THE METHOD OF DENSE CYCLE CONDITIONING, ITS APPLICATION, COMPUTATION AND A RESULT ON CONCENTRATION

Debapratim Banerjee

A DISSERTATION

in

Statistics

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Zongming Ma, Associate Professor of statistics

Graduate Group Chairperson

Nancy Zhang, Professor of statistics

Dissertation Committee

Mark G. Low, Professor of Statistics

Robin Pemantle, Professor of Mathematics

Jian Ding, Associate Professor of Statistics

THE METHOD OF DENSE CYCLE CONDITIONING, ITS APPLICATION, COMPUTATION AND A RESULT ON CONCENTRATION

Dedicated to my parents and my brother

ACKNOWLEDGEMENT

I would like to thank everyone who has been supportive for the completion my dissertation work. First of all I would like to thank my family for being extremely supportive all through my graduate career. I am grateful to my adviser Prof. Zongming Ma for providing all necessary helps and support to complete my dissertation. I am grateful to the friends I made here during these four years. To mention a few I would like to thank my first year office mates especially Arun and Shaokun with whom I shared my office for my rest of my years here. I am grateful to Arun, Somabha, Mo and Matteo for having always cheerful academic discussions. I have taken some interesting courses throughout my graduate studies here. To mention a few I learned introduction to stochastic calculus from Prof. Mike Steele, I learned about various directions in probability theory from the courses of Prof. Robin Pemantle, I also enjoyed the courses from Prof. Alexander Rakhlin, Prof. Elchanan Mossel and many other courses. I am grateful to people from the administrative office specially Lauren Rudick and Gidget Murray who have been extremely helpful to complete my graduate works. Finally I would like to thank the non academic staffs for their helps on different occasions.

ABSTRACT

THE METHOD OF DENSE CYCLE CONDITIONING, ITS APPLICATION, COMPUTATION AND A RESULT ON CONCENTRATION

Debapratim Banerjee

Zongming Ma

This dissertation contains works on three different directions. In the first direction, three different problems have been solved. The fundamental theme of these problems are to consider the log-likelihood ratio of certain processes under local perturbations. It is shown that in these cases below certain threshold the log-likelihood ratio can be approximated by log-likelihood ratio restricted to a certain class of statistics called the "signed cycles". These statistics were considered by the author in order to study contiguity for planted partition model in dense case. Details are given in Chapter 2 The sparse case is known in the literature by a paper of Mossel et al. These statistics found further applications in statistics and statistical physics where two other problems were solved. One might look at Chapters 3 and 1 for details. The second direction of this thesis is to show computability of these cycle statistics. It is proved that the "signed cycles" statistics can be approximated by certain linear spectral statistics of high dimensional random matrices. The proof techniques are highly motivated by a paper of Anderson and Zeitouni. One can have a look at Chapter 4 for details. In the third direction a problem of concentration inequality is considered. A Bernstein type concentration inequality is proved for statistics which are generalizations of a statistics introduced by Hoeffding. It is proven using the method exchangeable pairs introduced by Chatterjee. One might look at Chapter 6 for details.

TABLE OF CONTENTS

ACKNOWLEDGEMENT iv

ABSTRACT v

LIST OF TABLES ix

LIST OF ILLUSTRATIONS x

PREFACE xi

CHAPTER 1 : Fluctuation of the free energy of Sherrington-Kirkpatrick model with Curie-Weiss interaction: the paramagnetic regime 1

1.1 Overview 1

1.2 Introduction 1

1.3 Main result 5

1.4 Proof techniques and related definitions 6

1.5 Construction of $\mathbb{P}_n$ and $\mathbb{Q}_n$ and asymptotic distribution of signed cycles . . 9

1.6 Proof of Theorem 1 13

1.7 Appendix 25

CHAPTER 2 : Contiguity and non-reconstruction results for planted partition models: the dense case 33

2.1 Overview 33

2.2 Introduction 33

2.3 Our results 37

2.4 A result on contiguity 39

2.5 Signed cycles and their asymptotic distributions 44

2.6 Calculation of second moment and completion of the proof of Theorem 2 . . 57

2.7 Proof of non reconstructability 62
2.8 Appendix 69

CHAPTER 3 : Optimal signal detection in some spiked random matrix models:likelihood ratio tests and linear spectral statistics 76
3.1 Introduction 76
3.2 Preliminaries 81
3.3 Main results 84
3.4 Asymptotic normality of bipartite signed cycles 88
3.5 Proof of Theorems 4 and 5 90
3.6 Proof of Theorem 6 100
3.7 Proof of Proposition 1 102
3.8 Proof of Proposition 2 107

CHAPTER 4 : Optimal hypothesis testing for planted partition model with growing degrees 122
4.1 Introduction 122
4.2 Definitions and notation 129
4.3 Linear spectral statistics and likelihood ratio tests 131
4.4 Outline of proofs 142
4.5 Preliminary combinatorics results 150
4.6 Proofs of main results 168

CHAPTER 5 : Non backtracking matrices and optimal hypothesis testing for planted partition models with growing degrees 230
5.1 overview 230
5.2 Introduction 230
5.3 Preliminaries 233
5.4 Results from Banerjee (2018) and Banerjee and Ma (2017a) 235
5.5 Our results 236

5.6 Proofs 238

CHAPTER 6 : A Bernstein type inequality for sums of choices in three dimensional arrays 252

6.1 Overview 252

6.2 Arrays and Concentration Inequalities 252

6.3 On the method of exchangeable pair 253

6.4 Strategy of the Proofs 254

6.5 Proofs of the results 255

CHAPTER 7 : Future Scopes 263

APPENDIX 264

BIBLIOGRAPHY 267

CHAPTER 1 : Fluctuation of the free energy of Sherrington-Kirkpatrick model with Curie-Weiss interaction: the paramagnetic regime

1.1. Overview

We consider a spin system with pure two spin Sherrington-Kirkpatrick Hamiltonian with Curie-Weiss interaction. The model where the spins are spherically symmetric was considered by Baik and Lee (2017) and Baik et al. (2018) which shows a two dimensional phase transition with respect to temperature and the coupling constant. In this paper we prove a result analogous to Baik and Lee (2017) in the "paramagnetic regime" when the spins are i.i.d. Rademacher. We prove the free energy in this case is asymptotically Gaussian and can be approximated by a suitable linear spectral statistics. Unlike the spherical symmetric case the free energy here can not be written as a function of the eigenvalues of the corresponding interaction matrix. The method in this paper relies on a dense sub-graph conditioning technique introduced by Banerjee (2018). The proof of the approximation by the linear spectral statistics part is taken from Banerjee and Ma (2017a).

1.2. Introduction

1.2.1. The model description

We at first give the description of the model. We start with a symmetric matrix $A = (A_{i,j})_{i,j=1}^n$ where the entries in the strict upper triangular part of A are i.i.d. standard Gaussian and for simplicity one might take $A_{i,i} = 0$. The Hamiltonian corresponding to the Sherrington-Kirkpatrick model without any external field is given by

$$H_n^{SK}(\sigma) := \frac{1}{\sqrt{n}}\langle \sigma, A\sigma\rangle = \frac{1}{\sqrt{n}}\sum_{i,j} A_{i,j}\sigma_i\sigma_j = \frac{2}{\sqrt{n}}\sum_{1\le i<j\le n} A_{i,j}\sigma_i\sigma_j. \tag{1.2.1}$$

Here σ_i's are called spins and in this paper we shall only consider the case when $\sigma_i \in \{-1, 1\}$ for each i. In particular, one might consider the case when the spins σ_i's are i.i.d. Rademacher random variables. This is known as the classical Sherrington- Kirkpatrick model. This model has got significant amount interest in the study of spin glasses over the last few decades. Celebrated results like the proof of Parisi formula is considered one of the major advancements in this field. One might look at Panchenko (2013), Talagrand (2006) for some information in this regard.

However the main focus of this paper is the following Hamiltonian

$$H_n(\sigma) := H_n^{SK}(\sigma) + H_n^{CW}(\sigma) \tag{1.2.2}$$

where the Curie-Weiss Hamiltonian with coupling constant J is defined by

$$H_n^{CW}(\sigma) := \frac{J}{n}\sum_{i,j=1}^{n}\sigma_i\sigma_j = \frac{J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2. \tag{1.2.3}$$

Note that the Hamiltonian $H_n^{CW}(\sigma)$ is large in magnitude when all σ_i have the same sign. The Hamiltonian H_n is similar to the SK model with external field,

$$H_n^{\text{ext}}(\sigma) := H_n^{SK}(\sigma) + h\sum_{i=1}^{n}\sigma_i. \tag{1.2.4}$$

The main result of this paper is whenever σ_i's are i.i.d. Rademacher variable we obtain a limit theorem for the free energy corresponding to the Hamiltonian $H_n(\sigma)$ when $\beta < \frac{1}{2}$ and $\beta J < \frac{1}{2}$. If the spins $\sigma = (\sigma_1, \ldots, \sigma_n)$ are distributed according to the uniform measure on the sphere S_{n-1} where $S_{n-1} := \left\{\sigma \in \mathbb{R}^n \mid ||\sigma||^2 = n\right\}$, then the analogous Hamiltonian was considered in Baik and Lee (2017) and Baik et al. (2018). However the results in Baik and Lee (2017) are much more general than the current paper in the sense they are able to consider any $\beta > 0, J > 0$. Depending on the values of β, J, there are three distinct regimes where the free energy shows different behaviors. In particular, the regime $\beta < \frac{1}{2}$ and $\beta J < \frac{1}{2}$ is known as the para-magnetic regime where the result analogous to this paper was obtained in Baik and Lee (2017). The regime when $\beta > \frac{1}{2}$ and $J < 1$ is known as the spin glass regime and the other case ($\beta J > \frac{1}{2}$ and $J > 1$) is known as the ferromagnetic regime. Although the results in Baik and Lee (2017) are much more general than the current paper in terms of possible choices of (β, J), the technique of that paper is restricted to the case when the spins $\sigma = (\sigma_1, \dots, \sigma_n)$ are distributed according to the uniform measure on the sphere S_{n-1} which does not cover the case when σ_i's are i.i.d. Rademacher random variables. This is the problem we consider in this paper.

We now give a very brief overview of the literature for the fluctuation of free energy of classical Sherrington-Kirkpatrick model in presence or absence of an external field.

The classical Sherrington-Kirkpatrick model with no external field ($h = 0$) under goes a phase transition at $\beta = \frac{1}{2}$. When the spins σ_i's are i.i.d. Rademacher and $\beta < \frac{1}{2}$ the free energy has a Gaussian limiting distribution. One might look at Aizenman et al. (1987) and Comets and Neveu (1995) for some references. The case $\beta > \frac{1}{2}$ is known as the low temperature regime. To the best of our limited knowledge, very few things are known about the fluctuations of the free energy in this regime. One might look at Chatterjee (2017) where it is proved that the fluctuation of the free energy of the Sherrington-Kirkpatrick model is at least 0(1). When the spins are uniformly distributed on S_{n-1}, the free energy analogously undergoes a phase transition at $\beta = \frac{1}{2}$. When $\beta < \frac{1}{2}$, the free energy has a Gaussian limiting distribution and can be approximated by a linear spectral statistics of the eigenvalues. The case low temperature case ($\beta > \frac{1}{2}$) is also well-known in this case where the free energy has a limiting GOE Tracy-Widom distribution with $O\left(n^{-\frac{2}{3}}\right)$ fluctuations. One might look at Baik and Lee (2016) for a reference.

1.2.2. Preliminary definitions

We now give some preliminary definitions. We start with defining a Hamiltonian which generalizes the one defined in (1.2.2).

