frac{\mathbb{E}[T_n(G)]}{n}$$

exists and is positive. Moreover, Proposition 3.2.3 gives that

$$0 < c_1 \leq \frac{\mathrm{Var}(T_n(G))}{n} \leq c_2$$

for all n for some constants $c_1, c_2 > 0$ depending on G. The next lemma says that in fact we can say more. Assume that $v(G)$ is the number of vertices in G, $k(G)$ is the number of edges in G and $D = D(G)$ is the diameter of G.

Corollary 3.8.1. *Let G be a finite connected graph. Then we have*

$$|\,\mathbb{E}[T_n(G)] - n\nu(G)| \leq \mu D \text{ for all } n$$

and the limit

$$\sigma^2(G) := \lim_{n\to\infty} \frac{\sigma_n^2(G)}{n}$$

exists and is positive.

Proof. Let $\tilde{\mu}_n = \mu_n/n$ and $\tilde{\sigma}_n^2 = \sigma_n^2/n$. Using the proof given in corollary 3.7.2 we have

$$|n\tilde{\mu}_n - (ml\tilde{\mu}_l + r\tilde{\mu}_r)| \leq m\mu D \text{ and } \left|(n\tilde{\sigma}_n^2)^{1/2} - (ml\tilde{\sigma}_l^2 + r\tilde{\sigma}_r^2)^{1/2}\right| \leq mbD \qquad (3.36)$$

for all $n = ml + r$ with $0 \leq r < l$ where $b = (\mu^2 + \sigma^2)^{1/2}$. Thus for any m, k we have $|\tilde{\mu}_{mk} - \tilde{\mu}_m| \leq \mu D/m$. Reversing the roles of m and k, and combining, we see that for any m, k, we have

$$|\tilde{\mu}_m - \tilde{\mu}_k| \leq \mu D/k + \mu D/m.$$

Taking limits as $k \to \infty$ we have, for any m,

$$|\tilde{\mu}_m - \lim_{n\to\infty} \tilde{\mu}_n| \leq \mu D/m.$$

For the variance, we take $n = 2l$ in equation (3.36) to have

$$|\tilde{\sigma}_{2l} - \tilde{\sigma}_l| \leq bD(2/l)^{1/2}.$$

Hence, it follows that $\tilde{\sigma}_{2^k}$ is Cauchy and $\lim_{k\to\infty} \tilde{\sigma}_{2^k}$ exists.

Now take any $l \geq 1$. There exists a unique positive integer $k = k(l)$ such that $2l^{3/2} \leq 2^k < 4l^{3/2}$ ($k(l) = 1 + \lceil \log_2 l^{3/2} \rceil$). Suppose $2^k = ml + r$ where $0 \leq r < l$. Clearly $\sqrt{l} \leq m \leq 4\sqrt{l}$. Now from (3.36) we have,

$$\left|(2^k \tilde{\sigma}_{2^k}^2)^{1/2} - (ml\tilde{\sigma}_l^2 + r\tilde{\sigma}_r^2)^{1/2}\right| \leq mbD.$$

Dividing by $2^{k/2}$ on both sides, we get

$$\left|\tilde{\sigma}_{2^k} - \left(\tilde{\sigma}_l^2 + \frac{r(\tilde{\sigma}_r^2 - \tilde{\sigma}_l^2)}{ml + r}\right)^{1/2}\right| \leq \frac{mbD}{\sqrt{ml + r}} \leq 2bDl^{-1/4}.$$

Note that k, m, r are functions of l in the above expression. Among these, $m(l) \geq l^{1/2}$ and $r(l) < l$. Taking $l \to \infty$, and using the fact that the sequence $\{\tilde{\sigma}_n^2\}_{n\geq 1}$ is uniformly bounded (see Proposition 3.5.1), we get that $\lim_{m\to\infty} \tilde{\sigma}_m$ exists and equals $\lim_{k\to\infty} \tilde{\sigma}_{2^k}$. Positivity of the limit follows from the variance lower bound given in Proposition 3.2.3. □

Note that, if we consider the point-to-point cylinder first-passage time $t_n(G)$ in $[0, n] \times G$, the same results given in Lemma 3.8.1 hold for $\mathbb{E}[t_n(G)]$ and $\mathrm{Var}(t_n(G))$.

Now we consider the process $X(m)$ where $X(m) = t_m(G) - m\nu(G)$ for $m \in \{0, 1, \ldots\}$ and $X_n(t) = X_m + (t - m)(X_{m+1} - X_m)$ for $t \in (m, m+1)$. Note that when G is the trivial graph consisting of a single vertex, $X(n)$ corresponds to random walk with linear interpolation and by Donsker's theorem $\{(n\sigma^2)^{-1/2} X(nt)\}_{t\geq 0}$ converges to Brownian motion. The next lemma says that for general G we also have the same behavior. We assume that $\mathbb{E}[\omega^p] < \infty$ for some $p > 2$ where $\omega \sim F$.

Corollary 3.8.2. *The scaled process $\{(n\sigma^2(G))^{-1/2} X(nt)\}_{t\geq 0}$ converges in distribution to standard Brownian motion as $n \to \infty$.*

Proof. Consider the continuous process X' defined as $X'(n) := T_n(G) - n\nu(G)$ for $n \in \{0, 1, \ldots\}$ and extended by linear interpolation. By Lemma 3.3.1 it is enough to prove Brownian convergence for $\{Y_n(t) := (n\sigma^2(G))^{-1/2} X'(nt) : 0 \leq t \leq T\}$ for any fixed $T > 0$. To prove the result it suffices to show that the finite dimensional distributions of $Y_n(t)$ converge weakly to those of B_t and that $\{Y_n\}$ is tight.

First of all note that for any $s > 0$, we have

$$\begin{aligned} |Y_n(s) - (n\sigma^2(G))^{-1/2} X(\lfloor ns \rfloor)| &\leq (n\sigma^2(G))^{-1/2} |X'(1 + \lfloor ns \rfloor) - X'(\lfloor ns \rfloor)| \\ &\leq (n\sigma^2(G))^{-1/2} (Z + \nu(G)) \xrightarrow{\mathrm{P}} 0 \end{aligned}$$

where Z is the maximum of all the edge weights connecting $\{\lfloor ns \rfloor\} \times G$ to $\{1 + \lfloor ns \rfloor\} \times G$, which has the distribution of maximum of $v(G)$ many i.i.d. random variables each having distribution F. Thus it is enough to prove finite dimensional distributional convergence of the process $\{W_n(t) := (n\sigma^2(G))^{-1/2} X'(\lfloor nt \rfloor)\}_{t\geq 0}$. For a fixed $t > 0$, using Theorem 3.2.1 we have $W_n(t) \xrightarrow{\mathrm{w}} N(0, t)$ since $\lfloor nt \rfloor / n \to t$.

For $0 = t_0 < t_1 < t_2 < \cdots < t_l < \infty$, define $V_i = T_{\lfloor nt_{i-1}\rfloor,\lfloor nt_i\rfloor}(G) - (\lfloor nt_i\rfloor - \lfloor nt_{i-1}\rfloor)\nu(G)$ for $i = 1, 2, \ldots, l$. Clearly V_i's are independent for all i. Moreover using Lemma 3.7.1 we have

$$\mathbb{E}[|W_n(t_i) - W_n(t_{i-1}) - (n\sigma^2(G))^{-1/2}V_i|] \to 0$$

as $n \to \infty$ for all i. Thus by independence and by CLT for $(n\sigma^2(G))^{-1/2}V_i$, we have

$$(W_n(t_i) - W_n(t_{i-1}))_{i=1}^l \xrightarrow{\text{w}} (B_{t_i} - B_{t_{i-1}})_{i=1}^l \text{ as } n \to \infty.$$

To prove tightness for $\{Y_n(\cdot)\}$, first of all note that certainly $\{Y_n(0)\}$ is tight as $Y_n(0) \equiv 0$. Also it is enough to prove tightness for $\{W_n(\cdot)\}$. We will prove tightness via the following lemma.

Corollary 3.8.3 (Billingsley [13], page 87-91)**.** *The sequence $\{W_n\}$ is tight if there exist constants $C \geq 0$ and $\lambda > 1/2$ such that for all $0 \leq t_1 < t_2 < t_3$ and for all n, we have*

$$\mathbb{E}[|W_n(t_2) - W_n(t_1)|^{2\lambda}|W_n(t_3) - W_n(t_2)|^{2\lambda}] \leq C|t_2 - t_1|^{\lambda}|t_3 - t_2|^{\lambda}.$$

Using the Cauchy-Schwarz inequality and Proposition 3.5.1, it is easy to that Lemma 3.8.3 holds with $\lambda = p/4$. Thus we are done. □

3.9 Numerical results

In this section we report some numerical simulation results which support Conjecture 3.1.3 and 3.1.6. We consider two-dimensional rectangles $[n] \times [-h_n, h_n]$ with $h_n = n^{\alpha}$ for h_n ranging from 30 to 60 and α from the sequence $2/3, 1/2, 2/5$ and $1/3$. For the edge weight distribution we take Bernoulli(p) for different values of p. For each configuration we simulate 1000 observations for $a_n(h_n)$ to estimate the variance.

We assume that there are two constants $\beta, \gamma > 0$ depending only on the distribution of edge weights such that

$$\text{Var}(a_n(h_n)) \approx \beta n h_n^{-\gamma}$$

for $h_n \leq n^{2/3}$. Note that we have the rigorous result that $\gamma \in [0, 1]$ if it exists. However it is not clear how to define the approximation properly. Our conjecture is that γ exists in some "appropriate" sense (for example the ratio of the logarithms of both sides are bounded) and satisfies the following:

Conjecture 3.9.1. *For two-dimension, we have*

$$\gamma = 1/2$$

when $h_n = \Theta(n^{\alpha})$ and $\alpha \leq 2/3$.

To estimate the numbers β, γ we use the simple linear regression model

$$\log \mathrm{Var}(a_n(h_n)) = \log \beta + \log n - \gamma \log(h_n) + \text{ Gaussian error}$$

and least square estimates. The results are summarized in Table 3.1 and 3.2. For each α, the first two columns show estimated values of γ and β. The third column gives the R^2-values for the linear fit. The row for a given value of p corresponds to taking Bernoulli(p) as the edge-weight distribution. In figure 3.1 the estimated values of γ are plotted against p for different values α, which shows that γ is close to 1/2 for all values of p.

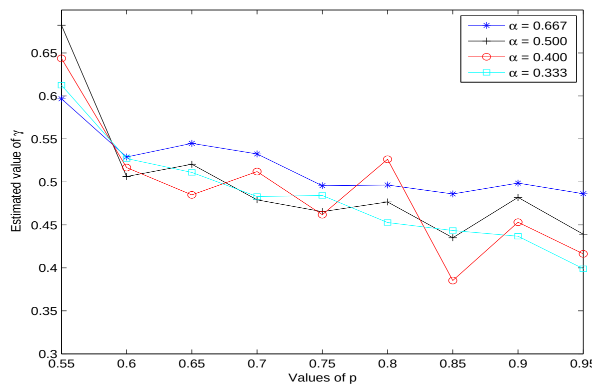

Figure 3.1: Plot of estimated value of γ vs. p for different values of α.

p	$\alpha = 2/3$			$\alpha = 1/2$		
	γ estimate	β estimate	R-squared	γ estimate	β estimate	R-squared
.55	0.59665	0.33373	0.9899	0.68224	0.38860	0.9890
.60	0.52898	0.34687	0.9936	0.50626	0.27719	0.9825
.65	0.54485	0.44715	0.9944	0.52052	0.33806	0.9902
.70	0.53255	0.48762	0.9939	0.47911	0.32135	0.9853
.75	0.49552	0.42032	0.9943	0.46539	0.31256	0.9850
.80	0.49639	0.42626	0.9913	0.47664	0.31795	0.9854
.85	0.48601	0.37961	0.9953	0.43500	0.24835	0.9897
.90	0.49857	0.35201	0.9952	0.48197	0.24066	0.9765
.95	0.48624	0.23308	0.9923	0.43909	0.13365	0.9887

Table 3.1: Simulation results for $\alpha = 2/3$ and 1/2.

Figure 3.2 shows QQ plots based on the above simulation data for $a_n(h_n)$ for $n = h_n^2 = 55$ against an appropriately fitted normal distribution, supporting the conjecture

p	$\alpha = 2/5$			$\alpha = 1/3$		
	γ estimate	β estimate	R-squared	γ estimate	β estimate	R-squared
.55	0.64363	0.32603	0.9954	0.61249	0.28677	0.9965
.60	0.51667	0.28690	0.9965	0.52718	0.28627	0.9964
.65	0.48483	0.29860	0.9950	0.51104	0.31453	0.9975
.70	0.51208	0.34944	0.9962	0.48300	0.30865	0.9972
.75	0.46182	0.30366	0.9968	0.48419	0.31650	0.9954
.80	0.52628	0.34583	0.9953	0.45275	0.27474	0.9964
.85	0.38531	0.20287	0.9967	0.44346	0.24726	0.9953
.90	0.45303	0.20414	0.9955	0.43682	0.18604	0.9970
.95	0.41627	0.11022	0.9949	0.39896	0.10018	0.9967

Table 3.2: Simulation results for $\alpha = 2/5$ and $1/3$.

of asymptotic normality. We will investigate asymptotic normality of $a_n(h_n)$ for $h_n \ll n^{2/3}$ in future research.