Definition 1. (interactions) Suppose $A_{i,j}$, $1 \leq i \leq j \leq n$ be i.i.d. standard Gaussian random variables. Set $A_{j,i} = A_{i,j}$ for $i < j$. Let $M_{i,j} = \frac{1}{\sqrt{n}} A_{i,j} + \frac{J}{n}$ and $M_{i,i} = \frac{1}{\sqrt{n}} A_{i,i} + \frac{J'}{n}$ for some n independent non negative fixed constants J and J'. One considers the Hamiltonian

$H_n(\sigma) = \langle \sigma, M\sigma \rangle$. The defined Hamiltonian is more general than the one defined in (1.2.2) in the following sense. Here one also allows the random variables $A_{i,i}$ to be standard Gaussian and one also allows J' to be any positive constant.

Given any Hamiltonian $H_n(\sigma)$ one of the most important aspects of it is its free energy. We now define it formally.

Definition 2. (Partition function and Free energy) Given any Hamiltonian $H_n(\sigma)$ where $\sigma = (\sigma_1, \ldots, \sigma_n)$ are distributed according to a measure μ_n, the partition function and free energy at an inverse temperature β is denoted by $Z_n(\beta)$ and $F_n(\beta)$ respectively and defined as follows.

$$Z_n(\beta) := \int \exp\left\{\beta H_n(\sigma)\right\} d\mu_n(\sigma)$$

and

$$F_n(\beta) := \frac{1}{n} \log\left(Z_n(\beta)\right).$$

In our case we take μ_n to be the uniform measure on the Hypercube $\{-1, +1\}^n$.

Definition 3. (Chebyshev Polynomial) We need the definition of Chebyshev Polynomial of first kind of degree m is defined to be a polynomial $S_m(x)$ which takes $\cos(\theta)$ to $\cos(m\theta)$. In particular $S_m(\cos(\theta)) = \cos(m\theta)$. We need a slight variant of this polynomial S_m which is called P_m is defined as

$$P_m(x) = 2S_m(x/2).$$

In particular, one might check that $P_m\left(z + z^{-1}\right) = z^m + z^{-m}$.

Finally we define the Wasserstein distance between two distribution functions.

Definition 4. We at first fix $p \geq 1$. Suppose F^X and F^Y are two distribution functions such that $\int_{x\in\mathbb{R}} |x|^p dF^X(x) < \infty$ and $\int_{x\in\mathbb{R}} |x|^p dF^Y(x) < \infty$. Then the Wasserstein distance for p between F^X and F^Y is is denoted by W_p and defined to be

$$W_p\left(F^X, F^Y\right) := \left[\inf_{X\sim F^X; Y\sim F^Y} \mathrm{E}\left[|X - Y|^p\right]\right]^{\frac{1}{p}}.$$

The following result is well known.

Proposition 1. *Suppose X_n be a sequence of random variables and X be a random variable. Then $X_n \xrightarrow{d} X$ and $\mathrm{E}[X_n^2] \to \mathrm{E}[X]$ if $W_2\left(F^{X_n}, F^X\right) \to 0$.*

One might see Mallows (1972) for a reference.

1.3. Main result

We are ready to state the main result of this paper.

Theorem 1.

1. *(Asymptotic normality) Consider the Hamiltonian $H_n(\sigma)$ as defined in Definition 1. Let $F_n(\beta)$ be the free energy corresponding to the Hamiltonian $H_n(\sigma)$. When $\beta < \frac{1}{2}$ and βJ the following result holds:*

$$n\left(F_n(\beta) - F(\beta)\right) \xrightarrow{d} N(f_1, \alpha_1) \tag{1.3.1}$$

   where $F(\beta) = \beta^2$,

$$\alpha_1 = -\beta^2 - \frac{1}{2}\log\left(1 - 4\beta^2\right)$$

   and

$$f_1 = -\frac{1}{2}\log\left(1 - 2\beta J\right) + \beta(J' - J) - \frac{1}{2}\alpha_1 - \frac{3}{2}\beta^2.$$

2. *(Approximation by signed cycle counts) For any sequence m_n diverging to infinity such that $m_n = o\left(\sqrt{\log n}\right)$, one also has the following approximation result for the log partition function $\log\left(Z_n(\beta)\right)$.*

$$\log\left(Z_n(\beta)\right) + \frac{1}{2}\log\left(1 - 2\beta J\right) - (n-1)\beta^2 + \beta(J - J') - \beta C_{n,1} - \sum_{k=2}^{m_n} \frac{2\mu_k\left(C_{n,k} - (n-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k} \xrightarrow{p} 0 \tag{1.3.2}$$

   with $\mu_k = (2\beta)^k$.

Remark 1. (Approximation of cycles by linear spectral statistics) Let $\tilde{A}$ be the matrix obtained by putting 0 on the diagonal of the matrix A. Let P_k be as defined in Definition 3. Then to following is true for any $3 \leq k = o\left(\sqrt{\log n}\right)$ under $\mathbb{P}_n$.

$$C_{n,k} - \left\{ \mathsf{Tr}\left(P_k\left(\frac{1}{\sqrt{n}}\tilde{A} \right) \right) - \mathrm{E}\left[\mathsf{Tr}\left(P_k\left(\frac{1}{\sqrt{n}}\tilde{A} \right) \right) \right] \right\} \stackrel{p}{\to} 0.$$

Here for any function f and a matrix A

$$\mathsf{Tr}\left[f(A)\right] = \sum_{i=1}^{n} f(\lambda_i)$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the matrix A. The proof is similar to the proof of Theorem 3.4 in Banerjee and Ma (2017a).

1.4. Proof techniques and related definitions

The fundamental technique of the proof of Theorem 1 is completely different from that of Baik and Lee (2017). The proof in the current paper is based on the dense sub graph conditioning technique introduced in Banerjee (2018). The fundamental idea is to view the free energy as the log of the Radon-Nikodym derivative $\left(\log \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)$ of two suitably defined sequences of measures $\mathbb{P}_n$ and $\mathbb{Q}_n$. Now one introduce a class of random variables called the signed cycles (Definition 5) and prove that these variables asymptotically determined the the full Radon-Nikodym derivative. This is done by a fine second moment argument. The argument in this part is highly motivated from a paper by Janson (1995) where it is proved that a similar kind of argument holds for random regular graphs where the signed cycle counts are replaced by normal cycle counts. The technique of cycle conditioning was also used in Mossel et al. (2015) in their proof of contiguity of the probability measures induced by a planted partition model and the Erdős- Rényi model in the sparse regime.

We now start with defining the signed cycles random variables.

Definition 5. Let A be a $n \times n$ symmetric matrix with i.i.d. mean 0 and variance 1. For

$k \geq 2$, we define the signed cycles random variables $C_{n,k}$ as follows:

$$C_{n,k} := \left(\frac{1}{\sqrt{n}}\right)^k \sum_{i_0,i_1,\ldots,i_{k-1}} A_{i_0,i_1} A_{i_1,i_2} \ldots A_{i_{k-1},i_0}.$$

Here $i_0, \ldots, i_{k-1}$ are taken to be all distinct. For $k = 1$, $C_{n,k}$ is simply defined as follows:

$$C_{n,1} := \left(\frac{1}{\sqrt{n}}\right) \sum_i A_{i,i}.$$

In this paper we require the concept of mutual contiguity of two sequence of measures heavily. Now we define these concepts.

Definition 6. (Contiguity) For two sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ defined on σ-fields $(\Omega_n, \mathcal{F}_n)$, we say that $\mathbb{Q}_n$ is contiguous with respect to $\mathbb{P}_n$, denoted by $\mathbb{Q}_n \triangleleft \mathbb{P}_n$, if for any event sequence A_n, $\mathbb{P}_n(A_n) \to 0$ implies $\mathbb{Q}_n(A_n) \to 0$. We say that they are (asymptotically) mutually contiguous, denoted by $\mathbb{P}_n \triangleleft \triangleright \mathbb{Q}_n$, if both $\mathbb{Q}_n \triangleleft \mathbb{P}_n$ and $\mathbb{P}_n \triangleleft \mathbb{Q}_n$ hold.

If someone is interested one might have a look at Le Cam (2012) and Le Cam and Yang (2012) for general discussions on contiguity.

The following result gives an useful way to study mutual contiguity:

Proposition 2. *Suppose that $L_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$, regarded as a random variable on $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, converges in distribution to some random variable L as $n \to \infty$. Then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous if and only if $L > 0$ a.s. and $\mathrm{E}[L] = 1$.*

This result is a direct consequence of so called Le Cam's first lemma. One might look at Le Cam (2012) for a reference.

We now state a result on mutual contiguity of measures.

Proposition 3. *(Janson's second moment method): Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures such that for each n, both are defined on the common σ-algebra $(\Omega_n, \mathcal{F}_n)$. Suppose that for each $i \geq 1$, $W_{n,i}$ are random variables defined on $(\Omega_n, \mathcal{F}_n)$.*

Then the probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are asymptotically mutually contiguous if the following conditions hold simultaneously:

(i) $\mathbb{Q}_n$ *is absolutely continuous with respect to* $\mathbb{P}_n$ *for each* n*;*

(ii) For any fixed $k \geq 1$*, one has* $(W_{n,1}, \ldots, W_{n,k})\,|\mathbb{P}_n \xrightarrow{d} (Z_1, \ldots, Z_k)$ *and* $(W_{n,1}, \ldots, W_{n,k})\,|\mathbb{Q}_n \xrightarrow{d} (Z'_1, \ldots, Z'_k)$.

(iii) $Z_i \sim N(0, \sigma_i^2)$ *and* $Z'_i \sim N(\mu_i, \sigma_i^2)$ *are sequences of independent random variables.*

(iv) The likelihood ratio statistic $Y_n = \frac{\mathrm{d}\mathbb{Q}_n}{\mathrm{d}\mathbb{P}_n}$ *satisfies*

$$\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[Y_n^2\right] \leq \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\} < \infty. \tag{1.4.1}$$

(v) Under $\mathbb{P}_n$*,* $W_{n,i}$*'s are uncorrelated and there exists a sequence* $m_n \to \infty$ *such that*

$$\mathrm{Var}\left[\sum_{i=1}^{m_n} \frac{\mu_i}{\sigma_i^2} W_{n,i}\right] \to C < \infty$$

Here the Var *is considered with respect to the measure* $\mathbb{P}_n$.

In addition, we have that under $\mathbb{P}_n$*,*

$$Y_n \xrightarrow{d} \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i Z_i - \frac{1}{2}\mu_i^2}{\sigma_i^2}\right\}. \tag{1.4.2}$$

Furthermore, given any $\epsilon, \delta > 0$ *there exists a natural number* $K = K(\delta, \epsilon)$ *such that for any sequence* n_l *there is a further subsequence* n_{l_m} *such that*

$$\limsup_{m\to\infty} \mathbb{P}_{n_{l_m}}\left(\left|\log(Y_{n_{l_m}}) - \sum_{k=1}^{K} \frac{2\mu_k W_{n_{l_m},k} - \mu_k^2}{2\sigma_k^2}\right| \geq \epsilon\right) \leq \delta. \tag{1.4.3}$$

Proposition 3 is one of the most important results required for the proof of Theorem 1. In particular, the rest of the proof relies on defining the measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ and $W_{n,i}$'s properly. It is worth noting that in this context the statistics $C_{n,i}$'s serve as $W_{n,i}$'s.

1.5. Construction of $\mathbb{P}_n$ and $\mathbb{Q}_n$ and asymptotic distribution of signed cycles

1.5.1. Construction of the measure $\mathbb{Q}_n$

We at first give the construction of measures $\mathbb{P}_n$ and $\mathbb{Q}_n$.