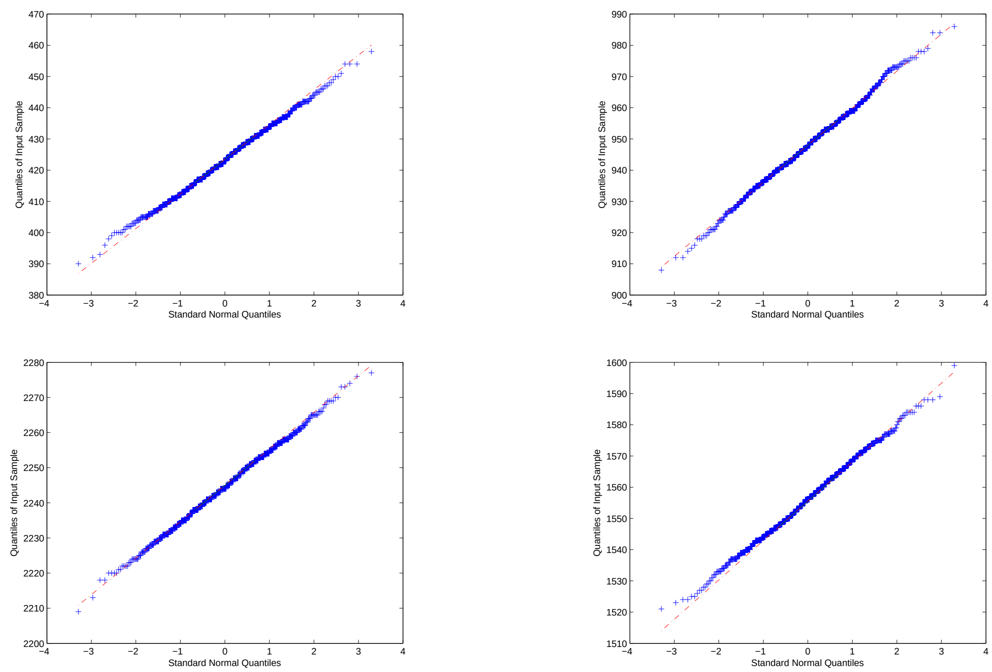

Figure 3.2: QQ plots based on simulation data for $a_n(n^{1/2})$ for $n = 3000$ against an appropriately fitted normal distribution for Bernoulli(p) edge weights, $p = 0.6, 0.7, 0.8, 0.9$ in clockwise direction starting from top left.

Chapter 4

Spectra of random linear combinations of projection matrices

4.1 Introduction

For a symmetric $n \times n$ matrix A, let $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ denote its eigenvalues arranged in nonincreasing order. The spectral measure Λ_A of A is defined as the empirical measure of its eigenvalues which puts mass $1/n$ to each of its eigenvalues, *i.e.,*

$$\Lambda_A = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}$$

where δ_x is the dirac measure at x. In particular when the matrix A is random we have a random spectral measure corresponding to A.

In his seminal paper [111] Wigner proved that the spectral measure for a large class of random matrices converges to the semi-circular law, as the dimension grows to infinity. Much work has since been done on various aspects of eigenvalues for different ensembles of large real symmetric or complex hermitian random matrices. In many cases, the random matrix has a simple linear structure. Moreover, there is also a big literature on the asymptotic spectral measure of random matrices coming from Haar measure on classical groups (e.g., orthogonal, unitary, simplectic group). Some of the results are surveyed in [53, 85]. The results are not only of interest to statisticians or to physicists but also to mathematicians, because of it its relation to combinatorics, geometry and algebra.

Many new results have been proved in the last few years for understanding liming spectral distribution of large random matrices having more complicated algebraic structure. In [23] the authors considered the spectra of large random Hankel, Markov and Toeplitz matrices, which was motivated by an open problem in [5] (see also [55]). We briefly describe their result for Markov matrices.

Let $\{X_{ij} : j \geq i \geq 1\}$ be an infinite upper triangular array of i.i.d. random variables and define $X_{ji} = X_{ij}$ for $j > i \geq 1$. Let M_n be the random $n \times n$ symmetric matrix given by

$$(M_n)_{ij} = \begin{cases} X_{ij} & \text{if } i \neq j \\ -\sum_{l \neq i} X_{il} & \text{if } i = j. \end{cases}$$

Note that each of the rows of $\mathbf{M}_n$ has zero sum. Their result says the following:

Theorem 4.1.1 (Theorem 1.3 in [23]). *Let $\{X_{ij} : j \geq i \geq 1\}$ be a collection of i.i.d. random variables with $E(X_{12}) = 0$ and $\mathrm{Var}(X_{12}) = 1$. With probability 1, $\Lambda_{n^{-1/2}M_n}$ converges weakly as $n \to \infty$ to the free convolution γ_M of the semicircle and standard normal measures. This measure γ_M is a nonrandom symmetric probability measure with smooth bounded density and does not depend on the distribution of X_{12} and has unbounded support.*

Note that $M_n = X_n - D_n$ where $(X_n)_{ij} = X_{ij}$ and D_n is the diagonal matrix with i-th diagonal entry given by $\sum_{j=1}^n X_{ij}$. By Wigner's result $\Lambda_{n^{-1/2}X_n}$ converges to the semicircular law and each $n^{-1/2}(D_n)_{ii}$ converges to i.i.d. standard Gaussian random variable. Thus the result is intuitively clear, however it is hard to prove because of the strong dependence between X_n and D_n. Now note that, M_n can also be written as follows

$$M_n = \sum_{1 \leq i < j \leq n} X_{ij}(I - P_{ij})$$

where P_{ij} is the permutation matrix (which is also a projection matrix) corresponding to the permutation (i, j) that interchanges i and j.

Recently, in [43] the author considered linear combinations of matrices defined via representations and coxeter generators of the symmetric group. The result is described in Theorem 4.3.4. Here the matrices involved are all self-adjoint unitary matrices. However it is easy to check that a matrix U is self-adjoint and unitary iff $(I - U)/2$ is a projection matrix.

In many other cases also the random matrix, when it has a linear structure, can be written as a linear function $\sum_\alpha X_\alpha M_\alpha^{(n)}$ of i.i.d. random variables $\{X_\alpha\}$ where $M_\alpha^{(n)}$'s are deterministic matrices. For example Wigner matrices can be written as $\sum_{i \leq j} X_{ij} M_{ij}^{(n)}$ where $M_{ij}^{(n)}$ is the $n \times n$ matrix with 1 at the (i, j) and (j, i)-th position and zero everywhere else.

In this chapter, we are interested in the case when $M_\alpha^{(n)}$'s are affine transformation of projection matrices, that is, $M_\alpha^{(n)}$ can be written as a linear combination of a projection matrix and the identity matrix. Note that, the Markov random matrix example in [23] and the result in [43] fall in this category.

Motivated by the result in [43] we will investigate sufficient conditions under which the limiting measure exists and we also identify the limit.

Let $X_1, X_2, \dots$ be a sequence of i.i.d. real random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}(X_1^2) = 1$. Given n, suppose we have $k = k(n)$ many $n \times n$ symmetric matrices

$$M_1^{(n)}, M_2^{(n)}, \dots, M_k^{(n)}.$$

Without loss of generality (by appropriate scaling) we will always assume that the spectral radius of $M_i^{(n)}$ is one for all $i = 1, 2, \dots, k$.

Now consider the random matrix

$$A_n = \sum_{i=1}^{k} a_i^{(n)} X_i M_i^{(n)}$$

where $\{a_i^{(n)}\}$ is a sequence of nonnegative real numbers. Let $\Lambda_n = \Lambda_{A_n}$ be the spectral measure of A_n. Clearly Λ_n is a random measure on $\mathbb{R}$.

For a $n \times n$ symmetric matrix A define its trace norm by

$$\|A\|_{\text{tr}} := \frac{1}{n} \sum_{i=1}^{n} |\lambda_i(A)|. \tag{4.1}$$

where $\lambda_i(A)$'s are the eigenvalues of A (counting multiplicity). Our first result says that the limit of Λ_n, if exists, is universal under minimal assumptions.

Lemma 4.1.2. *Suppose $\Lambda_{A_n} \xrightarrow{\text{w}} \Lambda_\infty$ (w.r.t. the topology of weak convergence of measures) in distribution as $n \to \infty$ where $A_n = \sum_{i=1}^{k(n)} a_i^{(n)} Z_i M_i^{(n)}$, Z_i's are i.i.d. standard normal random variables and Λ_∞ is a random probability measure on $\mathbb{R}$. Assume that*

$$\max_{1 \le i \le k(n)} \left| a_i^{(n)} \right| \to 0 \text{ and } \|\mathbf{a}\|^2 \max_{1 \le i \le k(n)} \left\| M_i^{(n)} \right\|_{tr} \text{ is uniformly bounded}$$

as $n \to \infty$. Then $\Lambda_{B_n} \xrightarrow{\text{w}} \Lambda_\infty$ in distribution as $n \to \infty$ where $B_n = \sum_{i=1}^{k(n)} a_i^{(n)} X_i M_i^{(n)}$ and X_i's are independent uniformly square integrable random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}(X_1^2) = 1$.

For simplicity, we assume that all $M_i^{(n)}$'s are projection matrices, that is

$$\left(M_i^{(n)}\right)^2 = M_i^{(n)}.$$

We also assume that $\text{Tr}(M_{i_1}^{(n)} M_{i_2}^{(n)} \cdots M_{i_k}^{(n)})$ depends only on k, n when $i_1, i_2, \dots, i_k$'s are distinct integers such that $M_{i_1}^{(n)}, M_{i_2}^{(n)}, \dots, M_{i_k}^{(n)}$ commute with each other. Define $\mu_k(n)$ as the above number $\text{Tr}(M_{i_1}^{(n)} M_{i_2}^{(n)} \cdots M_{i_k}^{(n)})$. Our main result says the following.

Theorem 4.1.3. *Assume that*

$$\sum_{i=1}^{k(n)} \left(a_i^{(n)}\right)^2 = 1$$

and

$$\max_{1\le i\le k(n)} |a_i^{(n)}| \to 0, \sum_{(i,j)\in E_n} \left(a_i^{(n)} a_j^{(n)}\right)^2 \to 0 \text{ as } n\to\infty$$

where $E_n := \{(i,j) : M_i^{(n)}$ *does not commute with* $M_j^{(n)}\}$*. Also assume that*

$$\frac{\mu_1(n)}{n} \to \theta \text{ and } \frac{\mu_2(n)}{n} \to \theta^2 \text{ as } n\to\infty$$

for some real number $\theta \in [0,1]$*. Let* Λ_n *be the empirical spectral distribution of*

$$A_n = \sum_{i=1}^{k(n)} a_i^{(n)} Z_i M_i^{(n)}$$

where Z_i*'s are i.i.d. standard Gaussian random variables. Then* Λ_n *converges in distribution (with respect to the topology of weak convergence of probability measures on* $\mathbb{R}$*) to a random distribution* Λ_∞ *in probability where* $\Lambda_\infty = \nu_Z, Z$ *is* $N(0,1)$ *and* ν_z *is the distribution* $N(\theta z, \theta(1-\theta))$*.*

In Section 4.2 we state the main results. We will provide several examples from representation theory of symmetric groups in Section 4.3. Section 4.4 gives generalization of results from Section 4.2. Finally in Section 4.5 we will prove the results.

4.2 Results

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean zero and variance one. Suppose we have a sequence of k many $d\times d$ symmetric matrices $\mathbf{M} = (M_1, M_2, \ldots, M_k)$ and a sequence of real numbers $\mathbf{a} = (a_1, a_2, \ldots, a_k)$. We consider the random matrix

$$A = \sum_{i=1}^{k} a_i X_i M_i.$$

Since A is symmetric all of its eigenvalues are real. Hence the empirical spectral measure Λ_A is a probability measure on the real line.

We will always assume that there is an underlying parameter n, such that $k, d, \mathbf{M}, \mathbf{a}$ all depend on n. We will write $k(n)$, $\mathbf{M}(n), \mathbf{a}(n)$, $M_i^{(n)}$, $a_i^{(n)}, A_n$ instead of $k, \mathbf{M}, \mathbf{a}$, M_i, a_i, A respectively when the dependence on n need to be shown explicitly. Here we will investigate the limiting behavior of Λ_{A_n} under appropriate assumptions as $n\to\infty$.

Before stating the result we need some definitions. Given a $d\times d$ matrix A we define its mean trace by

$$\overline{\mathrm{Tr}}(A) = \frac{1}{d}\sum_{i=1}^{d} A_{ii}.$$

We also denote the L^2 operator norm of A by

$$\|A\| = \sup\{\|Ax\|_2 : \|x\|_2 = 1\} \tag{4.2}$$

and its trace norm by

$$\|A\|_{\mathrm{tr}} = \overline{\mathrm{Tr}}(\sqrt{A^*A}). \tag{4.3}$$

If $\lambda_1, \lambda_2, \ldots, \lambda_d$ are the d eigenvalues (which can be complex) of A counting multiplicity, then we have $\|A\| = \max_{1\le i\le d}|\lambda_i|$ and $\|A\|_{\mathrm{tr}} = \frac{1}{d}\sum_{i=1}^d |\lambda_i|$.

By changing the a_i's if necessary, w.l.g. we may assume that $\|M_i\| \le 1$ (any uniform bound is enough) for all $1 \le i \le k$, that is all the eigenvalues of M_i are in the interval $[-1, 1]$. First we will prove that under quite general condition of the limiting spectral measure of A_n is universal w.r.t. the distribution of X when it exists. Define

$$\begin{aligned} \|\mathbf{a}\| &:= \left(a_1^2 + a_2^2 + \cdots + a_k^2\right)^{1/2}, \\ c_n = \max_{1\le i\le k} |a_i| &\text{ and } b_n = \max_{1\le i\le k} \|M_i\|_{\mathrm{tr}}. \end{aligned} \tag{4.4}$$

Recall that, $k, \mathbf{a}, a_i, M_i$ depend on n, so b_n, c_n depend on n too.

Lemma 4.2.1. *Suppose $\Lambda_{A_n} \xrightarrow{\mathrm{w}} \Lambda_\infty$ (w.r.t. the topology of weak convergence of measures) in distribution as $n \to \infty$ where $A_n = \sum_{i=1}^k a_i Z_i M_i$, Z_i's are i.i.d. standard normal random variables and Λ_∞ is a random probability measure on $\mathbb{R}$. Assume that $c_n \to 0$ and*

$$\max_{1\le i\le k} \|M_i\|, \|\mathbf{a}\|^2 b_n \text{ are uniformly bounded}$$

as $n \to \infty$. Then $\Lambda_{B_n} \xrightarrow{\mathrm{w}} \Lambda_\infty$ in distribution as $n \to \infty$ where $B_n = \sum_{i=1}^k a_i X_i M_i$ and X_i's are independent uniformly square integrable random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}(X_1^2) = 1$.

In the classical Wigner ensemble or Markov random matrix case, $k(n) \approx n^2/2, c_n = n^{-1/2}, \|\mathbf{a}\|^2 \approx n$ and $b_n \le 2/n$. In most of our later examples, we will have $\|\mathbf{a}\| = 1$ and $c_n = o(1)$ as $n \to \infty$ and so the result in Lemma 4.2.1 holds.

By Lemma 4.2.1 it is enough to prove the limits for standard Gaussian random variables. In our examples M_i will be of the form $aP + bI$ where P is a projection matrix, I is the identity matrix and a, b are real numbers with $|a + b| \le 1$.

Given a sequence $\mathbf{M} = (M_1, M_2, \ldots, M_k)$ of k matrices we define the *"interaction graph"* of $\mathbf{M}$ as follows:

Definition 4.2.2. *The graph $G := ([k], E)$ with vertex set $[k]$ and edge set $E = \{(i, j) : M_iM_j \ne M_jM_i\}$ is called the interaction graph of $\mathbf{M}$.*

We also define,

Definition 4.2.3. *The* sparseness *of* $\mathbf{M}$ *w.r.t. the sequence* $\mathbf{a}$ *is defined as*

$$N(\mathbf{a};\mathbf{M}) = \sum_{(i,j)\in E} a_i^2 a_j^2$$

where E *is the set of edges in the interaction graph of* $\mathbf{M}$.

Note that if all a_i's are equal to $k^{-1/2}$ then $N(\mathbf{a};\mathbf{M})$ is $k^{-2}|E|$ and if all elements in $\mathbf{M}$ commute with each other then G is the empty graph $([k],\varnothing)$. So $N(\mathbf{a};\mathbf{M})$ measures the size of the interaction graph $G(\mathbf{M})$ w.r.t. the weight sequence $\mathbf{a}$. When needed, to stress the dependence on n, we will write E_n, N_n instead of $E, N(\mathbf{a}(n);\mathbf{M}(n))$.

We say that condition $\mathbf{A}$ holds if

Condition A. For every integer $s,t \geq 0$ there is a real number $\mu_{s,t}$ such that

$$\Delta_{s,t}(n) := \sup|\,\overline{\mathrm{Tr}}(M_{i_1}M_{i_2}\cdots M_{i_s}M_{i_{s+1}}^2 M_{i_{s+2}}^2\cdots M_{i_{s+t}}^2) - \mu_{s,t}| \to 0 \qquad (4.5)$$

as $n\to\infty$ where the supremum is taken over all distinct $i_1,i_2,\ldots,i_{s+t}\in[k]^{s+t}$ such that $M_{i_1},M_{i_2},\ldots,M_{i_{s+t}}$ commute with each other.

In Lemma 4.4.1 we will show that if condition $\mathbf{A}$ holds then there is a random vector (θ,γ) taking values in $[-1,1]\times[0,1]$ such that $\mu_{s,t}=\mathbb{E}[\theta^s\gamma^t]$ for all $s,t\geq 0$.

Now we are ready to state our main theorem.

Theorem 4.2.4. *Assume that* $\|\mathbf{a}(n)\|_2 = 1, \max|a_i^{(n)}| \to 0, N(\mathbf{a}(n);\mathbf{M}(n)) \to 0$ *as* $n\to\infty$. *Also assume that condition* ***A*** *holds with* $\mu_{s,t}=\mathbb{E}[\theta^s\gamma^t]$ *for some random vector* (θ,γ) *such that* $\gamma\geq\theta^2$ *a.s. Let* Λ_n *be the empirical spectral distribution of*

$$A_n = \sum_{i=1}^{k(n)} a_i^{(n)} Z_i M_i^{(n)}$$

where Z_i*'s are i.i.d. standard Gaussian random variables. Then* Λ_n *converges in distribution (with respect to the topology of weak convergence of probability measures on* $\mathbb{R}$*) to a random distribution* Λ_∞ *in probability where* $\Lambda_\infty = \nu_Z, Z$ *is* $N(0,1)$ *and* ν_z *is the unconditional distribution of* Y *where* $Y\sim N(\theta z,\gamma-\theta^2)$ *conditional on* (θ,γ).

Note that, condition $\mathbf{A}$ is not very easy to check. But there is one case in which it is easier to check that condition.

Lemma 4.2.5. *Suppose that* M_i*'s are affine transformation of projection matrices. Suppose that* $\Delta_{1,0}(n),\Delta_{2,0}(n)\to 0$ *as* $n\to\infty$ *for some numbers* $\mu_{1,0},\mu_{2,0}$ *where* $\Delta_{s,t}(n)$ *is as defined in* (4.5). *Suppose that* $\mu_{2,0}=\mu_{1,0}^2$. *Also assume that* $\overline{\mathrm{Tr}}(M_{i_1}M_{i_2}\cdots M_{i_s})$ *depends only on* s *when* $1\leq i_1,i_2,\ldots,i_s\leq k$ *are distinct and* $M_{i_1},M_{i_2},\ldots,M_{i_s}$ *commute with each other. Then condition* ***A*** *holds for all* $s,t\geq 1$.

The proof of Lemma 4.2.5 is an easy consequence of Lemma 4.4.1 using subsequence argument. We also note that if Condition **A** is satisfied for M_i's with $\mu_{s,t} = \mathbb{E}[\theta^s\gamma^t]$, then Condition **A** is also satisfied for $pI + qM_i$'s with $\mu_{s,t} = \mathbb{E}[(p + q\theta)^s(p^2 + 2pq\theta + q^2\gamma)^t]$.

We will use method of moments to prove convergence in distribution in Theorem 4.2.4. For a nonnegative integer s, define

$$\Lambda_n(x^s) = \int_{\mathbb{R}} x^s d\Lambda_n(x) = \overline{\mathrm{Tr}}(A_n^s).$$

First we show the following result. Recall that

$$N_n := \sum_{(i,j)\in E} a_i^2 a_j^2$$

where $E = \{(i,j) : i < j, M_iM_j \neq M_jM_i\}$. For noninteger t define $\mu_{s,t} = 0$.

Lemma 4.2.6. *Let $s \geq 1$ be fixed. Then we have*

$$\mathbb{E}\left(\Lambda_n(x^s) - \sum_{r=0}^{s} \binom{s}{r} \nu_{s-r}\mu_{r,(s-r)/2} W^{s-r} Y_r\right)^2 \leq C_s(c_n^2 + N_n + \max_{0\leq r\leq s/2} \Delta_{s-2r,r}(n))$$

where

$$W_n = \left(\sum_{i=1}^{k} a_i^2 Z_i^2\right)^{1/2} \quad \textit{and } Y_r(n) := r! \sum_{1\leq i_1<i_2<\cdots<i_r\leq k} \prod_{j=1}^{r} a_{i_j} Z_{i_j}.$$

Now using standard results about convergence of W_n and $Y_r(n)$ we will complete the proof of Theorem 4.2.4.

4.3 Examples

All our examples involve matrices arising from finite dimensional irreducible representations of permutation group. Similar results can also be proved for other classical Coxeter groups of which permutation group is one example and it will be developed in a future research. Let $\mathfrak{S}_n$ denote the permutation group on the set $[n] := \{1, 2, \ldots, n\}$. We will write the elements of $\mathfrak{S}_n$ in cycle notation following the usual convention of omitting cycles of length one, so that $(1, 2)$ will denote the permutation that interchanges 1 and 2 while keeping other numbers fixed. We will denote the identity permutation by e.

Suppose that ρ is a d-dimensional unitary representation of $\mathfrak{S}_n$. That is ρ is a group homomorphism from $\mathfrak{S}_n$ to the group $\mathrm{GL}_d(\mathbb{R})$, the group of $d \times d$ invertible matrices so that ρ maps identity element to identity element and $\rho(\sigma\tau) = \rho(\sigma)\rho(\tau)$ for all $\sigma, \tau \in \mathfrak{S}_n$.

In order to describe the examples we need some basic results from the representation theory of symmetric groups. The results are available in any standard sources, such as [44, 59, 77, 78, 99, 102]. First of all, following the arguments in [43] we will consider only irreducible unitary representation ρ of $\mathfrak{S}_n$, as any unitary representation may be decomposed into a direct sum of irreducible representations. Secondly, as we are only interested in the spectra, we need the equivalence class of the representation ρ, where two representations ρ and π are unitarily equivalent if there is a unitary matrix U such that $\rho(g) = U\pi(g)U^*$ for all $g \in \mathfrak{S}_n$.

The equivalence classes of irreducible representations of $\mathfrak{S}_n$ are indexed by the partitions λ of $[n]$. A partition is a non-increasing sequence of integers $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ and $|\lambda| := \lambda_1 + \lambda_2 + \cdots + \lambda_k = n$. We will use the standard notation $\lambda \vdash n$ to denote λ is a partition of n. Let ρ_λ be the irreducible unitary representation indexed by λ. We first look at the dimension d_λ of ρ_λ.

Any partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k) \vdash n$ can be visualized as an "Young diagram", that is, a left-justified array of boxes with k rows, where the top or the first row contains λ_1 many boxes, the second row contains λ_2 many boxes and so on. We number the boxes using the usual matrix convention, so that (i, j) denote the j-th box in the i-th row, we denote it by $(i, j) \in \lambda$. The hook length of the (i, j)-th box is defined as $h_{ij} = 1+$number of boxes strictly right of the (i, j)-th box + number of boxes strictly below the (i, j)-th box. The dimension d_λ is then given by (see [102])

$$d_\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h_{ij}}.$$

Given a partition $\lambda \vdash n$, its conjugate partition λ' is defined by transposing the Young diagram of λ, that is, $\lambda' = (\lambda'_1, \lambda'_2, \ldots, \lambda'_l)$ where $l = \lambda_1$ and $\lambda'_j := \max\{i : (i, j) \in \lambda\}$.

To describe the results we will use the following version of the Frobenius coordinates of a Young diagram λ:

$$f_m(\lambda) := \max\{i : (m, i) \in \lambda\} - m + \frac{1}{2} \tag{4.6}$$

$$g_m(\lambda) := \max\{i : (i, m) \in \lambda\} - m + \frac{1}{2} \tag{4.7}$$

where $m = 1, 2, \ldots, r(\lambda)$ and $r(\lambda) = \max\{k : (k, k) \in \lambda\}$ is the length of the main diagonal of λ. Note that $f_m(\lambda) = \lambda_m - m + 1/2$, $g_m(\lambda) = \lambda'_m - m + 1/2$ and $\sum_{m=1}^{r(\lambda)}(f_m(\lambda) + g_m(\lambda)) = n$.