In this paper $\mathbb{P}_n$ is simply taken to be the measure induced by $(A_{i,j})_{1\le i<j\le n}$. We now define the measure $\mathbb{Q}_n$ in the following way: At first for any given $\sigma \in \{-1,+1\}^n$, we define the measure $\mathbb{Q}_{n,\sigma}$ by

$$\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n} := \exp\left\{\sum_{i<j}\left(\frac{2\beta}{\sqrt{n}}\sigma_i\sigma_j A_{i,j} - \frac{2\beta^2}{n}\right) + \frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2\right\}. \tag{1.5.1}$$

Observe that $\mathbb{Q}_{n,\sigma}$ is not in general a probability measure. In particular,

$$\int_{\Omega_n} d\mathbb{Q}_{n,\sigma} = \exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2\right\}.$$

Finally, we define

$$\mathbb{Q}_n = \frac{1}{\mathrm{E}_{\mu_n}\left[\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2\right\}\right]}\sum_{\sigma\in\{-1,+1\}^n}\frac{1}{2^n}\mathbb{Q}_{n,\sigma}. \tag{1.5.2}$$

Observe that $\mathbb{Q}_n$ is a valid probability measure on Ω_n. We shall prove later that

$$\tau_n := \mathrm{E}_{\mu_n}\left[\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2\right\}\right] \to \frac{1}{\sqrt{1-2\beta J}}.$$

It is worth noting that:

$$
\begin{aligned}
\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} &= \frac{1}{\tau_n} \sum_{\sigma\in\{-1,+1\}^n} \frac{1}{2^n} \exp\left\{ \sum_{i<j} \left(\frac{2\beta}{\sqrt{n}} \sigma_i\sigma_j A_{i,j} - \frac{2\beta^2}{n} \right) + \frac{\beta J}{n} \left(\sum_{i=1}^{n} \sigma_i \right)^2 \right\} \\
&= \frac{1}{\tau_n} \exp\left\{ -(n-1)\beta^2 + \beta J \right\} \exp\left\{ -\frac{\beta}{\sqrt{n}} \sum_{i=1}^{n} A_{i,i} - \beta J' \right\} Z_n(\beta).
\end{aligned}
\tag{1.5.3}
$$

So in order to prove Theorem 1 it is enough to prove a central limit theorem for $\log\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)$ and to prove that $\log\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)$ is asymptotically independent of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_{i,i}$.

1.5.2. Asymptotic distribution of $C_{n,i}$'s under $\mathbb{P}_n$ and $\mathbb{Q}_n$

In order to derive the limiting distribution of $C_{n,i}$'s under $\mathbb{Q}_n$ we at first need to define another sequence of measure $\mathbb{Q}'_n$. We shall at first derive the limiting distribution of $C_{n,i}$'s under $\mathbb{Q}'_n$ and then we shall find the limiting distribution of $C_{n,i}$'s under $\mathbb{Q}_n$.
Let for any given $\sigma \in \{-1,+1\}^n$, $\mathbb{Q}'_{n,\sigma}$ be defined as

$$
\frac{d\mathbb{Q}'_{n,\sigma}}{d\mathbb{P}_n} = \exp\left\{ \sum_{i<j} \left(\frac{2\beta}{\sqrt{n}} \sigma_i\sigma_j A_{i,j} - \frac{2\beta^2}{n} \right) \right\}.
$$

Observe that $\mathbb{Q}'_{n,\sigma}$ is a probability measure. In fact $(A_{i,j})_{1\le i<j\le n}\Big|_{\mathbb{Q}'_{n,\sigma}}$ are independent normal random variables with $A_{i,j}\Big|_{\mathbb{Q}'_{n,\sigma}} \sim N\left(\frac{2\beta}{\sqrt{n}}\sigma_i\sigma_j, 1\right)$. Finally

$$
\mathbb{Q}'_n := \frac{1}{2^n} \sum_{\sigma\in\{-1,1\}^n} \mathbb{Q}'_{n,\sigma}.
$$

The first result in this section gives the asymptotic distribution of $C_{n,i}$'s under $\mathbb{P}_n$ and $\mathbb{Q}_n$.

Proposition 4. *1. Under $\mathbb{P}_n$, we have for any $2 \le k_1 < k_2 \ldots < k_l = o\left(\sqrt{\log(n)}\right)$ with l fixed,*

$$
\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2}}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l}}{\sqrt{2k_l}} \right) \xrightarrow{d} N_l(0, I_l).
$$

2. Let Ψ_n be the uniform probability measure on the hyper cube $\{-1,+1\}^n$. Then there *exists a set S_n with $\Psi_n(S_n) \to 0$, we have for all $\sigma \in S_n^c$, under $\mathbb{Q}'_{n,\sigma}$*

$$\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \dots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l)$$

where $\mu_i := (2\beta)^i$. This implies under $\mathbb{Q}'_n$,

$$\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \dots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l).$$

3. *Finally, $C_{n,1} \xrightarrow{d} N(0,1)$ under $\mathbb{P}_n$ and is asymptotically independent of the process $\{C_{n,k} - (n-1)\mathbb{I}_{k=2}\}_{k\geq 2}$.*

The proof of Proposition 4 is similar to the proof of Proposition 4.1 of Banerjee (2018). We omit the details. With Proposition 4, we now give the asymptotic distribution of $C_{n,i}$'s under $\mathbb{Q}_n$.

Proposition 5. *Under $\mathbb{Q}_n$, we have for any $2 \leq k_1 < k_2 \ldots < k_l = o\left(\sqrt{\log(n)}\right)$ with l fixed,*

$$\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \dots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l).$$

Proof. We assume Proposition 4 and give the proof. We need to prove for any bounded continuous function $f : \mathbb{R}^l \to \mathbb{R}$,

$$\int f\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \dots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) d\mathbb{Q}_n \to \mathrm{E}\left[f(Z_{k_1}, \dots, Z_{k_l})\right]$$

where $Z_{k_1}, \ldots, Z_{k_l}$ are independent standard Gaussian random variables. Now

$$\begin{aligned}
&\int_{\Omega_n} f\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) d\mathbb{Q}_n \\
&= \frac{1}{2^n} \sum_{\sigma \in \{-1,+1\}^n} \int_{\Omega_n} f\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) d\mathbb{Q}_{n,\sigma} \\
&= \frac{1}{2^n} \sum_{\sigma \in \{-1,+1\}^n} \int_{\Omega_n} f\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) \frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n} d\mathbb{P}_n \\
&= \frac{1}{\tau_n}\frac{1}{2^n} \sum_{\sigma \in \{-1,+1\}^n} \int_{\Omega_n} f\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) \exp\left\{\frac{\beta J}{n}\left(\sum \sigma_i\right)^2\right\} \frac{d\mathbb{Q}'_{n,\sigma}}{d\mathbb{P}_n} d\mathbb{P}_n \\
&= \frac{1}{\tau_n}\frac{1}{2^n} \sum_{\sigma \in \{-1,+1\}^n} \exp\left\{\frac{\beta J}{n}\left(\sum \sigma_i\right)^2\right\} F(\sigma) \\
&= \frac{1}{\tau_n} \mathrm{E}_{\Psi_n}\left[\exp\left\{\frac{\beta J}{n}\left(\sum \sigma_i\right)^2\right\} F(\sigma)\right]
\end{aligned} \tag{1.5.4}$$

Here $F(\sigma) = \int_{\Omega_n} f\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) \frac{d\mathbb{Q}'_{n,\sigma}}{d\mathbb{P}_n} d\mathbb{P}_n$. From Proposition 4, we know that under the measure $\Psi_n(\cdot)$, $F(\sigma) \xrightarrow{p} \mathrm{E}\left[f(Z_{k_1}, \ldots, Z_{k_l})\right]$. Now from central limit theorem,

$$\frac{1}{n}\left(\sum \sigma_i\right)^2 \xrightarrow{d} Y$$

where Y is a Chi-squared random variable with 1 degree of freedom. So by Slutsky's theorem we have under the measure Ψ_n

$$F(\sigma) \exp\left\{\frac{\beta J}{n}\left(\sum \sigma_i\right)^2\right\} \xrightarrow{d} \mathrm{E}\left[f(Z_{k_1}, \ldots, Z_{k_l})\right] \exp\left\{\beta J Y\right\}.$$

Further, from Hoeffding's inequality we also have when $\beta J < \frac{1}{2}$, the sequence $\exp\left\{\frac{\beta J}{n}\left(\sum \sigma_i\right)^2\right\}$ is uniformly integrable. Since the random variables $F(\sigma)$'s are uniformly bounded, the sequence $F(\sigma)\exp\left\{\frac{\beta J}{n}\left(\sum \sigma_i\right)^2\right\}$ is also uniformly integrable. As a consequence,

$$\mathrm{E}_{\Psi_n}\left[\exp\left\{\frac{\beta J}{n}\left(\sum \sigma_i\right)^2\right\} F(\sigma)\right] \to \mathrm{E}\left[f(Z_{k_1}, \ldots, Z_{k_l})\right] \frac{1}{\sqrt{1 - 2\beta J}}.$$

□

1.6. Proof of Theorem 1

As mentioned in subsection 1.5.1, we at first prove a central limit theorem for $\log\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)|\mathbb{P}_n$ and finally proving $\log\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)|\mathbb{P}_n$ is asymptotically independent of $C_{n,1}$. The main idea is to use Proposition 3 to a class of measure $\tilde{\mathbb{Q}}_n$ which is close to $\mathbb{Q}_n$ in total variation distance. We now give a formal proof of Theorem 1.

Proof of Theorem 1:

We at first prove the central limit theorem for $\log\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)|\mathbb{P}_n$. The proof is broken into two steps as follows.

Step 1 (Construction of the measure $\tilde{\mathbb{Q}}_n$) : To begin with we shall consider a set $\Omega(\sigma)_n \subset \{-1,+1\}^n$ such that $\Psi_n\left(\Omega(\sigma)_n\right) \to 1$. The precise definition of $\Omega(\sigma)_n$ will be provided later. Now we consider the measure $\tilde{\mathbb{Q}}_n$ as follows

$$\tilde{\mathbb{Q}}_n = \frac{1}{\mathrm{E}_{\Psi_n}\left[\mathbb{I}_{\Omega(\sigma)_n}\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}\right]}\sum_{\sigma\in\Omega(\sigma)_n}\frac{1}{2^n}\mathbb{Q}_{n,\sigma} = \frac{1}{\tilde{\tau}_n}\sum_{\sigma\in\Omega(\sigma)_n}\frac{1}{2^n}\mathbb{Q}_{n,\sigma}$$

where we define

$$\tilde{\tau}_n := \mathrm{E}_{\Psi_n}\left[\mathbb{I}_{\Omega(\sigma)_n}\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}\right].$$

Since the sequence of random variables $\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}$ is uniformly integrable it follows that for any sequence of sets $\Omega_n(\sigma)$ such that $\Psi_n\left[\Omega_n(\sigma)\right] \to 1$,

$$\mathrm{E}_{\Psi_n}\left[\mathbb{I}_{\Omega(\sigma)_n}\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}\right] \to \frac{1}{\sqrt{1-2\beta J}}.$$

Now we prove the sequences of measures $\mathbb{Q}_n$ and $\tilde{\mathbb{Q}}_n$ are close in the total variation sense.