Given a pair of sequences $\alpha = (\alpha_1, \alpha_2, \ldots), \beta = (\beta_1, \beta_2, \ldots)$ (finite or infinite), define

$$p_m(\alpha, \beta) := \sum_{i\geq 1} \alpha_i^m + (-1)^{m+1} \sum_{i\geq 1} \beta_i^m$$

for $m \geq 1$. Also let σ_m denote the cyclic permutation $(12\cdots m)$ consisting of one cycle of length m. The following result is from [66, Lemma 2, pp. 77].

Lemma 4.3.1. *The following are equivalent for a sequence of partitions $\lambda^{(n)} \vdash n$.*

1. *For each $m \geq 2$ the limit $\lim_{n\to\infty} \overline{\mathrm{Tr}}(\rho_{\lambda^{(n)}}(\sigma_m))$ exists.*
2. *For each $i = 1, 2, \ldots$ the limits*

$$\lim_{n\to\infty} \frac{f_i(\lambda^{(n)})}{n} = \alpha_i, \lim_{n\to\infty} \frac{g_i(\lambda^{(n)})}{n} = \beta_i \tag{4.8}$$

exist.

Moreover, if these conditions are satisfied then

$$\lim_{n\to\infty} \overline{\mathrm{Tr}}(\rho_{\lambda^{(n)}}(\sigma_m)) = p_m(\alpha, \beta).$$

Note that the numbers α_i and β_i denotes, respectively, the frequencies of the boxes in the i-th row and i-th column in the growing Young diagram and they satisfy $\sum_{i\geq 1} \alpha_i + \sum_{i\geq 1} \beta_i \leq 1$. We also have the following result concerning asymptotic multiplicativity of irreducible characters with respect to cycles.

Lemma 4.3.2. *Suppose* (4.8) *holds. Then for any fixed permutation σ we have*

$$\lim_{n\to\infty} \overline{\mathrm{Tr}}(\rho_{\lambda^{(n)}}(\sigma)) = \prod_{m\geq 1} (p_m(\alpha, \beta))^{k_m}$$

where σ contains k_m cycles of length m in its disjoint cycle decomposition.

In fact, Thoma's theorem (see [109]) states that all normalised irreducible characters of the infinite symmetric group $\mathfrak{S}_\infty = \cup_{n\geq 1} \mathfrak{S}_n$ arises in the above way. Lemma 4.3.2 can be proved directly using the explicit expression for characters evaluated at a fixed permutation (see [72, 73]). From now on we will assume that we have a sequence of partitions $\lambda^{(n)} \vdash n$ satisfying (4.8).

For $n \geq 1$, let $\mathcal{C}_n(2)$ denote the conjugacy class of all two cycles in $\mathfrak{S}_n$, that is $\mathcal{C}_n(2) = \{(i,j) : 1 \leq i < j \leq n\} \subset \mathfrak{S}_n$. Note that $|\mathcal{C}_n(2)| = n(n-1)/2$. Define

$$M_{i,j} = \frac{1}{2}\left(I - \rho_{\lambda^{(n)}}((i,j))\right)$$

for $1 \leq i < j \leq n$. Since $(i,j)^2 = e$, the identity permutation and $\rho_{\lambda^{(n)}}$ is a unitary representation, we have

$$M_{i,j}^* = M_{i,j} \text{ and } M_{i,j}^2 = M_{i,j}$$

for all $i < j$. Hence $M_{i,j}$'s are projection matrices. By Lemma 4.3.2, it is easy to see that condition **A** is satisfied with $\mu_{s,t} = \theta^{s+t}$ where $\theta = (1 - p_2(\alpha, \beta))/2$. Now note that (i,j) and (p,q) does not commute iff $|\{i,j\} \cap \{p,q\}| = 1$. Thus we have the following result as a corollary of Lemma 4.2.1 and Theorem 4.2.4. Also note the remark after Lemma 4.2.5.

Lemma 4.3.3. *For $n \geq 1$, let $\lambda^{(n)} \vdash n$ and let ρ_n be an irreducible unitary representation of $\mathfrak{S}_n$ corresponding to $\lambda^{(n)}$. Let Λ_n be the empirical spectral distribution of the random matrix*

$$A_n = \sum_{1 \leq i < j \leq n} a_{ij}^{(n)} X_{ij} \big(pI + q\rho_n((i,j)) \big)$$

where p, q are two fixed real numbers, X_{ij}'s are i.i.d. random variables with mean zero and variance one and $\{a_{ij}^{(n)} : 1 \leq i < j \leq n\}$ is a sequence of real numbers satisfying

$$\sum_{1 \leq i < j \leq n} \left(a_{ij}^{(n)} \right)^2 = 1 \text{ and } \sum_{i=1}^{n} (\sum_{j=1}^{n} \left(a_{ij}^{(n)} \right)^2)^2 \to 0$$

as $n \to \infty$. Suppose that (4.8) *holds. Then Λ_n converges in distribution (with respect to the topology of weak convergence of probability measure on $\mathbb{R}$) to a random probability measure ν_Z where Z is standard Gaussian and ν_z is the distribution $N((p + q\theta)z, q^2(1 - \theta^2))$ where*

$$\theta = \sum_{i \geq 1} \alpha_i^2 - \sum_{i \geq 1} \beta_i^2.$$

One example where the above lemma is applicable is when all $a_{ij}^{(n)}$'s are equal. Note that, taking $p = 0, q = 1$ and $a_{ij}^{(n)} = (n-1)^{-1/2}$ when $j = i + 1$ and 0 otherwise, we get back Theorem 1.1 in [43] which is stated in Theorem 4.3.4.

Theorem 4.3.4. *For $n \geq 1$, let $\lambda^{(n)}$ be a partition of some positive integer N_n. Let ρ_n be an irreducible unitary representation of $\mathfrak{S}_{N_n}$ corresponding to $\lambda^{(n)}$. Let Λ_n be the empirical spectral distribution of the random matrix*

$$\frac{1}{\sqrt{N_n - 1}} \sum_{k=1}^{N_n - 1} Z_{n,k} \rho_n((k, k+1)),$$

where $Z_{n,1}, Z_{n,2}, \ldots, Z_{n,N_n-1}$ are independent standard Gaussian random variables. Suppose that $N_n \to \infty$ as $n \to \infty$ and

$$\lim_{n \to \infty} \frac{\sum_i \binom{\lambda_i}{2} - \sum_j \binom{\lambda'_j}{2}}{\binom{N_n}{2}} = \theta$$

exists. Then Λ_n converges in distribution (with respect to the topology of weak convergence of probability measures on $\mathbb{R}$) to a random probability measure Λ_∞ that is Gaussian with mean θZ and variance $1 - \theta^2$, where Z is a standard Gaussian random variable. In particular, the (non-random) expectation measure $\mathbb{E}(\Lambda_\infty)$ is standard Gaussian.

Also taking $q = 1, p = -\theta$ it is easy to see that the random Gaussian mean part (θZ) in the limiting distribution of Λ_n is coming from the nonzero trace of $\rho_n((i,j))$'s.

Now, we look at other conjugacy classes. The conjugacy classes of permutation groups are indexed by cycle structures. Let $\mathcal{C}_n(2^{k_2}3^{k_3}\cdots m^{k_m})$ denote the conjugacy class in $\mathfrak{S}_n$ of the permutations with k_i many cycles of length i for $i = 2, 3, \ldots, m$. Let $l = \sum_{i=2}^{m} ik_i$. Then it is easy to see that $n^{-l}|\mathcal{C}_n(2^{k_2}3^{k_3}\cdots m^{k_m})|$ converges to a constant as $n \to \infty$. Now given a permutation σ, define its support to be the set

$$s(\sigma) = \{i : s(i) \neq i\}.$$

Clearly for $\sigma \in \mathcal{C}_n(2^{k_2}3^{k_3}\cdots m^{k_m})$, $|s(\sigma)| = l$.

Let $k_1, k_2, \ldots, k_m$ be a fixed sequence of nonnegative integers. We consider the congugacy class $\mathcal{C}_n = \mathcal{C}_n(2^{k_2}3^{k_3}\cdots m^{k_m})$ of $\mathfrak{S}_n$. For $n \geq 1$, let $\lambda^{(n)} \vdash n$ and let ρ_n be an irreducible unitary representation of $\mathfrak{S}_n$ corresponding to $\lambda^{(n)}$. Also assume that $\lambda^{(n)}$ satisfies condition (4.8). Define $r = \text{l.c.m.}\{i : k_i > 0\}$. Then it is easy to see that r is the order of any element of $\mathcal{C}_n$, that is r is the smallest positive integer such that $\sigma^r = $id for $\sigma \in \mathcal{C}_n$. Also one can easily verify that the matrix

$$M_n(\sigma) = \frac{1}{r}\sum_{i=1}^{r}\left(I - \rho_n(\sigma^i)\right)$$

is a projection matrix for $\sigma \in \mathcal{C}_n$. Let $[\sigma]$ denote the cyclic subgroup generated by σ. Clearly $M_n(\sigma)$ depends only on $[\sigma]$. Also note that $\tau \in [\sigma]$ and τ, σ are conjugates imply that $[\tau] = [\sigma]$ and $s(\sigma) = s(\tau)$. Now given $n \geq 1$ consider the random matrix

$$A_n = \sum_{[\sigma]\in\mathcal{C}_n} a_{[\sigma]}^{(n)} X_{[\sigma]} M_n(\sigma)$$

where the sum is over distinct cyclic subgroups with generator from $\mathcal{C}_n$, $X_{[\sigma]}$'s are i.i.d. r.v.'s with mean zero and variance one, and $\{a_{[\sigma]}^{(n)}\}$ is a sequence of real numbers such that

$$\sum_{[\sigma]\in\mathcal{C}_n}\left(a_{[\sigma]}^{(n)}\right)^2 = 1$$

and

$$\sum_{[\sigma],[\tau]\in\mathcal{C}_n, s(\sigma)\cap s(\tau)\neq\varnothing}\left(a_{[\tau]}^{(n)}a_{[\sigma]}^{(n)}\right)^2 \to 0 \text{ as } n \to \infty.$$

Then we have the following.

Lemma 4.3.5. *Assume the conditions above. Let Λ_n be the empirical spectral distribution of A_n. Then Λ_n converges in distribution (with respect to the topology of weak* *convergence of probability measure on* $\mathbb{R}$*) to a random probability measure* ν_Z *where* Z *is standard Gaussian and* ν_z *is the distribution* $N(\theta z, \theta(1-\theta))$ *where*

$$\theta = 1 - \frac{1}{r}\sum_{i=1}^{r} K(\sigma^i)$$

where σ *is a permutation with* k_i *many cycles of length* i*,* $i = 2, 3, \ldots, m$ *and* $K(\tau)$ *is defined as*

$$K(\tau) := \prod_{i\geq 1} (p_i(\alpha, \beta))^{l_i}$$

where $2^{l_2}3^{l_3}\cdots m^{l_m}$ *is the cycle structure of* τ*.*

Clearly, the hypothesis of Lemma 4.3.5 is satisfied when all $a_{[\sigma]}^{(n)}$'s are equal. As another example, where the matrices involved are symmetric but not projection matrices, we consider the case $M_n(\sigma) = (\rho_n(\sigma) + \rho_n(\sigma^{-1}))/2$ for $\sigma \in \mathcal{C}_n$. Here we redefine the class $[\sigma] = \{\sigma, \sigma^{-1}\}$. Then under the conditions stated in Lemma 4.3.5 (with the new definition of $[\sigma]$) the same conclusion holds with ν_z the distribution $N(\theta z, (1+\gamma)/2 - \theta^2)$ where $\theta = K(\sigma), \gamma = K(\sigma^2)$ and σ is a permutation with k_i many cycles of length i, $i = 2, 3, \ldots, m$.

4.4 Generalizations

In the previous section we considered the case when all M_i's have asymptotically equal average trace. Here we generalize the result to the case when this is not the case. First we define a notation. For an index $\boldsymbol{t} = (t_1, t_2, \ldots, t_c) \in \mathbb{Z}_+^c$ and a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_c) \in \mathbb{R}^c$, define $\boldsymbol{\theta}^{\boldsymbol{t}} := \theta_1^{t_1}\theta_2^{t_2}\cdots\theta_c^{t_c}$.

Fix a positive integer c. For every n, suppose we have a sequence of positive integers $k_1(n), k_2(n), \ldots, k_c(n)$ and for each $i = 1, 2, \ldots, c$, suppose we have a sequence of real numbers $a_{i,j}^{(n)}, j = 1, 2, \ldots, k_i(n)$ and a sequence of matrices $M_{i,j}^{(n)}, j = 1, 2, \ldots, k_i(n)$ each with spectral radius smaller than 1. Assume that

$$\sum_{i=1}^{c}\sum_{j=1}^{k_i(n)} \left(a_{i,j}^{(n)}\right)^2 = 1$$

and for all $i = 1, 2, \ldots, c$

$$\sum_{j=1}^{k_i(n)} \left(a_{i,j}^{(n)}\right)^2 \to p_i$$

as $n \to \infty$. Define the set E_n as

$$E_n := \{((i,j),(k,l)) : M_{i,j}^{(n)}, M_{k,l}^{(n)} \text{ does not commute}\}.$$

Assume that

$$\max_{\substack{1\le i\le c\\ 1\le j\le k_i(n)}} |a_{i,j}^{(n)}| \to 0 \text{ and } \sum_{((i,j),(k,l))\in E_n} \left(a_{i,j}^{(n)} a_{k,l}^{(n)}\right)^2 \to 0$$

as $n \to \infty$.

Finally, instead of condition **A** we assume the following condition:

Condition B. For every sequence of integers $\boldsymbol{s} = (s_1, s_2, \ldots, s_c)$, $\boldsymbol{t} = (t_1, t_2, \ldots, t_c)$ there is a real number $\mu_{\boldsymbol{s},\boldsymbol{t}}$ such that

$$\Delta_{\boldsymbol{s},\boldsymbol{t}}(n) := \sup \left| \overline{\mathrm{Tr}} \left(\prod_{i=1}^{c} \prod_{j=1}^{s_i} M_{i,f(i,j)}^{(n)} \prod_{j=1}^{t_i} \left(M_{i,g(i,j)}^{(n)} \right)^2 \right) - \mu_{\boldsymbol{s},\boldsymbol{t}} \right| \to 0 \tag{4.9}$$

as $n \to \infty$ where the supremum is taken over all distinct indices $(i, f(i,j))$, $(i, g(i,j))$, $1 \le i \le c, 1 \le j \le k_i(n)$ such that the corresponding matrices commute with each other.

Before going to the main result of this section, let us state a lemma which identifies the constants $\mu_{\boldsymbol{s},\boldsymbol{t}}$.

Lemma 4.4.1. *Suppose condition **B** holds. Then there is a random vector $(\boldsymbol{\theta}, \boldsymbol{\gamma})$ taking values in $[-1,1]^c \times [0,1]^c$ such that $\mu_{\boldsymbol{s},\boldsymbol{t}} = \mathbb{E}[\boldsymbol{\theta}^{\boldsymbol{s}} \boldsymbol{\gamma}^{\boldsymbol{t}}]$ for all $\boldsymbol{s}, \boldsymbol{t}$.*

Our main result in this section says the following.

Theorem 4.4.2. *Assume the conditions stated above. Also assume that condition **B** holds with $\mu_{\boldsymbol{s},\boldsymbol{t}} = \mathbb{E}[\boldsymbol{\theta}^{\boldsymbol{s}} \boldsymbol{\gamma}^{\boldsymbol{t}}]$ for some random vector $(\boldsymbol{\theta}, \boldsymbol{\gamma})$. Let Λ_n be the empirical spectral distribution of*

$$A_n = \sum_{i=1}^{c} \sum_{j=1}^{k_i(n)} a_{i,j}^{(n)} Z_{i,j}^{(n)} M_{i,j}^{(n)}$$

where $Z_{i,j}^{(n)}$'s are i.i.d. standard Gaussian random variables. Then Λ_n converges in distribution (with respect to the topology of weak convergence of probability measures on $\mathbb{R}$) to a random distribution Λ_∞ in probability where $\Lambda_\infty = \nu_{\boldsymbol{Z}}$, $\boldsymbol{Z}$ is a c-dimensional standard normal random vector and $\nu_{\boldsymbol{z}}$ is the unconditional distribution of Y where $Y \sim N(\sum_{i=1}^{c} p_i \theta_i z_i, \sum_{i=1}^{c} p_i(\gamma_i - \theta_i^2))$ conditional on $(\boldsymbol{\theta}, \boldsymbol{\gamma})$.

The proof follows the same line of proof used in the the proof of Theorem 4.2.4 and so we will omit the proof. Note that if all the matrices $M_{i,j}^{(n)}$ are projection matrices or affine transformations of projection matrices, then it is enough to prove for $\boldsymbol{t} = \boldsymbol{0}$ to prove condition **B** for all $\boldsymbol{s}, \boldsymbol{t}$. Moreover a result similar to Lemma 4.2.5 also holds here. Note that Theorem 4.4.2 can be used to prove convergence results for matrices arising from random linear combinations of group representations of symmetric matrices (as in Section 4.3), when more than one conjugacy classes are involved.

Proof. First of all note that A_n can be written in the following way, by combining the terms that involve the same product of Z_σ's,

$$A_n = N_n^{-1/2} 2^{l-2k_2} \sum_{[\sigma]:\sigma\in\mathcal{C}_n} \prod_{i=2}^{m} \prod_{j=1}^{k_i} Z^{(n)}_{\sigma_{i,j}} \left(\frac{\rho_n(\sigma_{i,j}) + \rho_n(\sigma_{i,j}^{-1})}{2} - \theta_i I \right)$$

where for $\sigma \in \mathcal{C}_n$ with disjoint cycles $\sigma_{i,j}, 1 \le j \le k_i; 2 \le i \le m$, we define $[\sigma]$ as the set of two element sets $(\sigma_{i,j}, \sigma_{i,j}^{-1}), 1 \le j \le k_i; 2 \le i \le m$. □

4.5 Proofs

For a positive integer $c \ge 1$, define

$$\nu_c = \begin{cases} (c-1)(c-3)\cdots 1 & \text{if } c \text{ is even} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $\mathbb{E}[Z^c] = \nu_c$ where $Z \sim N(0,1)$. For an index $\boldsymbol{t} = (t_1, t_2, \ldots, t_c) \in \mathbb{Z}_+^c$ and a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_c) \in \mathbb{R}^c$, define $\boldsymbol{\theta}^{\boldsymbol{t}} := \theta_1^{t_1}\theta_2^{t_2}\cdots\theta_c^{t_c}$. Also define the size of $\boldsymbol{t}$ by $|\boldsymbol{t}| := t_1 + t_2 + \cdots + t_c$. We will write $a_{\boldsymbol{t}}$ instead of $a_{t_1}a_{t_2}\cdots a_{t_c}$, $X_{\boldsymbol{t}}$ instead of $X_{t_1}X_{t_2}\cdots X_{t_c}$ and $M_{\boldsymbol{t}}$ instead of $M_{t_1}M_{t_2}\cdots M_{t_c}$. We will use the notation $[n]$ to denote the set $\{1, 2, \ldots, n\}$. For the constants we will use the following convention: $C, K, \ldots$ will denote universal constants that may change from line to line, C_s will denote constants depending only on s. We will define other notations as we go along.

The following standard lemma will be useful. For completeness we give a short proof.

Lemma 4.5.1. *Let $M_1, M_2, \ldots, M_k$ be a sequence of $n \times n$ matrices. Then we have*

$$\overline{\mathrm{Tr}}(M_1 M_2 \cdots M_k) \le \|M_i\|_{tr} \prod_{j\ne i} \|M_j\| \ \text{ for all } 1 \le i \le k.$$

Proof. Let $M_1 = UDV$ be the singular value decomposition of M_1 where U, V are unitary matrices and D is a diagonal matrix consisting of absolute values of the eigenvalues of M_1. Let d_i be the i-th diagonal entry of D. For a matrix A define its max-norm by

$$\|A\|_{\max} = \max_{1\le i,j\le n} |a_{ij}|.$$

Then clearly we have $(DA)_{ij} \le |d_{ii}| \, \|A\|_{\max}$ for all $i, j \in [n]$. Thus

$$\begin{aligned}\overline{\mathrm{Tr}}(M_1 M_2 \cdots M_k) &= \overline{\mathrm{Tr}}(DVM_2\cdots M_k U) \\ &= \frac{1}{n}\sum_{i=1}^{n} d_{ii}(VM_2\cdots M_k U)_{ii} \le \frac{1}{n}\sum_{i=1}^{n} |d_{ii}| \, \|VM_2\cdots M_k U\|_{\max}.\end{aligned}$$

Now it is easy to see that $\|A\|_{\max} \leq \|A\|$ and $\|\cdot\|$ is submultiplicative. Hence simplifying we have

$$\overline{\mathrm{Tr}}(M_1M_2\cdots M_k) \leq \|D\|_{\mathrm{tr}}\|VM_2\cdots M_kU\| \leq \|M_1\|_{\mathrm{tr}}\prod_{j\neq 1}\|M_j\|.$$

Since $\overline{\mathrm{Tr}}(AB) = \overline{\mathrm{Tr}}(BA)$ the result is true for all i and we are done. □

Now we give a proof of Lemma 4.2.1. We use Steiltjes transform and the invariance result from [26] to complete the proof.

4.5.1 Proof of Lemma 4.2.1: Universality

For a probability measure μ on $\mathbb{R}$, its Steiltjes transform is defined as the function

$$m_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z}\,\mu(dx) \text{ for } z \in \mathbb{C}\setminus\mathbb{R}.$$

It is a standard fact in probability that, μ_n converges to μ weakly as $n\to\infty$ if and only if $m_{\mu_n}(z) \to m_\mu(z)$ as $n\to\infty$ for every $z\in\mathbb{C}\setminus\mathbb{R}$. For a symmetric matrix A, let m_A denote the Steiltjes transform of the empirical spectral measure Λ_A of A. It is easy to see that

$$m_A(z) = \overline{\mathrm{Tr}}((A-zI)^{-1}) \text{ for } z\in\mathbb{C}\setminus\mathbb{R}$$

where I is the identity matrix.

Hence to prove the lemma, it is enough to prove that

$$\begin{aligned} &\mathbb{E}[g(\Re(m_{B_n}(z))) - g(\Re(m_{A_n}(z)))] \to 0\\ \text{and }\ &\mathbb{E}[g(\Im(m_{B_n}(z))) - g(\Im(m_{A_n}(z)))] \to 0\end{aligned}$$

as $n\to\infty$ for every $g\in\mathcal{G}, z\in\mathbb{C}\setminus\mathbb{R}$ where $\mathcal{G}$ be the set of all thrice differentiable functions $g:\mathbb{R}\to\mathbb{R}$ such that $|g^{(i)}(x)|\leq 1$ for all $x\in\mathbb{R}$ and $i=1,2,3$ where $g^{(i)}$ is the i-th derivative of g. Here $\Re(z),\Im(z)$ denote, respectively, the real and complex part of z. By similarity it is enough to prove the first one.

Fix $n\geq 1$ and $z=u+iv\in\mathbb{C}\setminus\mathbb{R}$ where $v\neq 0$. Recall that

$$c_n = \max_{1\leq i\leq k}|a_i| \text{ and } b_n = \max_{1\leq i\leq k}\|M_i\|_{\mathrm{tr}}.$$

Define the functions $A:\mathbb{R}^k\to\mathbb{R}^{k\times k}, G:\mathbb{R}^k\to\mathbb{C}^{k\times k}$ and $f:\mathbb{R}^k\to dR$ as follows

$$A(\mathbf{x}) = \sum_{i=1}^k a_ix_iM_i,\ G(\mathbf{x}) = (A(\mathbf{x})-zI)^{-1} \text{ and } f(\mathbf{x}) := \Re\,\overline{\mathrm{Tr}}(G(\mathbf{x}))$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$. Clearly A, G, f are infinitely differentiable functions of $\mathbf{x}$. We also have

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= -a_i \Re\,\overline{\mathrm{Tr}}(GM_iG) \\ \frac{\partial^2 f}{\partial x_i^2} &= 2a_i^2 \Re\,\overline{\mathrm{Tr}}(GM_iGM_iG) \\ \text{and } \frac{\partial^3 f}{\partial x_i^3} &= -6a_i^3 \Re\,\overline{\mathrm{Tr}}(GM_iGM_iGM_iG) \end{aligned}$$

for all $i = 1, 2, \ldots, k$. Using Lemma (4.5.1) and the fact that $\|M_i\|_{\mathrm{tr}} \le b_n \le 1$, $\|M_i\| \le 1$ and $\|G(\mathbf{x})\| \le |v|^{-1}$ for all $\mathbf{x} \in \mathbb{R}^k$, we have

$$\left|\frac{\partial f}{\partial x_i}\right| \le b_n|a_i||v|^{-2}, \left|\frac{\partial^2 f}{\partial x_i^2}\right| \le 2b_na_i^2|v|^{-3} \text{ and } \left|\frac{\partial^3 f}{\partial x_i^3}\right| \le 6b_n|a_i|^3|v|^{-4}$$

for all $\mathbf{x} \in \mathbb{R}^k$. Thus using the Lindeberg technique from Theorem 1.1 of [26] we have

$$\begin{aligned} &|\mathbb{E}(g(\Re(m_{B_n}(z))) - g(\Re(m_{A_n}(z)))| = |\mathbb{E}\, g(f(\mathbf{X})) - \mathbb{E}\, g(f(\mathbf{Z}))| \\ &\le C_v b_n \sum_{i=1}^k a_i^2 \left[\mathbb{E}(X_i^2; |X_i| \ge L) + \mathbb{E}(Z_i^2; |Z_i| \ge L)\right] \\ &\quad + C_v b_n \sum_{i=1}^k |a_i|^3 \left[\mathbb{E}(|X_i|^3; |X_i| < L) + \mathbb{E}(|Z_i|^3; |Z_i| < L)\right] \end{aligned}$$

where $C_v = 6\max\{|v|^{-3}, |v|^{-6}\}$. Now using the fact that $\mathbb{E}[X_i^2] = \mathbb{E}[Z_i^2] = 1$ we have, taking $L = c_n^{-1/2}$

$$\begin{aligned} &|\mathbb{E}(g(\Re(m_{B_n}(z))) - g(\Re(m_{A_n}(z)))| \\ &\le C_v \|\mathbf{a}\|^2 b_n \left[\max_{1\le i\le k} \mathbb{E}(X_i^2; |X_i| \ge c_n^{-1/2}) + \mathbb{E}(Z_1^2; |Z_1| \ge c_n^{-1/2}) + c_n^{1/2}\right]. \end{aligned} \tag{4.10}$$

By our assumption that $c_n \to 0$ and $\|\mathbf{a}\|^2 b_n$ is uniformly bounded as $n \to \infty$ the right hand side of equation (4.10) converges to zero as $n \to \infty$. Thus we are done.