Let $A_n \in \mathcal{F}_n$ be a sequence of measurable sets. We have

$$\begin{aligned}
&\left|\mathbb{Q}_n(A_n) - \tilde{\mathbb{Q}}_n(A_n)\right| \\
&= \left|\frac{1}{\tau_n}\sum_{\sigma\in\{-1,+1\}^n}\frac{1}{2^n}\int_{A_n}\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}d\mathbb{P}_n - \frac{1}{\tilde{\tau}_n}\sum_{\sigma\in\Omega_n(\sigma)}\frac{1}{2^n}\int_{A_n}\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}d\mathbb{P}_n\right| \\
&\leq \left|\frac{1}{\tau_n}\sum_{\sigma\in\Omega_n(\sigma)^c}\frac{1}{2^n}\int_{A_n}\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}d\mathbb{P}_n\right| + \left|\left(\frac{1}{\tau_n} - \frac{1}{\tilde{\tau}_n}\right)\sum_{\sigma\in\Omega_n(\sigma)}\frac{1}{2^n}\int_{A_n}\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}d\mathbb{P}_n\right| \\
&\leq \left|\frac{1}{\tau_n}\sum_{\sigma\in\Omega_n(\sigma)^c}\frac{1}{2^n}\int_{\Omega_n}\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}d\mathbb{P}_n\right| + \left|\left(\frac{1}{\tau_n} - \frac{1}{\tilde{\tau}_n}\right)\right|\left|\sum_{\sigma\in\Omega_n(\sigma)}\frac{1}{2^n}\int_{\Omega_n}\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}d\mathbb{P}_n\right| \\
&\leq \left|\frac{1}{\tau_n}\mathrm{E}_{\Psi_n}\left[\mathbb{I}_{\Omega(\sigma)_n^c}\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^n\sigma_i\right)^2\right\}\right]\right| + \left|\left(\frac{1}{\tau_n} - \frac{1}{\tilde{\tau}_n}\right)\right|\mathrm{E}_{\Psi_n}\left[\mathbb{I}_{\Omega(\sigma)_n}\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^n\sigma_i\right)^2\right\}\right]
\end{aligned} \tag{1.6.1}$$

Observe that the final expression in (1.6.1) does not depend on the set A_n and also it has been argued earlier that the final expression in (1.6.1) converges to 0. As a consequence, by Proposition 5 under the measure $\tilde{\mathbb{Q}}_n$ the random variables for any $2 \leq k_1 < k_2 \ldots < k_l = o\left(\sqrt{\log(n)}\right)$ with l fixed,

$$\left(\frac{C_{n,k_1} - (n-1)\mathbb{I}_{k_1=2} - \mu_{k_1}}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l).$$

Now we prove that $\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[\left(\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)^2\right] \leq \exp\left\{\sum_{k=2}^{\infty}\frac{\mu_k^2}{\sigma_k^2}\right\}$ where $\mu_k = (2\beta)^k$. This will allow us to use Proposition 3 for $\tilde{\mathbb{Q}}_n$. In particular, we shall get $\left(\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)|\mathbb{P}_n$ has a normal limiting distribution. Once this is done, the limiting distribution of $\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}|\mathbb{P}_n$ can be derived by the following arguments which proves

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} - \frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}|\mathbb{P}_n \xrightarrow{p} 0.$$

Since both τ_n and $\tilde{\tau}_n$ have the same finite limit, the random variable

$$\tilde{Y}_n := \frac{\tilde{\tau}_n}{\tau_n}\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\,|\mathbb{P}_n$$

has the same limiting distribution as $\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\,|\mathbb{P}_n$. In particular,

$$\left(\tilde{Y}_n - \frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)|\mathbb{P}_n \xrightarrow{p} 0.$$

So it is enough to prove

$$\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} - \tilde{Y}_n\right)|\mathbb{P}_n \xrightarrow{p} 0.$$

However,

$$\begin{aligned} 0 \le \frac{d\mathbb{Q}_n}{d\mathbb{P}_n} - \tilde{Y}_n &= \frac{1}{\tau_n}\left(\sum_{\sigma\in\Omega_n(\sigma)^c}\frac{1}{2^n}\frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}\right) \\ \Rightarrow \mathrm{E}_{\mathbb{P}_n}\left[\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} - \tilde{Y}_n\right] &= \frac{1}{\tau_n}\mathrm{E}_{\Psi_n}\left[\mathbb{I}_{\Omega_n(\sigma)^c}\exp\left\{\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}\right] \to 0. \end{aligned} \tag{1.6.2}$$

This completes the proof of

$$\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} - \frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)|\mathbb{P}_n \xrightarrow{p} 0.$$

Step 2 $\left(\textbf{Upper bounding } \mathrm{E}_{\mathbb{P}_n}\left[\left(\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)^2\right]\right)$**:**

We know that

$$\begin{aligned}
\left(\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)^2 &= \left(\frac{1}{\tilde{\tau_n}}\right)^2 \frac{1}{4^n} \sum_{\sigma\in\Omega(\sigma)_n} \sum_{\sigma'\in\Omega(\sigma)_n} \frac{d\mathbb{Q}_{n,\sigma}}{d\mathbb{P}_n}\frac{d\mathbb{Q}_{n,\sigma'}}{d\mathbb{P}_n} \\
&= \left(\frac{1}{\tilde{\tau_n}}\right)^2 \frac{1}{4^n} \sum_{\sigma\in\Omega(\sigma)_n} \sum_{\sigma'\in\Omega(\sigma)_n} \exp\Bigg\{\sum_{i<j}\left(\frac{2\beta}{\sqrt{n}}A_{i,j}\left(\sigma_i\sigma_j+\sigma'_i\sigma'_j\right)-\frac{4\beta^2}{n}\right) \\
&\qquad + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2 + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma'_i\right)^2\Bigg\} \\
\Rightarrow \mathrm{E}_{\mathbb{P}_n}&\left[\left(\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)^2\right] \\
&= \left(\frac{1}{\tilde{\tau_n}}\right)^2 \frac{1}{4^n} \sum_{\sigma\in\Omega(\sigma)_n} \sum_{\sigma'\in\Omega(\sigma)_n} \exp\Bigg\{\sum_{i<j}\left(\frac{2\beta^2}{n}\left(\sigma_i\sigma_j+\sigma'_i\sigma'_j\right)^2-\frac{4\beta^2}{n}\right) \\
&\qquad + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2 + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma'_i\right)^2\Bigg\} \\
&= \left(\frac{1}{\tilde{\tau_n}}\right)^2 \frac{1}{4^n} \sum_{\sigma\in\Omega(\sigma)_n} \sum_{\sigma'\in\Omega(\sigma)_n} \exp\Bigg\{\sum_{i<j}\left(\frac{4\beta^2}{n}\sigma_i\sigma_j\sigma'_i\sigma'_j\right) + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2 \\
&\qquad + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma'_i\right)^2\Bigg\} \\
&= \left(\frac{1}{\tilde{\tau_n}}\right)^2 \frac{1}{4^n} \sum_{\sigma\in\Omega(\sigma)_n} \sum_{\sigma'\in\Omega(\sigma)_n} \exp\Bigg\{\frac{2\beta^2}{n}\left(\sum_{i=1}^n \sigma_i\sigma'_i\right)^2 - 2\beta^2 \\
&\qquad + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2 + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma'_i\right)^2\Bigg\} \\
&= \exp\left\{-2\beta^2\right\}\left(\frac{1}{\tilde{\tau_n}}\right)^2 \mathrm{E}_{\Psi_n\otimes\Psi_n}\Bigg[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\mathbb{I}_{\sigma'\in\Omega(\sigma)_n}\exp\Bigg\{\frac{2\beta^2}{n}\left(\sum_{i=1}^n \sigma_i\sigma'_i\right)^2 + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2 \\
&\qquad\qquad + \frac{\beta J}{n}\left(\sum_{i=1}^n \sigma'_i\right)^2\Bigg\}\Bigg]
\end{aligned} \tag{1.6.3}$$

Here $\Psi_n \otimes \Psi_n$ denote the two fold product of the uniform probability measure on $\{-1,1\}^n \times \{-1,1\}^n$.

Observe that the random variable

$$\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\mathbb{I}_{\sigma'\in\Omega(\sigma)_n}\exp\left\{\frac{2\beta^2}{n}\left(\sum_{i=1}^n\sigma_i\sigma_i'\right)^2+\frac{\beta J}{n}\left(\sum_{i=1}^n\sigma_i\right)^2+\frac{\beta J}{n}\left(\sum_{i=1}^n\sigma_i'\right)^2\right\}\xrightarrow{d} \exp\left\{2\beta^2Y_1+\beta JY_2+\beta JY_3\right\} \tag{1.6.4}$$

where Y_1, Y_2, Y_3 are three independent chi-square random variables each with one degree of freedom. Our target is to prove the random variable in the L.S. of (1.6.4) is uniformly integrable. This done by proving

$$\limsup_{n\to\infty}\mathrm{E}_{\Psi_n\otimes\Psi_n}\left[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\mathbb{I}_{\sigma'\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\left(\frac{2\beta^2}{n}\left(\sum_{i=1}^n\sigma_i\sigma_i'\right)^2 +\frac{\beta J}{n}\left(\sum_{i=1}^n\sigma_i\right)^2+\frac{\beta J}{n}\left(\sum_{i=1}^n\sigma_i'\right)^2\right\}\right)\right]<\infty$$

for sufficiently small η. We at first write

$$
\begin{aligned}
&= \mathrm{E}_{\Psi_n \otimes \Psi_n}\left[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\mathbb{I}_{\sigma'\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\left(\frac{2\beta^2}{n}\left(\sum_{i=1}^n \sigma_i\sigma_i'\right)^2+\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2+\right.\right.\right.\\
&\qquad\left.\left.\left.\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i'\right)^2\right)\right\}\right]\\
&= \mathrm{E}\left[\mathrm{E}\left[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\mathbb{I}_{\sigma'\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\left(\frac{2\beta^2}{n}\left(\sum_{i=1}^n \sigma_i\sigma_i'\right)^2+\right.\right.\right.\right.\\
&\left.\left.\left.\left.\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2+\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i'\right)^2\right)\right\}|\sigma\right]\right]\\
&= \mathrm{E}\left[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}\right.\\
&\qquad\qquad\left.\mathrm{E}\left[\mathbb{I}_{\sigma'\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\left(\frac{2\beta^2}{n}\left(\sum_{i=1}^n \sigma_i\sigma_i'\right)^2+\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i'\right)^2\right)\right\}|\sigma\right]\right]\\
&= \mathrm{E}\left[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}\right.\\
&\quad\left.\mathrm{E}\left[\mathbb{I}_{\sigma'\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\frac{1}{n}\left(\sigma'\right)^{\mathrm{T}}A'A\left(\sigma'\right)\right\}|\sigma\right]\right]\\
&\leq \mathrm{E}\left[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\exp\left\{(1+\eta)\frac{\beta J}{n}\left(\sum_{i=1}^n \sigma_i\right)^2\right\}\mathrm{E}\left[\exp\left\{(1+\eta)\frac{1}{n}\left(\sigma'\right)^{\mathrm{T}}A^{\mathrm{T}}A\left(\sigma'\right)\right\}|\sigma\right]\right].
\end{aligned}
\tag{1.6.5}
$$

Here T denotes the transpose of a matrix and the matrix $A_{2\times n}$ is given by

$$
A = \begin{pmatrix} \beta J & \beta J & \ldots & \beta J \\ 2\beta^2\sigma_1 & 2\beta^2\sigma_2 & \ldots & 2\beta^2\sigma_n \end{pmatrix}. \tag{1.6.6}
$$

Since $\mathrm{E}\left[\exp\left\{\alpha^{\mathrm{T}}\sigma'\right\}\right] \leq \exp\left\{\frac{1}{2}||\alpha||^2\right\}$ for any $\alpha \in \mathbb{R}^n$, we have the following tail estimate by Theorem 1 and Remark 1 of Hsu et al. (2012):

$$
\mathbb{P}\left[\frac{1}{n}\left(\sigma'\right)^{\mathrm{T}}A^{\mathrm{T}}A\left(\sigma'\right) \geq \mathrm{tr}(\Sigma) + 2\sqrt{\mathrm{tr}(\Sigma^2)t} + 2||\Sigma||t\,|\sigma\right] \leq e^{-t}
$$

where $\Sigma = \frac{1}{\sqrt{n}} A$. Observe that the nonzero eigenvalues of Σ is same as the nonzero eigenvalues of $\frac{1}{n} AA^{\mathrm{T}}$. Now

$$\frac{1}{n} AA^{\mathrm{T}} = \begin{pmatrix} \beta J & 2\beta^3 J \left(\frac{1}{n}\sum_{i=1}^{n} \sigma_i\right) \\ 2\beta^3 J \left(\frac{1}{n}\sum_{i=1}^{n} \sigma_i\right) & 2\beta^2 \end{pmatrix}.$$