4.5.2 Proof of the main theorem: Theorem 4.2.4

Before delving into the proof let us recall some facts about Hermite polynomials and multiple Wiener integeral (See [57, 88]). The Hermite polynomials H_n are defined by the generating function

$$\sum_{n=0}^{\infty} t^n H_n(x) := \exp\left(tx - \frac{x^2}{2}\right).$$

Equivalently,

$$H_n(x) = \frac{(-1)^n}{n!} \exp\left(\frac{x^2}{2}\right) \frac{d^n}{dx^n} \exp\left(-\frac{x^2}{2}\right).$$

The polynomial H_n has degree n with leading coefficient $\frac{1}{n!}$. If Z is a standard Gaussian random variable, then

$$\mathbb{E}[H_m(Z)H_n(Z)] = \begin{cases} \frac{1}{n!}, & \text{if } m = n, \\ 0, & \text{otherwise.} \end{cases}$$

Let W be the usual white noise on $[0,1]$. For an $L^2([0,1]^m)$ function g, define $I_m(g)$ to be the *multiple Wiener integral*

$$\int_{[0,1]^m} g(t_1, t_2, \cdots, t_m) W(dt_1) \cdots W(dt_m)$$

for $m \geq 1$. I_m is an continuous operator on $L^2([0,1]^m)$. For an $L^2([0,1])$ function f we have $n!H_m(W(f)) = I_m(f^{\otimes m})$ where $f^{\otimes m}(t_1, t_2, \cdots, t_m) = f(t_1)f(t_2)\cdots f(t_m)$ and $W(f)$ is the Ito integral $\int_0^1 f(t)W(dt)$.

Let us recall the setup first. Let Λ_n be the empirical spectral distribution of the matrix

$$A_n := \sum_{i=1}^{k} a_i Z_i M_i$$

where Z_i's are i.i.d. standard Gaussian random variables, M_i's are $d \times d$ symmetric matrices with $\|M_i\| \leq 1$ and $\|M_i\|_{\text{tr}} \leq t_n$ and $\{a_i\}$ is a sequence of real numbers with $\sum_{i=1}^k a_i^2 = 1$. We also have

$$c_n := \max_{1 \leq i \leq k} |a_i| \to 0 \text{ and } N_n := \sum_{(i,j) \in E} a_i^2 a_j^2 \to 0$$

as $n \to \infty$ where $E = \{(i,j) : i < j, M_iM_j \neq M_jM_i\}$. Moreover, by our assumption condition **A** (4.5) is satisfied with $\mu_{s,t} = \mathbb{E}[\theta^s \gamma^t]$ for $s, t \geq 0$. Fix an integer $s \geq 1$. Using Lemma 4.2.6 we have

$$\mathbb{E}\left(\Lambda_n(x^s) - \sum_{r=0}^{s} \binom{s}{r} \nu_{s-r}\mu_{r,(s-r)/2} W^{s-r} Y_r\right)^2 \\ \leq C_s(c_n^2 + N_n + \max_{0 \leq r \leq s/2} \Delta_{s-2r,r}(n))$$

where

$$W_n = \left(\sum_{i=1}^{k} a_i^2 Z_i^2\right)^{1/2} \text{ and } Y_r(n) := r! \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq k} \prod_{j=1}^{r} a_{i_j} Z_{i_j}.$$

To prove convergence of $\Lambda_n(x^s)$ we will use a coupling argument. Let n be fixed.

The grand coupling: Let W be a white noise on $[0,1]$. For simplicity we define

$$b_i^{(n)} = \sum_{j=1}^{i} \left(a_j^{(n)}\right)^2$$

for $i = 1, 2, \ldots, k$ to be the partial sum of the a_i-squared. Define

$$Z_i^{(n)} := \begin{cases} W((b_{i-1}^{(n)}, b_i^{(n)}])/a_i^{(n)} & \text{if } a_i^{(n)} \neq 0 \\ X_i & \text{if } a_i^{(n)} = 0 \end{cases}$$

for $i = 1, 2, \ldots, k(n)$ where X_i's are i.i.d. $\mathrm{N}(0,1)$ random variable independent of W. It is easy to see that using the same white noise for all n this definition gives a valid coupling of all the A_n's. Note that $Z_i^{(n)}$ always appears as $a_i^{(n)} Z_i^{(n)}$ for all $i = 1, 2, \ldots, k(n)$. Hence what really matters is the $a_i^{(n)} \neq 0$ case. Fix $r, t \geq 0$. To find the limit of $\Lambda_n(x^s)$ we will use the following standard lemma.

Lemma 4.5.2. *Under the above coupling we have*

$$\sum_{1 \leq i_1 < \cdots < i_r < k(n)} \prod_{i=1}^{r} a_{i_j}^{(n)} Z_{i_j}^{(n)} \to H_r(V)$$

in L^2 and hence in probability where $V = W((0,1])$ and H_r is the Hermite polynomial of degree r.

Proof. For $i = 1, 2, \ldots, k(n)$, define the set $A_{i,n} = (s_{i-1,n}, s_{i,n}]$ where

$$s_{i,n} = \sum_{j=1}^{i} \left(a_j^{(n)}\right)^2 \text{ for } 1 \leq i \leq k(n).$$

Recall the grand coupling. If we define the function

$$f_n(x_1, x_2, \ldots, x_r) = \sum_{\substack{1 \leq i_1, i_2, \ldots, i_r < k(n) \\ \text{all are distinct}}} \mathbf{1}_{A_{i_1,n} \times A_{i_2,n} \times \cdots A_{i_r,n}}(x_1, x_2, \ldots, x_r) \tag{4.11}$$

then we have

$$\sum_{1 \leq i_1 < \cdots < i_r < k(n)} \prod_{i=1}^{r} a_{i_j}^{(n)} Z_{i_j}^{(n)} = \frac{1}{r!} I_r(f_n) \tag{4.12}$$

where $I_r(f)$ is the multiple Wiener integral of f w.r.t. the white noise W. It is easy to see that

$$||f_n - \mathbb{1}_{(0,1]^r}||_2^2 \leq C_r c_n^2 \tag{4.13}$$

where C_r is a universal constant depending only on r. Now note that

$$I_t(\mathbb{1}_{(0,1]^r}) = r! H_r(W((0,1]))$$

where H_r is the Hermite polynomial of degree r. Now the proof is complete by L^2-continuity of the I_r operator. □

Clearly $V \sim N(0,1)$. Now note that $W_n^2 = \sum_{i=1}^{k_n} (a_i^{(n)} Z_i^{(n)})^2$ converges to 1 in L^2 under the condition $c_n = \max_i |a_i^{(n)}| \to 0$ as $n \to \infty$. Hence W_n^s converges to 1 in probability for any $s \geq 0$. Combining these results we have

$$\Lambda_n(x^s) \to \sum_{r=0}^{s} \frac{s!}{(s-r)!} \nu_{s-r} \mu_{r,(s-r)/2} H_r(V).$$

in probability. Define the function b_s by

$$b_s(z) = \sum_{r=0}^{s} \frac{s!}{(s-r)!} \nu_{s-r} \mu_{r,(s-r)/2} H_r(z).$$

Recall that $\nu_s = \mathbb{E}[Z^s]$ where Z is a N$(0,1)$ random variable and $\mu_{s,t} = \mathbb{E}[\theta^s \gamma^t]$ if s, t are nonnegative integers and zero otherwise. Thus we have

$$\begin{aligned}\sum_{s=0}^{\infty} b_s(z) \frac{x^s}{s!} &= \mathbb{E}\left[\sum_{s=0}^{\infty}\sum_{r=0}^{s} \frac{x^s \nu_{s-r}}{(s-r)!} \gamma^{(s-r)/2} \theta^r H_r(z)\right] \\ &= \mathbb{E}\left[\left(\sum_{s=0}^{\infty} \frac{\nu_s \gamma^{s/2} x^s}{s!}\right)\left(\sum_{r=0}^{\infty} x^r \theta^r H_r(z)\right)\right].\end{aligned}$$

Now note that $\sum_{s=0}^{\infty} \gamma^{s/2} x^s \nu_s / s! = \mathbb{E}[e^{\gamma^{1/2} x Z}] = e^{\gamma x^2/2}$. And the second term in the product can be written as

$$\sum_{r=0}^{\infty} \theta^r x^r H_r(z) = \exp\left[\theta x z - \theta^2 x^2/2)\right].$$

Hence we have

$$\sum_{s=0}^{\infty} b_s(z) \frac{x^s}{s!} = \mathbb{E} \exp\left[x\theta z + \left(\gamma - \theta^2\right) x^2/2\right]$$

which we recognize as the moment generating function of a probabiliy distribution which conditional on (θ, γ) is normal with mean θz and variance $\gamma - \theta^2$. This completes the proof.

Proof of Lemma 4.2.6. Fix $s \geq 1$. Recall that $||\boldsymbol{a}||_2 = 1$. We say that a random variable X is "negligible" if $\mathbb{E}[X^2] \leq C_s(c_n^2 + N_n)$. Consider the s-th moment under the spectral measure Λ_n,

$$\Lambda_n(x^s) = \int x^s \Lambda_n(x) = \frac{1}{d} \operatorname{Tr}(A_n^s) = \sum_{\boldsymbol{i} \in [k]^s} a_{\boldsymbol{i}} Z_{\boldsymbol{i}} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}}). \tag{4.14}$$

Here recall that $x_{\boldsymbol{i}}$ is a shorthand for $x_{i_1} x_{i_2} \cdots$. Given an index set $\boldsymbol{i} = (i_1, i_2, \ldots, i_s)$ define the edge-labeled graph $H_{\boldsymbol{i}} = ([s], E_{\boldsymbol{i}})$ as follows:

$$(p, q) \in E_{\boldsymbol{i}} \text{ iff } (i_p, i_q) \in E \text{ or } i_p = i_q$$

and the edge (p, q) is marked zero if $i_p = i_q$ and one otherwise. For an edge labeled graph H, $\hat{H}$ will denote the *skeleton* of H, the graph H without the edge labels. Let $\mathfrak{C}_s$ be the set of all graphs with vertex set $[s]$ where each edge is labeled with either 0 or 1. Clearly $|\mathfrak{C}_s| = 3^{\binom{s}{2}}$. Since s is fixed, $\mathfrak{C}_s$ doesn't depend on n. Note that

$$\Lambda(x^s) = \sum_{H \in \mathfrak{C}_s} \left[\sum_{\boldsymbol{i} \in [k]^s : H_{\boldsymbol{i}} = H} a_{\boldsymbol{i}} Z_{\boldsymbol{i}} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}}) \right]. \tag{4.15}$$

We will prove the lemma in four steps.

First Reduction: First of all we will prove that the contribution from $H \in \mathfrak{C}_s$ which has at least one connected component of size 3 or more, is negligible. Fix $H \in \mathfrak{C}_s$ such that H has at least one component of size 3 or more. Recall that $\overline{\operatorname{Tr}}(M_{\boldsymbol{i}}) \leq 1$ for all $\boldsymbol{i}$. Hence we have

$$\mathbb{E} \left[\sum_{\boldsymbol{i} : H_{\boldsymbol{i}} = H} a_{\boldsymbol{i}} Z_{\boldsymbol{i}} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}}) \right]^2 \leq \sum_{\boldsymbol{i}, \boldsymbol{t} : H_{\boldsymbol{i}} = H_{\boldsymbol{t}} = H} |a_{\boldsymbol{i}} a_{\boldsymbol{t}}| \left| \mathbb{E}(Z_{\boldsymbol{i}} Z_{\boldsymbol{t}}) \right|. \tag{4.16}$$

Now note that $\mathbb{E}[Z_{\boldsymbol{i}} Z_{\boldsymbol{t}}] = \mathbb{E} \prod_{j=1}^s Z_{i_j} Z_{t_j} = 0$ unless all indices i_j, t_j occur even number of times. Also $\mathbb{E}\,|Z_{\boldsymbol{i}} Z_{\boldsymbol{t}}|$ is uniformly bounded for $\boldsymbol{i}, \boldsymbol{t} \in [k]^s$. Since $H_{\boldsymbol{i}}$ contains one component of size more than 3 either there are distinct $p, q, r \in [s]$ such that $i_p = i_q = i_s$ or there are distinct $p, q \in [s]$ such that $i_p \neq i_q, (i_p, i_q) \in E$. Hence we have

$$\begin{aligned} \mathbb{E} \left[\sum_{\boldsymbol{i} : H_{\boldsymbol{i}} = H} a_{\boldsymbol{i}} Z_{\boldsymbol{i}} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}}) \right]^2 &\leq C_s \left(\sum_{i=1}^k a_i^2 \right)^{s-2} \left(\sum_{(i,j) \in E} a_i^2 a_j^2 + \sum_{i=1}^k a_i^4 \right) \\ &\leq C_s(N_n + c_n^2). \end{aligned} \tag{4.17}$$

Second Reduction: From (4.17) we know that the main contribution comes graphs with connected components of size at most two. Fix a $H \in \mathfrak{C}_s$. Suppose H has r

components of size one, p components of size two with label zero and q components of size two with label one. Here we will prove that the contribution is negligible if $q > 0$. As before we have $\mathbb{E}[Z_{\boldsymbol{i}} Z_{\boldsymbol{t}}] \neq 0$ only if all indices occur even number of times. Now, under the assumption that $q > 0$, there are distinct $l, m \in [s]$ such that $i_l \neq i_m, (i_l, i_m) \in E$. Hence

$$\mathbb{E}\left[\sum_{\boldsymbol{i}:H_{\boldsymbol{i}}=H} a_{\boldsymbol{i}} Z_{\boldsymbol{i}} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}})\right]^2 \leq C_s \left(\sum_{i=1}^{k} a_i^2\right)^{s-2} \sum_{(i,j)\in E} a_i^2 a_j^2 = C(s) N_n.$$

Third Reduction: Hence the main contribution comes from graphs $H \in \mathfrak{C}_s$ whose connected components are either of size one or of size two with label zero. Let r be the number of components of size one and t be the number of components of size two in H. Clearly $s = r + 2t$. Note that if $H_{\boldsymbol{i}} = H$ then σ_{i_j}'s commute for all $j \in [s]$. Then the number of connected graphs on vertex set $[s]$ with r connected components of size one and t connected components of size two is $\binom{s}{r}\nu_{s-r}$, since there are $\binom{s}{r}$ ways to choose the vertices that will comprise the r connected components of size one and $(2t-1)(2t-3)\cdots 1$ ways to match the remaining $2t = s - r$ vertices into unordered pairs that will comprise the t connected components of size two. Let H_r be the graph with vertex set $[s]$ and edge set $\{(r+2i-1, r+2i) : i \geq 1\}$ and all the edge labels are zero. Then combining everything we have

$$\mathbb{E}\left[\Lambda_n(x^s) - \sum_{r=0}^{s} \binom{s}{r} \nu_{s-r} \left(\sum_{\boldsymbol{i}:H_{\boldsymbol{i}}=H_r} a_{\boldsymbol{i}} Z_{\boldsymbol{i}} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}})\right)\right]^2 \leq C_s(c_n^2 + N_n).$$

Fourth Reduction: Fix $r \in [s]$ such that 2 divides $s - r$. Let $2t = s - r$. Consider the term

$$\begin{aligned} Y_r' &:= \sum_{\boldsymbol{i}\in[k]^s:H_{\boldsymbol{i}}=H_r} a_{\boldsymbol{i}} Z_{\boldsymbol{i}} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}}) \\ &= \sideset{}{'}\sum_{(i_p,i_q)\notin E_S \text{ for all } 1\leq p,q\leq r+t} \overline{\operatorname{Tr}}(M_{\boldsymbol{i}}) \prod_{j=1}^{r} a_{i_j} Z_{i_j} \prod_{j=r+1}^{r+t} a_{i_j}^2 Z_{i_j}^2 \end{aligned}$$

where $\sum'$ denotes sum over distinct indices. Using condition **A** (eqn. (4.5)) we have

$$\mathbb{E}\,|Y_r' - \mu_{r,t} Y_r''|^2 \leq C_s \Delta_{s,t}(n)$$

where

$$Y_r'' = \sideset{}{'}\sum_{(i_p,i_q)\notin E_S \text{ for all } 1\leq p,q\leq r+t} \prod_{j=1}^{r} a_{i_j} Z_{i_j} \prod_{j=r+1}^{r+t} a_{i_j}^2 Z_{i_j}^2$$

If we define

$$Y_r''' := \left[\sum_{(i_p,i_q)\notin E_S}' \prod_{j=1}^{r} a_{i_j} Z_{i_j} \right] \left(\sum_{i=1}^{k} a_i^2 Z_i^2 \right)^t$$

calculations similar to the previous ones show that

$$\begin{aligned} \mathbb{E}(Y_r'' - Y_r''')^2 &\leq C_s \left(\sum_{i=1}^{k} a_i^2 \right)^{r+t-1} \left(\sum_{(i,j)\in E_S} a_i^2 a_j^2 + \sum_{i=1}^{k} a_i^4 \right) \\ &= C_s(c_n^2 + N_n). \end{aligned}$$

Let

$$Y_r := \sum_{(i_p,i_q)\notin E_S}' \prod_{j=1}^{r} a_{i_j} Z_{i_j}.$$

and $W = \sqrt{\sum_{i=1}^{k} a_i^2 Z_i^2}$. Then we have

$$\begin{aligned} &\mathbb{E}\left[\Lambda(x^s) - \sum_{r=0}^{s} \binom{s}{r} \nu_{s-r} \mu_{r,(s-r)/2} W^{s-r} Y_r \right]^2 \\ &\leq C_s(c_n^2 + N_n + \max_{0\leq r\leq s/2} \Delta_{s-2r,r}). \end{aligned} \tag{4.18}$$

Now note that if we drop the condition $\{(i_p, i_q) \notin E_S\}$ in the defining sum for Y_r the result (4.18) is still true. Combining all the results we have the proof. □

4.5.3 Proof of Lemma 4.4.1

Let us recall condition **B** first.

Condition B. For every sequence of integers $\boldsymbol{s} = (s_1, s_2, \ldots, s_c)$, $\boldsymbol{t} = (t_1, t_2, \ldots, t_c)$ there is a real number $\mu_{\boldsymbol{s},\boldsymbol{t}}$ such that

$$\Delta_{\boldsymbol{s},\boldsymbol{t}}(n) := \sup \left| \overline{\text{Tr}} \left(\prod_{i=1}^{c} \prod_{j=1}^{s_i} M_{i,f(i,j)}^{(n)} \prod_{j=1}^{t_i} \left(M_{i,g(i,j)}^{(n)} \right)^2 \right) - \mu_{\boldsymbol{s},\boldsymbol{t}} \right| \to 0 \tag{4.19}$$

as $n \to \infty$ where the supremum is taken over all distinct indices $(i, f(i,j))$, $(i, g(i,l))$, $1 \leq j \leq s_i; 1 \leq l \leq t_i; 1 \leq i \leq c$ such that the corresponding matrices commute with each other.

We will use the solution of the multidimensional Hausdorff problem (cf. proposition 4.6.11 from [11]) to prove that there is a $[-1,1]^c \times [-1,1]^c$ valued random vector $(\boldsymbol{\theta}, \boldsymbol{\gamma})$ with $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_c), \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_c)$ such that

$$\mu_{\boldsymbol{s},\boldsymbol{t}} = \mathbb{E}[\boldsymbol{\theta}^{\boldsymbol{s}} \boldsymbol{\gamma}^{\boldsymbol{t}}]$$

for all $\boldsymbol{s}, \boldsymbol{t} = (t_1, t_2, \ldots, t_c) \in \mathbb{N}^c$. The fact that $\gamma_i \geq 0$ a.s. follows easily from a similar calculation. For simplicity we only prove that

$$\mu_{\boldsymbol{s},0} = \mathbb{E}[\boldsymbol{\theta}^{\boldsymbol{s}}].$$

The general case follows by working with $2c$ classes and taking $\gamma_i = \theta_{c+i}$ for $i = 1, 2, \ldots, c$. Equip $\mathbb{N}^c$ with the usual partial order, *i.e.,* $\boldsymbol{p} \leq \boldsymbol{n}$ if $p_i \leq n_i$ for all $i = 1, 2, \ldots, c$.

The solution of the multidimensional Hausdorff problem says that, in order that the numbers $\psi(\boldsymbol{s})$ for $\boldsymbol{s} \in \mathbb{N}^c$ are the multivariate moments of some $[0,1]^c$ valued random vector, a necessary and sufficient condition is that for all $\boldsymbol{m}, \boldsymbol{n} \in \mathbb{N}^c$

$$\sum_{\mathbf{0} \leq \boldsymbol{p} \leq \boldsymbol{n}} (-1)^{|\boldsymbol{p}|} \binom{n_1}{p_1} \binom{n_2}{p_2} \cdots \binom{n_c}{p_c} \psi(\boldsymbol{m} + \boldsymbol{p}) \geq 0. \tag{4.20}$$

If this is the case, with $(\theta_1, \theta_2, \ldots, \theta_c)$ being the random vector, the sum appearing in (4.20) is $\mathbb{E}[\theta_1^{m_1} \cdots \theta_c^{m_c} (1-\theta_1)^{n_1} \cdots (1-\theta_c)^{n_c}]$. Since we want to prove the existence of a $[-1,1]^c$ valued random variable it is enough to check the condition (4.20) with

$$\psi(\boldsymbol{t}) := \frac{1}{2^{|\boldsymbol{t}|}} \sum_{\mathbf{0} \leq \boldsymbol{s} \leq \boldsymbol{t}} \binom{t_1}{s_1} \binom{t_2}{s_2} \cdots \binom{t_c}{s_c} \mu_{\boldsymbol{s},0}. \tag{4.21}$$

This corresponds to the transformation $f : [-1,1] \mapsto [0,1]$ given by $f(x) = (1+x)/2$.

Fix $\boldsymbol{m}, \boldsymbol{n}$. Choose n large enough so that we can find distinct indices $(i, f_n(i,j))$, $1 \leq j \leq m_i + n_i; 1 \leq i \leq c$ such that the corresponding matrices commute with each other. for every n fix such a sequence f_n. Clearly we have

$$\psi(\boldsymbol{t}) = \lim_{n \to \infty} \overline{\operatorname{Tr}} \left(\prod_{i=1}^{c} \prod_{j=1}^{t_i} \frac{1}{2} (I + M_{i, f_n(i,j)}^{(n)}) \right) \tag{4.22}$$

for $\boldsymbol{t} \leq \boldsymbol{m} + \boldsymbol{n}$. By assumption all the matrices involved in (4.22) commute. Therefore the matrices are simultaneously diagonalizable by a unitary matrix, say, U_n. Let

$$D_{i,j}^{(n)} := U_n^* \cdot \frac{1}{2} (I + M_{i, f_n(i,j)}^{(n)}) \cdot U_n.$$

Clearly $D_{i,j}^{(n)}$'s are diagonal matrices with all diagonal entries lying in the interval $[0,1]$. Thus we have

$$\psi(\boldsymbol{t}) = \lim_{n \to \infty} \overline{\operatorname{Tr}} \left(\prod_{i=1}^{c} \prod_{j=1}^{t_i} D_{i,j}^{(n)} \right)$$

for all $\boldsymbol{t} \leq \boldsymbol{m} + \boldsymbol{n}$. Now

$$\begin{aligned}
&\sum_{\boldsymbol{0} \leq \boldsymbol{p} \leq \boldsymbol{n}} (-1)^{|\boldsymbol{p}|} \binom{n_1}{p_1} \binom{n_2}{p_2} \cdots \binom{n_c}{p_c} \psi(\boldsymbol{m} + \boldsymbol{p}) \\
&= \lim_{n \to \infty} \overline{\mathrm{Tr}} \prod_{i=1}^{c} \left(\sum_{0 \leq p_i \leq n_i} (-1)^{p_i} \binom{n_i}{p_i} \prod_{j=1}^{m_i + p_i} D_{i,j}^{(n)} \right) \\
&= \lim_{n \to \infty} \overline{\mathrm{Tr}} \prod_{i=1}^{c} \left(\prod_{j=1}^{m_i} D_{i,j}^{(n)} \prod_{j=1}^{n_i} (I - D_{i,m_i+j}^{(n)}) \right) \geq 0
\end{aligned}$$

for all $\boldsymbol{m}, \boldsymbol{n} \in \mathbb{N}^c$. This completes the proof.

Bibliography

[1] Kenneth S. Alexander. Approximation of subadditive functions and convergence rates in limiting-shape results. *Ann. Probab.*, 25(1):30–55, 1997.

[2] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.

[3] Richard Arratia, Larry Goldstein, and Louis Gordon. Poisson approximation and the Chen-Stein method. *Statist. Sci.*, 5(4):403–434, 1990. With comments and a rejoinder by the authors.