We now choose the set

$$\Omega(\sigma)_n := \left\{ \frac{1}{n} \sum_{i=1}^{n} \sigma_i \leq \delta_n \right\}$$

for some $\delta_n \to 0$ as $n \to \infty$. The existence of such $\Omega(\sigma)_n$ is ensured by weak law of large numbers. Now by Weyl's interlacing inequality, we have the eigenvalue of $\frac{1}{n} AA^{\mathrm{T}}$ are given by $\left\{\beta J + O(\delta_n), 2\beta^2 + O(\delta_n)\right\}$. Also note that on $\Omega(\sigma)_n$, $\mathrm{tr}(\Sigma)$ and $\mathrm{tr}(\Sigma^2)$ remain uniformly bounded. So given any $\epsilon > 0$ we can find a t_0 large enough such that

$$\mathrm{tr}(\Sigma) + 2\sqrt{\mathrm{tr}(\Sigma^2) t} < \epsilon 2 ||\Sigma|| t$$

for all $t > t_0$. As a consequence, for all $t > t_0$

$$\begin{aligned} &\mathbb{P}\left[\frac{1}{n} \left(\sigma'\right)^{\mathrm{T}} A^{\mathrm{T}} A \left(\sigma'\right) \geq (1+\epsilon) 2 ||\Sigma|| t \,|\sigma \right] \\ &\leq \mathbb{P}\left[\frac{1}{n} \left(\sigma'\right)^{\mathrm{T}} A^{\mathrm{T}} A \left(\sigma'\right) \geq \mathrm{tr}(\Sigma) + 2\sqrt{\mathrm{tr}(\Sigma^2) t} + 2||\Sigma|| t \,|\sigma \right] < e^{-t} \\ &\Rightarrow \mathbb{P}\left[(1+\eta) \frac{1}{n} \left(\sigma'\right)^{\mathrm{T}} A^{\mathrm{T}} A \left(\sigma'\right) \geq \log(t) \right] \leq t^{\frac{-1}{2(1+\epsilon)(1+\eta)||\Sigma||}} \;\forall\; t > \tilde{t}_0. \end{aligned} \tag{1.6.7}$$

where $\tilde{t}_0$ is another deterministic constant. Since $\max\left\{\beta J, 2\beta^2\right\} < \frac{1}{2}$, we can choose ϵ and η small enough such that

$$\frac{1}{2(1+\epsilon)(1+\eta)||\Sigma||} > \alpha_0 > 1.$$

As a consequence,

$$\mathbb{I}_{\sigma \in \Omega(\sigma)_n} \mathrm{E}\left[\exp\left\{ (1+\eta) \frac{1}{n} \left(\sigma'\right)^{\mathrm{T}} A^{\mathrm{T}} A \left(\sigma'\right) \right\} |\sigma \right] \leq \tilde{t}_0 + \int_{t > \tilde{t}_0} \frac{1}{t^{\alpha_0}} dt = \tilde{t}_0 + \frac{1}{\alpha_0 - 1} \frac{1}{t^{\alpha_0 - 1}} \tag{1.6.8}$$

On the other hand we can choose η small enough such that $\beta J(1+\eta) < \gamma_0 < \frac{1}{2}$. Now it is enough to prove that

$$\limsup \mathrm{E}\left[\exp\left\{(1+\eta)\frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2\right\}\right] < \infty. \tag{1.6.9}$$

However we know that for any $t > 0$,

$$\begin{aligned}
&\mathrm{E}\left[\exp\left\{\frac{t}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i\right\}\right] \leq \exp\left\{\frac{t^2}{2}\right\}\\
&\Rightarrow \mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i\right| > t\right] = 2\mathbb{P}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i > t\right] = 2\mathbb{P}\left[\exp\left\{\frac{t}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i\right\} > \exp\left\{t^2\right\}\right]\\
&\leq 2\exp\left\{-\frac{t^2}{2}\right\}
\end{aligned} \tag{1.6.10}$$

Here the last inequality is a straight forward application of Markov's inequality. Now

$$\begin{aligned}
&\mathbb{P}\left[\exp\left\{\frac{\beta J(1+\eta)}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2\right\} > t\right] = \mathbb{P}\left[\frac{\beta J(1+\eta)}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2 > \log(t)\right]\\
&\quad = \mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sigma_i\right| > \sqrt{\frac{\log t}{\beta J(1+\eta)}}\right] \leq 2\exp\left\{-\frac{\log t}{2\beta J(1+\eta)}\right\} \leq 2\left(\frac{1}{t}\right)^{\frac{1}{2\beta J(1+\eta)}} < 2\left(\frac{1}{t}\right)^{\frac{1}{2\gamma_0}}.
\end{aligned} \tag{1.6.11}$$

Observe that $\frac{1}{2\gamma_0} > 1$. Hence by argument similar to (1.6.8) we have

$$\limsup \mathrm{E}\left[\exp\left\{(1+\eta)\frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2\right\}\right] < \infty.$$

This completes the proof of uniform integrability of the random variable in the L.S. of

(1.6.4). As a consequence,

$$\lim_{n\to\infty} \mathrm{E}\left[\mathbb{I}_{\sigma\in\Omega(\sigma)_n}\mathbb{I}_{\sigma'\in\Omega(\sigma)_n} \exp\left\{\frac{2\beta^2}{n}\left(\sum_{i=1}^{n}\sigma_i\sigma_i'\right)^2 + \frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i\right)^2 + \frac{\beta J}{n}\left(\sum_{i=1}^{n}\sigma_i'\right)^2\right\}\right]$$
$$= \mathrm{E}\left[2\beta^2 Y_1 + \beta J Y_2 + \beta J Y_3\right] = \frac{1}{\sqrt{1-4\beta^2}}\frac{1}{1-2\beta J}. \tag{1.6.12}$$

Plugging this into (1.6.3) we have

$$\begin{aligned}\lim_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[\left(\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)^2\right] &= \exp\left\{-2\beta^2\right\}(1-2\beta J)\frac{1}{\sqrt{1-4\beta^2}}\frac{1}{1-2\beta J} \\ &= \exp\left\{-2\beta^2\right\}\frac{1}{\sqrt{1-4\beta^2}} \\ &= \exp\left\{-2\beta^2\right\}\exp\left\{-\frac{1}{2}\log\left(1-4\beta^2\right)\right\} \\ &= \exp\left\{-2\beta^2\right\}\exp\left\{\frac{1}{2}\sum_{k=1}^{\infty}\frac{\left(4\beta^2\right)^k}{k}\right\} = \exp\left\{\sum_{k=2}^{\infty}\frac{\mu_k^2}{2k}\right\}\end{aligned} \tag{1.6.13}$$

where $\mu_k = (2\beta)^k$. Now using Proposition 3 with $W_{n,k} = C_{n,k+1} - (n-1)\mathbb{I}_{k=1}$, we have for the sequences of measures $\tilde{\mathbb{Q}}_n$ and $\mathbb{P}_n$

$$\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}|\mathbb{P}_n \xrightarrow{d} \exp\left\{\sum_{k=1}^{\infty}\frac{2\mu_{k+1}Z_k - \mu_{k+1}^2}{4(k+1)}\right\} \tag{1.6.14}$$

where $Z_k \sim N\left(0, 2(k+1)\right)$. Hence

$$\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}|\mathbb{P}_n \xrightarrow{d} \exp\left\{\sum_{k=1}^{\infty}\frac{2\mu_{k+1}Z_k - \mu_{k+1}^2}{4(k+1)}\right\}.$$

This completes the proof of the asymptotic normality of $\log\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)|\mathbb{P}_n$.

Proof of part (2) of Theorem 1: Before proving part (1) of Theorem 1, we prove part

(2). Since

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} = \frac{1}{\tau_n} \exp\left\{-(n-1)\beta^2 + \beta J\right\} \exp\left\{-\frac{\beta}{\sqrt{n}} \sum_{i=1}^{n} A_{i,i} - \beta J'\right\} Z_n(\beta),$$

in order to prove part (2) of Theorem 1, we need to prove that

$$\log\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right) - \sum_{k=2}^{m_n} \frac{2\mu_k\left(C_{n,k} - (n-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k} |\mathbb{P}_n \xrightarrow{p} 0. \tag{1.6.15}$$

We at first prove the result analogous to (1.6.15) for $\log\left(\frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n}\right)$. (1.6.15) then follows from the fact that $\frac{d\mathbb{Q}_n}{d\mathbb{P}_n} - \frac{d\tilde{\mathbb{Q}}_n}{d\mathbb{P}_n} |\mathbb{P}_n \xrightarrow{p} 0$ and an application of continuous mapping theorem.

By (1.4.3), for any given $\epsilon, \delta > 0$ there exists $K = K(\epsilon, \delta)$ and for any subsequence n_l there exists a further subsequence n_{l_q} such that

$$\mathbb{P}_{n_{l_q}}\left(\left|\log(\frac{d\tilde{\mathbb{Q}}_{n_{l_q}}}{d\mathbb{P}_{n_{l_q}}}) - \sum_{k=2}^{K} \frac{2\mu_k(C_{n_{l_q},k} - (n-1)\mathbb{I}_{k=2}) - \mu_k^2}{4k}\right| \geq \frac{\epsilon}{2}\right) \leq \frac{\delta}{2}. \tag{1.6.16}$$

Now choose $K' \geq K$ such that

$$\sum_{K'+1}^{\infty} \frac{\mu_k^2}{2k} \leq \max\left\{\frac{\delta\epsilon^2}{100}, \frac{\epsilon}{100}\right\}.$$

For any $K' < k_1 < k_2 < m_n = o(\sqrt{\log n})$, the proof of Proposition 4 implies that $\mathrm{E}_{\mathbb{P}_n}\left[C_{n,k_1}\right] = 0$, $\mathrm{Cov}(C_{n,k_1}, C_{n,k_2}) = 0$ and $\mathrm{Var}(C_{n,k_i}) = 2k_i(1 + O(k_i^2/n))$ for $i \in \{1, 2\}$. So

$$\mathrm{Var}\left(\sum_{k=K'+1}^{m_{n_{l_q}}} \frac{2\mu_k C_{n_{l_q},k} - \mu_k^2}{4k}\right) = (1 + o(1)) \sum_{k=K'+1}^{m_{n_{l_q}}} \frac{\mu_k^2}{2k} \leq \frac{\delta\epsilon^2}{100}.$$

Now for large values of n_{l_q},

$$\begin{aligned}
&\mathbb{P}_{n_{l_q}}\left(\left|\sum_{k=K+1}^{m_{n_{l_q}}}\frac{2\mu_k C_{n_{l_q},k}}{4k}\right|\geq\frac{\epsilon}{4}\right)\leq\frac{16\delta\epsilon^2}{100\epsilon^2}, \qquad \text{and so}\\
&\mathbb{P}_{n_{l_q}}\left(\left|\sum_{k=K+1}^{m_{n_{l_q}}}\frac{2\mu_k C_{n_{l_q},k}-\mu_k^2}{4k}\right|\geq\frac{\epsilon}{4}+\frac{\epsilon}{100}\right)\leq\frac{16\delta\epsilon^2}{100\epsilon^2}.
\end{aligned} \tag{1.6.17}$$

Plugging in the estimates of (1.6.16) and (1.6.17) we have for all large values of n_{l_q},

$$\mathbb{P}_{n_{l_q}}\left(\left|\log\left(\frac{d\tilde{\mathbb{Q}}_{n_{l_q}}}{d\mathbb{P}_{n_{l_q}}}\right)-\sum_{k=1}^{m_{n_{l_q}}}\frac{2\mu_k(C_{n_{l_q},k}-(n-1)\mathbb{I}_{k=2})-\mu_k^2}{4k}\right|\geq\epsilon\right)\leq\delta. \tag{1.6.18}$$

Since (1.6.18) occurs to any subsequence and any (ϵ,δ) pair, this completes the proof.

Proof of part (1) of Theorem 1: Consider the random variable

$$M:=W+\sum_{k=1}^{\infty}\frac{2\mu_{k+1}Z_k-\mu_{k+1}^2}{4(k+1)}$$

where $W\sim N(0,\beta^2)$ and is independent of the random variable

$$\sum_{k=1}^{\infty}\frac{2\mu_{k+1}Z_k-\mu_{k+1}^2}{4(k+1)}.$$

Observe that from the proof of part (2) we have

$$\begin{aligned}
&\log\left(Z_n(\beta)\right)+\frac{1}{2}\log\left(1-2\beta J\right)-(n-1)\beta^2+\beta(J-J')-\beta C_{n,1}\\
&\quad-\sum_{k=2}^{m_n}\frac{2\mu_k\left(C_{n,k}-(n-1)\mathbb{I}_{k=2}\right)-\mu_k^2}{4k}\,|\mathbb{P}_n\xrightarrow{p}0.
\end{aligned} \tag{1.6.19}$$

So it is enough to prove that

$$\beta C_{n,1}+\sum_{k=2}^{m_n}\frac{2\mu_k\left(C_{n,k}-(n-1)\mathbb{I}_{k=2}\right)-\mu_k^2}{4k}\xrightarrow{d}N\left(\beta^2+\frac{1}{4}\log(1-4\beta^2),-\beta^2-\frac{1}{2}\log(1-4\beta^2)\right).$$

On the other hand for any fixed K,

$$\beta C_{n,1} + \sum_{k=2}^{K} \frac{2\mu_k\left(C_{n,k} - (n-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k} |\mathbb{P}_n \xrightarrow{d} W + \sum_{k=1}^{K-1} \frac{2\mu_{k+1}Z_k - \mu_{k+1}^2}{4(k+1)}.$$

Since all the random variables $\beta C_{n,1}$, $\sum_{k=2}^{m_n} \frac{2\mu_k\left(C_{n,k}-(n-1)\mathbb{I}_{k=2}\right)-\mu_k^2}{4k}$ and $\sum_{k=2}^{K} \frac{2\mu_k\left(C_{n,k}-(n-1)\mathbb{I}_{k=2}\right)-\mu_k^2}{4k}$ are tight, we have any of their linear combination is also tight. Hence given any subsequence n_l there exists a further subsequence n_{l_q} such that

$$\beta C_{n_{l_q},1} + \sum_{k=2}^{m_{n_{l_q}}} \frac{2\mu_k\left(C_{n_{l_q},k} - (n_{l_q}-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k} \Big|\mathbb{P}_{n_{l_q}} \xrightarrow{d} M\{n_{l_q}\}.$$

On the other hand for every fixed K there is a further subsequence $n_{l_{q_m}}$ (possibly dependent on K) such that

$$\left(\beta C_{n_{l_{q_m}},1} + \sum_{k=2}^{m_{n_{l_{q_m}}}} \frac{2\mu_k\left(C_{n_{l_{q_m}},k} - (n_{l_{q_m}}-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k}, \beta C_{n_{l_{q_m}},1} + \right.$$
$$\left. \sum_{k=2}^{K} \frac{2\mu_k\left(C_{n_{l_{q_m}},k} - (n_{l_{q_m}}-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k}\right) \Big|\mathbb{P}_{n_{l_{q_m}}}$$
$$\xrightarrow{d} \left(M_1, M_{2,K}\right). \tag{1.6.20}$$

where $M_1 \stackrel{d}{=} M\{n_{l_q}\}$ and $M_{2,K} \stackrel{d}{=} W + \sum_{k=1}^{K-1} \frac{2\mu_{k+1}Z_k - \mu_{k+1}^2}{4(k+1)}$. Hence

$$\sum_{k=K+1}^{m_{n_{l_{q_m}}}} \frac{2\mu_k\left(C_{n_{l_{q_m}},k} - (n_{l_{q_m}}-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k} \Big|\mathbb{P}_{n_{l_{q_m}}} \xrightarrow{d} M_1 - M_{2,K}.$$

On the other hand by Fatou's lemma for in distributional convergence

$$\liminf \mathrm{E}_{\mathbb{P}_{n_{l_{q_m}}}} \left[\left(\sum_{k=K+1}^{m_n} \frac{2\mu_k\left(C_{n_{l_{q_m}},k} - (n-1)\mathbb{I}_{k=2}\right) - \mu_k^2}{4k}\right)^2\right] \geq \mathrm{E}\left[(M_1 - M_{2,K})^2\right]. \tag{1.6.21}$$

We know that for large enough value of $n_{l_{q_m}}$,

$$\begin{aligned}
&\mathrm{E}_{\mathbb{P}_{n_{l_{q_m}}}}\left[\left(\sum_{k=K+1}^{m_{n_{l_{q_m}}}} \frac{2\mu_k\left(C_{n_{l_{q_m}},k}-(n-1)\mathbb{I}_{k=2}\right)-\mu_k^2}{4k}\right)^2\right] \\
&= \mathrm{Var}\left(\sum_{k=K+1}^{m_{n_{l_{q_m}}}} \frac{2\mu_k\left(C_{n_{l_{q_m}},k}-(n-1)\mathbb{I}_{k=2}\right)}{4k}\right) + \left(\sum_{k=K+1}^{m_{n_{l_{q_m}}}} \frac{\mu_k^2}{4k}\right)^2 \\
&= (1+o(1))\sum_{k=K+1}^{m_{n_{l_{q_m}}}} \frac{\mu_k^2}{2k} + \left(\sum_{k=K+1}^{m_{n_{l_{q_m}}}} \frac{\mu_k^2}{4k}\right)^2 .
\end{aligned} \tag{1.6.22}$$

Given any $\epsilon > 0$, we now choose K large enough so that $\sum_{k=K+1}^{\infty} \frac{\mu_k^2}{2k} \leq \epsilon$, implying $\mathrm{E}\left[(M_1 - M_{2,K})^2\right] = \epsilon + \epsilon^2/4$. Hence the R.S. of (1.6.21) converges to 0 as $K \to \infty$. This implies $W_2\left(F^{M_1}, F^{M_{2,K}}\right) \to 0$ as $K \to \infty$. Here F^{M_1} and $F^{M_{2,K}}$ denote the distribution functions of M_1 and $M_{2,K}$ respectively. As a consequence we have

$$W + \sum_{k=1}^{K-1} \frac{2\mu_{k+1}Z_k - \mu_{k+1}^2}{4(k+1)} \xrightarrow{d} M\{n_{l_q}\}.$$

As a consequence, $M\{n_{l_q}\} \stackrel{d}{=} W + \sum_{k=1}^{\infty} \frac{2\mu_{k+1}Z_k - \mu_{k+1}^2}{4(k+1)}$ which does not depend on the specific choice of the subsequence $\{n_{l_q}\}$. This concludes the proof. □

1.7. Appendix

We now give proofs of Propositions 3 and 4

1.7.1. Proof of Proposition 3

Proof of mutual contiguity and (1.4.2) This proof is broken into two steps. We focus on proving (1.4.2). Given (1.4.2), mutual contiguity is a direct consequence of Le Cam's first lemma Le Cam (2012).

Step 1. We first prove the random variable on the right hand side of (1.4.2) is almost

surely positive and has mean 1. Let us define

$$L := \exp\left\{\sum_{i=1}^{\infty} \frac{2\mu_i Z_i - \mu_i^2}{2\sigma_i^2}\right\}, \qquad L^{(m)} := \exp\left\{\sum_{i=1}^{m} \frac{2\mu_i Z_i - \mu_i^2}{2\sigma_i^2}\right\}, \quad \forall m \in \mathbb{N}.$$

As $Z_i \sim N(0, \sigma_i^2)$, for any $i \in \mathbb{N}$, and so

$$\mathrm{E}\left[\exp\left\{\frac{2\mu_i Z_i - \mu_i^2}{2\sigma_i^2}\right\}\right] = 1.$$

So $\{L^{(m)}\}_{m=1}^{\infty}$ is a martingale sequence and

$$\mathrm{E}\left[\left(L^{(m)}\right)^2\right] = \prod_{i=1}^{m} \exp\left\{\frac{\mu_i^2}{\sigma_i^2}\right\} = \exp\left\{\sum_{i=1}^{m} \frac{\mu_i^2}{\sigma_i^2}\right\}.$$

Now by the righthand side of (1.4.1), $L^{(m)}$ is a L^2 bounded martingale. Hence, L is a well defined random variable with

$$\mathrm{E}[L] = 1, \qquad \mathrm{E}[L^2] = \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\}.$$

On the other hand $\log(L)$ is a limit of Gaussian random variables, hence $\log(L)$ is Gaussian with

$$\mathrm{E}[\log(L)] = -\frac{1}{2}\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}, \qquad \mathrm{Var}(\log(L)) = \sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}.$$

Hence $\mathbb{P}(L = 0) = \mathbb{P}(\log(L) = -\infty) = 0$.

Step 2. Now we prove $Y_n \xrightarrow{d} L$. Since

$$\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[Y_n^2\right] < \infty,$$

condition (iv) implies that the sequence Y_n is tight. Prokhorov's theorem further implies that there is a subsequence $\{n_k\}_{k=1}^{\infty}$ such that Y_{n_k} converge in distribution to some random variable $L(\{n_k\})$. In what follows, we prove that the distribution of $L(\{n_k\})$ does not depend on the subsequence $\{n_k\}$. In particular, $L(\{n_k\}) \stackrel{d}{=} L$. To start with, note that since Y_{n_k} converges in distribution to $L(\{n_k\})$, for any further subsequence $\{n_{k_l}\}$ of $\{n_k\}$, $Y_{n_{k_l}}$ also converges in distribution to $L(\{n_k\})$.

Given any fixed $\epsilon > 0$ take m large enough such that

$$\exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\} - \exp\left\{\sum_{i=1}^{m} \frac{\mu_i^2}{\sigma_i^2}\right\} < \epsilon.$$

For this fixed number m, consider the joint distribution of $(Y_{n_k}, W_{n_k,1}, \ldots, W_{n_k,m})$. This sequence of $m+1$ dimensional random vectors with respect to $\mathbb{P}_{n_k}$ is tight by condition (ii). So it has a further subsequence such that

$$(Y_{n_{k_l}}, W_{n_{k_l},1}, \ldots, W_{n_{k_l},m})|\mathbb{P}_{n_{k_l}} \xrightarrow{d} ((H_1, \ldots, H_{m+1}) \in (\Omega(\{n_{k_l}\}), \mathcal{F}(\{n_{k_l}\}), P(\{n_{k_l}\}))(say).) \,.$$

where $H_1 \overset{d}{=} L(\{n_k\})$ and $(H_2, \ldots, H_{m+1}) \overset{d}{=} (Z_1, \ldots, Z_m)$ We are to show that we can define the random variables $L^{(m)}$ and $L(\{n_k\})$ in such a way that there exist suitable σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ such that $L^{(m)} \in \mathcal{F}_1$, $L(\{n_k\}) \in \mathcal{F}_2$, and $\mathrm{E}\left[L(\{n_k\}) \mid \mathcal{F}_1\right] = L^{(m)}$.