[4] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19:357–367, 1967.

[5] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

[6] Jinho Baik and Toufic M. Suidan. A GUE central limit theorem and universality of directed first and last passage site percolation. *Int. Math. Res. Not.*, (6):325–337, 2005.

[7] P. Baldi, Y. Rinott, and C. Stein. A normal approximation for the number of local maxima of a random function on a graph. In *Probability, statistics, and mathematics*, pages 59–81. Academic Press, Boston, MA, 1989.

[8] A. D. Barbour, Lars Holst, and Svante Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1992. Oxford Science Publications.

[9] F. Barthe, P. Cattiaux, and C. Roberto. Concentration for independent random variables with heavy tails. *AMRX Appl. Math. Res. Express*, (2):39–60, 2005.

[10] Itai Benjamini, Gil Kalai, and Oded Schramm. First passage percolation has sublinear distance variance. *Ann. Probab.*, 31(4):1970–1978, 2003.

[11] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic analysis on semigroups*, volume 100 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1984. Theory of positive definite and related functions.

[12] Shankar Bhamidi, Guy Bresler, and Allan Sly. Mixing time of exponential random graphs. In *IEEE 49th Annual IEEE Symposium on Foundations of Computer Science, 2008. FOCS'08*, pages 803–812, 2008.

[13] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.

[14] S. G. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.

[15] Sergey G. Bobkov. Large deviations and isoperimetry over convex probability measures with heavy tails. *Electron. J. Probab.*, 12:1072–1100 (electronic), 2007.

[16] Thierry Bodineau and James Martin. A universality property for last-passage percolation paths close to the axis. *Electron. Comm. Probab.*, 10:105–112 (electronic), 2005.

[17] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[18] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, 66(3):379–386, 1984.

[19] E. Bolthausen. Laplace approximations for sums of independent random vectors. II. Degenerate maxima and manifolds of maxima. *Probab. Theory Related Fields*, 76(2):167–206, 1987.

[20] E. Bolthausen, F. Comets, and A. Dembo. Large deviations for random matrices and random graphs. *In preparation*, 2009.

[21] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 2005.

[22] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003.

[23] Włodzimierz Bryc, Amir Dembo, and Tiefeng Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34(1):1–38, 2006.

[24] Sourav Chatterjee. Concentration inequalities with exchangeable pairs, 2005. Stanford University Ph. D. Thesis. Available at arXiv:math/0507526.

[25] Sourav Chatterjee. A simple invariance theorem. 2005.

[26] Sourav Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.*, 34(6):2061–2076, 2006.

[27] Sourav Chatterjee. Concentration of Haar measures, with an application to random matrices. *J. Funct. Anal.*, 245(2):379–389, 2007.

[28] Sourav Chatterjee. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321, 2007.

[29] Sourav Chatterjee. The missing log in large deviations for subgraph counts. *Preprint. Available at arXiv:1003.3498*, 2010.

[30] Sourav Chatterjee and Q.-M. Shao. Stein's method of exchangeable pairs with application to the curie-weiss model. *Preprint. Available at arXiv:0907.4450*, 2009.

[31] Sourav Chatterjee and S. R. S. Varadhan. Large deviations for subgraph counts via szemerédi regularity lemma. *Preprint. Available at arXiv:1003.3498*, 2010. In preparation.

[32] J. T. Chayes, L. Chayes, and R. Durrett. Critical behavior of the two-dimensional first passage time. *J. Statist. Phys.*, 45(5-6):933–951, 1986.

[33] J.-R. Chazottes, P. Collet, C. Külske, and F. Redig. Concentration inequalities for random fields via coupling. *Probab. Theory Related Fields*, 137(1-2):201–225, 2007.

[34] Louis H. Y. Chen and Qi-Man Shao. A non-uniform Berry-Esseen bound via Stein's method. *Probab. Theory Related Fields*, 120(2):236–254, 2001.

[35] J. Theodore Cox and Richard Durrett. Some limit theorems for percolation processes with necessary and sufficient conditions. *Ann. Probab.*, 9(4):583–603, 1981.

[36] Amir Dembo. Information inequalities and concentration of measure. *Ann. Probab.*, 25(2):927–939, 1997.

[37] Persi Diaconis and Susan Holmes, editors. *Stein's method: expository lectures and applications*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 46. Institute of Mathematical Statistics, Beachwood, OH, 2004. Papers from the Workshop on Stein's Method held at Stanford University, Stanford, CA, 1998.

[38] Hanna Döring and Peter Eichelsbacher. Moderate deviations in a random graph and for the spectrum of Bernoulli random matrices. *Electron. J. Probab.*, 14:no. 92, 2636–2656, 2009.

[39] P. Eichelsbacher and M. Lowe. Stein's method for dependent random variables occurring in statistical mechanics. *preprint*, 2009.

[40] Richard S. Ellis. *Entropy, large deviations, and statistical mechanics*, volume 271 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.

[41] Richard S. Ellis and Charles M. Newman. Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Z. Wahrsch. Verw. Gebiete*, 44(2):117–139, 1978.

[42] Richard S. Ellis and Charles M. Newman. The statistics of Curie-Weiss models. *J. Statist. Phys.*, 19(2):149–161, 1978.

[43] Steven N. Evans. Spectra of random linear combinations of matrices defined via representations and Coxeter generators of the symmetric group. *Ann. Probab.*, 37(2):726–741, 2009.

[44] William Fulton and Joe Harris. *Representation theory*, volume 129 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. A first course, Readings in Mathematics.

[45] Ivan Gentil, Arnaud Guillin, and Laurent Miclo. Modified logarithmic Sobolev inequalities and transportation inequalities. *Probab. Theory Related Fields*, 133(3):409–436, 2005.

[46] Larry Goldstein. Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.*, 42(3):661–683, 2005.

[47] Larry Goldstein and Gesine Reinert. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, 7(4):935–952, 1997.

[48] Larry Goldstein and Yosef Rinott. Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.*, 33(1):1–17, 1996.

[49] Nathael Gozlan. Characterization of Talagrand's like transportation-cost inequalities on the real line. *J. Funct. Anal.*, 250(2):400–425, 2007.

[50] Nathael Gozlan. Poincaré inequalities and dimension free concentration of measure. *Ann. Inst. Henri Poincaré Probab. Stat. (To appear)*, 2009.

[51] Janko Gravner, Craig A. Tracy, and Harold Widom. Limit theorems for height fluctuations in a class of discrete space and time growth models. *J. Statist. Phys.*, 102(5-6):1085–1132, 2001.

[52] Geoffrey Grimmett and Harry Kesten. First-passage percolation, network flows and electrical resistances. *Z. Wahrsch. Verw. Gebiete*, 66(3):335–366, 1984.

[53] Alice Guionnet. *Large random matrices: lectures on macroscopic asymptotics*, volume 1957 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006.

[54] J. M. Hammersley and D. J. A. Welsh. First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In *Proc. Internat. Res. Semin., Statist. Lab., Univ. California, Berkeley, Calif*, pages 61–110. Springer-Verlag, New York, 1965.

[55] Christopher Hammond and Steven J. Miller. Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *J. Theoret. Probab.*, 18(3):537–566, 2005.

[56] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[57] Christian Houdré and Víctor Pérez-Abreu, editors. *Chaos expansions, multiple Wiener-Itô integrals and their applications*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1994. Papers from the workshop held in Guanajuato, July 27–31, 1992.

[58] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, 1925.

[59] Gordon James and Adalbert Kerber. *The representation theory of the symmetric group*, volume 16 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1981. With a foreword by P. M. Cohn, With an introduction by Gilbert de B. Robinson.

[60] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[61] Svante Janson, Krzysztof Oleszkiewicz, and Andrzej Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[62] Svante Janson and Andrzej Ruciński. The infamous upper tail. *Random Structures Algorithms*, 20(3):317–342, 2002. Probabilistic methods in combinatorial optimization.

[63] Kurt Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.

[64] Kurt Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)*, 153(1):259–296, 2001.

[65] M. Kardar, G. Parisi, and Y.C. Zhang. Dynamic scaling of growing interfaces. *Physical Review Letters*, 56(9):889–892, 1986.

[66] S. V. Kerov. *Asymptotic representation theory of the symmetric group and its applications in analysis*, volume 219 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2003. Translated from the Russian manuscript by N. V. Tsilevich, With a foreword by A. Vershik and comments by G. Olshanski.

[67] Harry Kesten. Aspects of first passage percolation. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 125–264. Springer, Berlin, 1986.

[68] Harry Kesten. On the speed of convergence in first-passage percolation. *Ann. Appl. Probab.*, 3(2):296–338, 1993.

[69] Harry Kesten and Yu Zhang. A central limit theorem for "critical" first-passage percolation in two dimensions. *Probab. Theory Related Fields*, 107(2):137–160, 1997.

[70] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24(2):166–174, 2004.

[71] J. Krug and H. Spohn. Kinetic roughening of growing surfaces. *Solids far from equilibrium*, pages 479–582, 1991.

[72] Michel Lassalle. Explicitation of characters of the symmetric group. *C. R. Math. Acad. Sci. Paris*, 341(9):529–534, 2005.

[73] Michel Lassalle. An explicit formula for the characters of the symmetric group. *Math. Ann.*, 340(2):383–405, 2008.

[74] R. Latała and K. Oleszkiewicz. Between Sobolev and Poincaré. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 147–168. Springer, Berlin, 2000.

[75] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[76] C. Licea, C. M. Newman, and M. S. T. Piza. Superdiffusivity in first-passage percolation. *Probab. Theory Related Fields*, 106(4):559–591, 1996.

[77] Dudley E. Littlewood. *The theory of group characters and matrix representations of groups*. AMS Chelsea Publishing, Providence, RI, 2006. Reprint of the second (1950) edition.

[78] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[79] Anders Martin-Löf. A Laplace approximation for sums of independent random variables. *Z. Wahrsch. Verw. Gebiete*, 59(1):101–115, 1982.

[80] K. Marton. Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 1996.

[81] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996.

[82] K. Marton. Measure concentration and strong mixing. *Studia Sci. Math. Hungar.*, 40(1-2):95–113, 2003.

[83] Pascal Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.

[84] Colin McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[85] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[86] V. D. Milman. The heritage of P. Lévy in geometrical functional analysis. *Astérisque*, (157-158):273–301, 1988. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987).

[87] Charles M. Newman and Marcelo S. T. Piza. Divergence of shape fluctuations in two dimensions. *Ann. Probab.*, 23(3):977–1005, 1995.

[88] David Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[89] L. Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117–149, 1944.

[90] Juyong Park and M. E. J. Newman. Statistical mechanics of networks. *Phys. Rev. E (3)*, 70(6):66117–66122, 2004.

[91] Juyong Park and M. E. J. Newman. Solution for the properties of a clustered network. *Phys. Rev. E (3)*, 72(2):26136–26137, 2005.

[92] Yuval Peres and Robin Pemantle. Planar first-passage percolation times are not tight. *Probability and phase transition (G. Grimmett, ed.)*, pages 261–264, 1994.

[93] Martin Raič. CLT-related large deviation bounds based on Stein's method. *Adv. in Appl. Probab.*, 39(3):731–752, 2007.

[94] Daniel Richardson. Random growth in a tessellation. *Proc. Cambridge Philos. Soc.*, 74:515–528, 1973.

[95] Haskell P. Rosenthal. On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.*, 8:273–303, 1970.

[96] Andrzej Ruciński. When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields*, 78(1):1–10, 1988.

[97] Gideon Schechtman. Lévy type inequality for a class of finite metric spaces. In *Martingale theory in harmonic analysis and Banach spaces (Cleveland, Ohio, 1981)*, volume 939 of *Lecture Notes in Math.*, pages 211–215. Springer, Berlin, 1982.

[98] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987.

[99] Barry Simon. *Representations of finite and compact groups*, volume 10 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[100] Barry Simon and Robert B. Griffiths. The $(\phi^4)_2$ field theory as a classical Ising model. *Comm. Math. Phys.*, 33:145–164, 1973.

[101] R. T. Smythe and John C. Wierman. *First-passage percolation on the square lattice*, volume 671 of *Lecture Notes in Mathematics*. Springer, Berlin, 1978.

[102] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[103] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, Berkeley, Calif., 1972. Univ. California Press.

[104] Charles Stein. *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

[105] Toufic Suidan. A remark on a theorem of Chatterjee and last passage percolation. *J. Phys. A*, 39(28):8977–8981, 2006.

[106] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995.

[107] Michel Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[108] Michel Talagrand. A new look at independence. *Ann. Probab.*, 24(1):1–34, 1996.

[109] Elmar Thoma. Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen, symmetrischen Gruppe. *Math. Z.*, 85:40–61, 1964.

[110] Van H. Vu. A large deviation result on the number of small subgraphs of a random graph. *Combin. Probab. Comput.*, 10(1):79–94, 2001.

[111] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

[112] Yu Zhang. Shape fluctuations are different in different directions. *Ann. Probab.*, 36(1):331–362, 2008.