Since $\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[Y_n^2\right] < \infty$, the sequence $Y_{n_{k_l}}$ is uniformly integrable. This, together with condition (i), leads to

$$\mathrm{E}[L(\{n_k\})] = \lim_{l\to\infty} \mathrm{E}_{\mathbb{P}_{n_{k_l}}}[Y_{n_{k_l}}] = 1. \tag{1.7.1}$$

Now take any positive bounded continuous function $f : \mathbb{R}^m \to \mathbb{R}$. By Fatou's lemma

$$\liminf_{l\to\infty} \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[f(W_{n_{k_l},1}, \ldots, W_{n_{k_l},m})Y_{n_{k_l}}\right] \geq \mathrm{E}\left[f\left(Z_1, \ldots, Z_m\right) L(\{n_k\})\right]. \tag{1.7.2}$$

However for any constant ξ, (1.7.1) implies $\xi = \xi \mathrm{E}_{\mathbb{P}_{n_{k_l}}}[Y_{n_{k_l}}] \to \xi \mathrm{E}[L(\{n_k\})] = \xi$. Observe that given any bounded continuous function f we can find ξ large enough so that $f + \xi$ is a positive bounded continuous function. So (1.7.2) is indeed implied by Fatou's lemma.

Now

$$\begin{aligned}
&\liminf \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[\left(f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m})+\xi\right)Y_{n_{k_l}}\right] \\
&=\liminf \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m})Y_{n_{k_l}}\right]+\xi \\
&\geq \mathrm{E}\left[\left(f(Z_1,\ldots,Z_m)+\xi\right)L(\{n_k\})\right]
\end{aligned} \tag{1.7.3}$$

So (1.7.2) holds for any bounded continuous function f. On the other hand, replacing f by $-f$ we have

$$\lim_{l\to\infty} \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m})Y_{n_{k_l}}\right]=\mathrm{E}\left[f(Z_1,\ldots,Z_m)L(\{n_k\})\right]. \tag{1.7.4}$$

Now condition (ii) leads to

$$\int f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m})Y_{n_{k_l}}\mathrm{d}\mathbb{P}_{n_{k_l}}=\int f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m})\mathrm{d}\mathbb{Q}_{n_{k_l}}\to\int f(Z'_1,\ldots,Z'_m)\mathrm{d}Q.$$

Here Q is the measure induced by $(Z'_1,\ldots,Z'_m)$. In particular, one can take the measure Q such that $(Z_1,\ldots,Z_m)$ themselves are distributed as $(Z'_1,\ldots,Z'_m)$ under the measure Q. This is true since

$$\int f(Z'_1,\ldots,Z'_m)\mathrm{d}Q=\mathrm{E}\left[f(Z_1,\ldots,Z_m)L^{(m)}\right].$$

for any bounded continuous function f, and so $\int_A \mathrm{d}Q=\mathrm{E}[\mathbf{1}_A L^{(m)}]$ for any $A\in\sigma(Z_1,\ldots,Z_m)$. Now looking back into (1.7.4), we have for any $A\in\sigma(Z_1,\ldots,Z_m)$, $\mathrm{E}[\mathbf{1}_A L^{(m)}]=\mathrm{E}\left[\mathbf{1}_A L(\{n_k\})\right]$. Since by definition $L^{(m)}$ is $\sigma(Z_1,\ldots,Z_m)$ measurable, we have

$$L^{(m)}=\mathrm{E}\left[L(\{n_k\})\mid\sigma(Z_1,\ldots,Z_m)\right].$$

From Fatou's lemma

$$\mathrm{E}[L(\{n_k\})^2]\leq\liminf_{n\to\infty}\mathrm{E}_{\mathbb{P}_n}[Y_n^2]=\exp\left\{\sum_{i=1}^{\infty}\frac{\mu_i^2}{\sigma_i^2}\right\}.$$

As a consequence, we have

$$0 \le \mathrm{E}|L(\{n_k\}) - L^{(m)}|^2 = \mathrm{E}[L(\{n_k\})^2] - \mathrm{E}[L^{(m)2}] < \epsilon.$$

So $L_2(F^{L^{(m)}}, F^{L(\{n_k\})}) < \sqrt{\epsilon}$. Here $F^{L^{(m)}}$ and $F^{L(\{n_k\})}$ denote the distribution functions corresponding to $L^{(m)}$ and $L(\{n_k\})$ respectively. As a consequence, $W_2(F^{L^{(m)}}, F^{L(\{n_k\})}) \to 0$ as $m \to \infty$. Hence $L^{(m)} \xrightarrow{d} L(\{n_k\})$ by the result stated after Definition 4. On the other hand, we have already proved $L^{(m)}$ converges to L in L^2. So $L(\{n_k\}) \stackrel{d}{=} L$.

Proof of (1.4.3) We start with a sub sequence $\{n_l\}$. We shall choose k large enough which shall be specified later. We also know that both the random variables $\log(Y_{n_l})$ and $\left\{\sum_{i=1}^{k} \frac{2\mu_i W_{n_l,i} - \mu_i^2}{2\sigma_i^2}\right\}$ are tight.

We now prove that there is a M invariant of k such that both the probabilities

$$\begin{aligned} &\mathbb{P}_{n_l}\left[-M \le \log(Y_{n_l}) \le M\right] \ge 1 - \frac{\delta}{100} \\ &\mathbb{P}_{n_l}\left[-M \le \left\{\sum_{i=1}^{k} \frac{2\mu_i W_{n_l,i} - \mu_i^2}{2\sigma_i^2}\right\} \le M\right] \ge 1 - \frac{\delta}{100} \end{aligned} \tag{1.7.5}$$

for all n_l. Since the random variable Y_{n_l} do not depend on k the first inequality is obvious. For the second inequality observe that

$$\mathrm{Var}\left[\left\{\sum_{i=1}^{k} \frac{2\mu_i W_{n_l,i} - \mu_i^2}{2\sigma_i^2}\right\}\right] \le \mathrm{Var}\left[\sum_{i=1}^{m_n} \frac{2\mu_i W_{n_l,i} - \mu_i^2}{2\sigma_i^2}\right]$$

where m_n is a sequence increasing to infinity as mentioned in Proposition 3. Now

$$\mathrm{Var}\left[\sum_{i=1}^{m_n} \frac{2\mu_i W_{n_l,i} - \mu_i^2}{2\sigma_i^2}\right] < C' \tag{1.7.6}$$

for all n_l. for a deterministic constant C'. Since $\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^i} < \infty$ we have for some deterministic constant C'',

$$\mathbb{P}_{n_l}\left[\left|\sum_{i=1}^{k}\frac{2\mu_i W_{n_l,i}-\mu_i^2}{2\sigma_i^2}\right|>M\right]\leq\frac{C''}{M^2}\leq\frac{\delta}{100} \tag{1.7.7}$$

where $M^2=\frac{100C''}{\delta}$.

$$\mathbb{P}_{n_l}\left[-M\leq\log\left(Y_{n_l}\right)\leq M\cap -M\leq\left\{\sum_{i=1}^{k}\frac{2\mu_i W_{n_l,i}-\mu_i^2}{2\sigma_i^2}\right\}\leq M\right]\geq 1-\frac{\delta}{50}.$$

Now $\log(\cdot)$ is an uniformly continuous function on $[e^{-M},e^{M}]$. So given $\epsilon>0$, there exists $\tilde{\epsilon}$ such that for any $x,y\in[e^{-M},e^{M}]$,

$$\begin{aligned}|x-y|\leq\tilde{\epsilon}&\Rightarrow|\log(x)-\log(y)|\leq\epsilon\\ \Leftrightarrow|x-y|>\tilde{\epsilon}&\Leftarrow|\log(x)-\log(y)|>\epsilon\end{aligned} \tag{1.7.8}$$

Observe that given We have seen that the sequence $(Y_{n_l},W_{n_l,1},\ldots,W_{n_l,k})$ is tight for any given k. We know that there is a further sub-sequence n_{l_m} such that $(Y_{n_{l_m}},W_{n_{l_m},1},\ldots,W_{n_{l_m},k})$ converges jointly in distribution to

$$(Y_{n_{l_m}},W_{n_{l_m},1},\ldots,W_{n_{l_m},k})\xrightarrow{d}(H_1,H_2,\ldots,H_{k+1})\in(\Omega\{n_{l_m}\},\mathcal{F}\{n_{l_m}\},\mathbb{P}\{n_{l_m}\}).$$

Let $\mathcal{F}\{n_{l_m},1\}\subset\mathcal{F}\{n_{l_m}\}$ be the sigma algebra generated by $(H_2,\ldots,H_{k+1})$. Here $H_1\stackrel{d}{=}L$ and $(H_2,\ldots,H_{k+1})\stackrel{d}{=}(Z_1,\ldots,Z_k)$. Using the arguments same as the previous proof we see that

$$\mathrm{E}\left[H_1\left|\mathcal{F}_{n_1,1}\right.\right]=\exp\left\{\sum_{i=1}^{k}\frac{2\mu_i H_{i+1}-\mu_i^2}{2\sigma_i^2}\right\}.$$

As a consequence, we have

$$0\leq\mathrm{E}\left(H_1-\exp\left\{\sum_{i=1}^{k}\frac{2\mu_i H_{i+1}-\mu_i^2}{2\sigma_i^2}\right\}\right)^2\leq\exp\left\{\sum_{i=1}^{\infty}\frac{\mu_i^2}{\sigma_i^2}\right\}-\exp\left\{\sum_{i=1}^{k}\frac{\mu_i^2}{\sigma_i^2}\right\}.$$

We shall choose this k large enough so that

$$\exp\left\{\sum_{i=1}^{\infty}\frac{\mu_i^2}{\sigma_i^2}\right\} - \exp\left\{\sum_{i=1}^{k}\frac{\mu_i^2}{\sigma_i^2}\right\} < \frac{\delta\tilde{\epsilon}^2}{100}.$$

Now by Chebyshev's inequality

$$\mathbb{P}\left[\left|H_1 - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2}\right\}\right| \geq \frac{\tilde{\epsilon}}{2}\right] \leq \frac{\delta\tilde{\epsilon}^2}{25\tilde{\epsilon}^2} = \frac{\delta}{25}.$$

Since

$$\left(Y_{n_{l_m}}, W_{n_{l_m},1}, \ldots, W_{n_{l_m},k}\right) \xrightarrow{d} (H_1, H_2, \ldots, H_{k+1})$$

by continuous mapping theorem for in distributional convergence, we have

$$Y_{n_{l_m}} - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2}\right\} \xrightarrow{d} H_1 - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2}\right\}.$$

Since the set $[\frac{\tilde{\epsilon}}{2}, \infty)$ is closed, we have by Portmanteau theorem,

$$\begin{aligned}
&\limsup_{n_{l_m}} \mathbb{P}_{n_{l_m}}\left[\left|Y_{n_{l_m}} - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2}\right\}\right| > \tilde{\epsilon}\right] \\
&\leq \limsup_{n_{l_m}} \mathbb{P}_{n_{l_m}}\left[\left|Y_{n_{l_m}} - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2}\right\}\right| \geq \frac{\tilde{\epsilon}}{2}\right] \\
&\leq \frac{\delta}{25}.
\end{aligned} \tag{1.7.9}$$

As a consequence,

$$\begin{aligned}
\frac{\delta}{25} &\geq \limsup \mathbb{P}_{n_{l_m}} \left[\left| Y_{n_{l_m}} - \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \tilde{\epsilon} \right] \\
&\geq \limsup \mathbb{P}_{n_{l_m}} \left[Y_{n_{l_m}} \in [e^{-M}, e^{M}] \cap \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \in [e^{-M}, e^{M}] \right. \\
&\qquad \left. \cap \left| Y_{n_{l_m}} - \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \tilde{\epsilon} \right] \\
&\geq \limsup \mathbb{P}_{n_{l_m}} \left[Y_{n_{l_m}} \in [e^{-M}, e^{M}] \cap \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \in [e^{-M}, e^{M}] \right. \\
&\qquad \left. \cap \left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \epsilon \right] \\
&\geq \limsup 1 - \mathbb{P}_{n_{l_m}} \left[\left(Y_{n_{l_m}} \in [e^{-M}, e^{M}] \cap \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \in [e^{-M}, e^{M}] \right)^c \right] \\
&\qquad\qquad - \mathbb{P}_{n_{l_m}} \left[\left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| \leq \epsilon \right] \\
&\geq \limsup \left(\mathbb{P}_{n_{l_m}} \left[\left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \epsilon \right] - \frac{\delta}{100} \right) \\
&\Rightarrow \limsup \mathbb{P}_{n_{l_m}} \left[\left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \epsilon \right] \leq \frac{\delta}{25} + \frac{\delta}{100} < \delta.
\end{aligned} \tag{1.7.10}$$

□

CHAPTER 2 : Contiguity and non-reconstruction results for planted partition models: the dense case

2.1. Overview

We consider the two block stochastic block model on n nodes with asymptotically equal cluster sizes. The connection probabilities within and between cluster are denoted by $p_n := \frac{a_n}{n}$ and $q_n := \frac{b_n}{n}$ respectively. Mossel et al. (2015) considered the case when $a_n = a$ and $b_n = b$ are fixed. They proved the probability models of the stochastic block model and that of Erdős–Rényi graph with same average degree are mutually contiguous whenever $(a-b)^2 < 2(a+b)$ and are asymptotically singular whenever $(a-b)^2 > 2(a+b)$. Mossel et al. (2015) also proved that when $(a-b)^2 < 2(a+b)$ no algorithm is able to find an estimate of the labeling of the nodes which is positively correlated with the true labeling. It is natural to ask what happens when a_n and b_n both grow to infinity. In this paper we consider the case when $a_n \to \infty$, $\frac{a_n}{n} \to p \in [0,1)$ and $(a_n - b_n)^2 = \Theta(a_n + b_n)$. Observe that in this case $\frac{b_n}{n} \to p$ also. We show that here the models are mutually contiguous if asymptotically $(a_n - b_n)^2 < 2(1-p)(a_n + b_n)$ and they are asymptotically singular if asymptotically $(a_n - b_n)^2 > 2(1-p)(a_n + b_n)$. Further we also prove it is impossible find an estimate of the labeling of the nodes which is positively correlated with the true labeling whenever $(a_n - b_n)^2 < 2(1-p)(a_n + b_n)$ asymptotically. The results of this paper justify the negative part of a conjecture made in Decelle et al. (2011) for dense graphs.

2.2. Introduction

In the last few years the stochastic block model has been one of the most active domains of modern research in statistics, computer science and many other related fields. In general a stochastic block model is a network with a hidden community structure where the nodes within the communities are expected to be connected in a different manner than the nodes between the communities. This model arises naturally in many problems of statistics, machine learning and data mining, but its applications further extends to population genetics

Pritchard et al. (2000) , where genetically similar sub-populations are used as the clusters, to image processing Shi and Malik (2000), Sonka et al. (2007) , where the group of similar images acts as cluster, to the study of social networks , where groups of like-minded people act as clusters Newman et al. (2002).

Recently a huge amount of effort has been dedicated to find out the clusters. Numerous different clustering algorithms have been proposed in literature. One might look at Johnson (1967),Dempster et al. (1977), Bui et al. (1987), Dyer and Frieze (1989), Boppana (1987), Bickel and Chen (2009), Condon and Karp (1999), Rohe et al. (2011), McSherry (2001) for some references. One might also look at the review paper by **?** for a detailed study of the literature.

One of the easiest examples of the stochastic block model is the planted partition model where one have only two clusters of more or less equal size. Formally,

Definition 7. For $n \in \mathbb{N}$, and $p, q \in [0, 1]$ let $\mathcal{G}(n, p, q)$ denote the model of random,$\pm$ labelled graphs in which each vertex u is assigned (independently and uniformly at random) a label $\sigma_u \in \{\pm 1\}$ and each edge between u and v are included independently with probability p if they have the same label and with probability q if they have different labels.

The case when p and q are sufficiently close to each other has got significant amount of interest in literature. Decelle et al. (2011) made a fascinating conjecture in this regard.

Conjecture 1. *Let $p = \frac{a}{n}$ and $q = \frac{b}{n}$ where a and b are fixed real numbers. Then the following are true.*

i) If $(a - b)^2 > 2(a + b)$ then one can find almost surely a bisection of the vertices which is positively correlated with the original clusters.

ii) If $(a - b)^2 < 2(a + b)$ then the problem is not solveable.

iii) Further, there are no consistent estimators of a and b if $(a - b)^2 < 2(a + b)$ and there are consistent estimators of a and b whenever $(a - b)^2 > 2(a + b)$.

Coja-Oghlan (2010) solved part $i)$ of the problem when $(a - b)^2 > C(a + b)$ for some large

C and finally part ii) and iii) of Conjecture 1 was proved by Mossel et al. (2015) and part i) was solved by Mossel et al. (2013) and Massoulié (2013) independently.

Typically the problem is much more delicate when more than two communities are present in the sparse case. To keep things simple let us consider the general stochastic block model with k asymptotically equal sized blocks with connection probabilities within and between blocks are given by $\frac{a}{n}$ and $\frac{b}{n}$ respectively. It was conjectured in Mossel et al. (2015) that for k sufficiently large, there is a constant $c(k)$ such that whenever

$$c(k) < \frac{(a-b)^2}{a+(k-1)b} < k$$

the reconstruction problem is solvable in exponential time, it is not solvable if $\frac{(a-b)^2}{a+(k-1)b} < c(k)$ and solvable in polynomial time if $k < \frac{(a-b)^2}{a+(k-1)b}$. The upper bound is known as Kesten-Stigum threshold. Bordenave et al. (2015) solved the reconstruction problem above a deterministic threshold by spectral analysis of non-backtraking matrix. One might look at Banks et al. (2016) for the non solvability part. They proved that the probability models of stochastic block model and that of Erdős–Rényi graph with same average degree are contiguous and the reconstruction problem is unsolvable if

$$d < \frac{2\log(k-1)}{k-1}\frac{1}{\lambda^2}.$$

Here $d = \frac{a+(k-1)b}{k}$ and $\lambda = \frac{a-b}{kd}$. Abbe and Sandon (2015) provides an efficient algorithm for reconstruction above the Kesten-Stigum threshold. Abbe and Sandon (2015) and Banks et al. (2016) also provide cases strictly below the Kesten-Stigum threshold where the problem is solvable in exponential time.

On the other hand, a different type of reconstruction problem was considered in Mossel et al. (2016) for denser graphs. They considered two different notions of recovery. The first one is weak consistency where one is interested in finding a bisection $\hat{\sigma}$ such that σ and $\hat{\sigma}$ have correlation going to 1 with high probability. The second one is called strong consistency. Here one is interested in finding a bisection $\hat{\sigma}$ such that $\hat{\sigma}$ is either σ or $-\sigma$ with probability tending to 1. Mossel et al. (2016) prove that weak consistency is possible if and only if $\frac{n(p_n-q_n)^2}{p_n+q_n} \to \infty$ and strong consistency is possible if and only if

$$\left(a_n + b_n - 2\sqrt{a_n b_n} - 1\right)\log n + \frac{1}{2}\log\log n \to \infty.$$

Here $a_n = \frac{np_n}{\log n}$ and $b_n = \frac{nq_n}{\log n}$ respectively. (Abbe et al., 2014) studied the same problem independently in the logarithmic sparsity regime. They prove that for $a = \frac{np_n}{\log n}$ and $b = \frac{nq_n}{\log n}$ fixed, $(a + b) - 2\sqrt{ab} > 1$ is sufficient for strong consistency and that $(a + b) - 2\sqrt{ab} \geq 1$ is necessary. We note that their results are implied by Mossel et al. (2016).

However, according to the best of our knowledge questions similar to part *ii*) and *iii*) of Conjecture 1 have not yet been addressed in dense case (i.e. when a and b increase to infinity). This is the main focus of this paper.

Before stating our results we mention that the results in Mossel et al. (2015) is more general than part *iii*) of Conjecture 1. Let $\mathbb{P}_n$ and $\mathbb{P}'_n$ be the sequences of probability measures induced by $\mathcal{G}(n, p, q)$ and $\mathcal{G}(n, \frac{p+q}{2}, \frac{p+q}{2})$ respectively. Then Mossel et al. (2015) prove that whenever a and b are fixed numbers and $(a - b)^2 < 2(a + b)$, the measures $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous i.e. for a sequence of events A_n, $\mathbb{P}_n(A_n) \to 0$ if and only if $\mathbb{P}'_n(A_n) \to 0$. Now part *iii*) of Conjecture 1 directly follows from the contiguity. The proof in Mossel et al. (2015) is based on calculating the limiting distribution of the short cycles and using a result on contiguity (Theorem 1 in Janson (1995) and Theorem 4.1 in Wormald (1999)). However, one should note that the result from Mossel et al. (2015) does not directly generalize to the denser case. Since, one requires the limiting distributions of short cycles to be independent Poisson in order to use Janson's result. In our proof instead of considering the short cycles we consider the "signed cycles"(to be defined later) which have asymptotic normal distributions. We also find a result analogous to Janson for the normal random variables in order to complete the proof. / On the other hand, the original proof of non-reconstruction from Mossel et al. (2015) relies on the coupling of $\mathbb{P}_n$ and $\mathbb{P}'_n$ with probability measure induced by Galton Watson trees of suitable parameters. However, it is well known that when the graph is sufficiently dense i.e. $a_n \gg n^{o(1)}$ the coupling argument does not work. So our proof is based on fine analysis of conditional probabilities. Technically, this proof is closely related to the non-reconstruction proof in section 6.2 of Banks et al. (2016) rather than the original proof given in Mossel et al. (2015).

A natural question arises how far the arguments in this paper generalize to the multi-community case. Unfortunately, we do not have a definite answer for this problem. The fundamental difficulty here is the absence of locally tree like structure which is the essence of all the proofs in the sparse regime. However, we believe the similar thresholds are true even in dense case also. In fact, it was shown in **?** that for the multi-community case the models are mutually singular much below the Kesten-Stigum threshold. We leave the problem for future research.

The paper is organized in the following manner. In Section 2.3 we build some preliminary notations and state our results. Section 2.4 is dedicated for building a result analogous to Theorem 1 in Janson (1995). In Section 2.5 we define signed cycles and find their asymptotic distributions. Section 2.6 is dedicated to complete the proofs of our contiguity results. In Section 2.7 we prove the non-reconstruction result. Finally, the paper concludes with an Appendix containing a proof of a result from random matrix theory used in this paper.

2.3. Our results

Through out the paper a random graph will be denoted by G and $x_{i,j}$ will be used to denote the indicator random variable corresponding to an edge between the nodes i and j. Further $\mathbb{P}_n$ and $\mathbb{P}'_n$ will be used to denote the sequence of probability measures induced by $\mathcal{G}(n, p_n, q_n)$ and $\mathcal{G}(n, \frac{p_n+q_n}{2}, \frac{p_n+q_n}{2})$ respectively. For notational simplicity we denote $\frac{p_n+q_n}{2}$ by $\hat{p}_n$.

In this paper we shall consider the case when $(a_n - b_n)^2 = \Theta(a_n + b_n)$. We shall use the following notations through out the paper. We denote $c_n := \frac{(a_n-b_n)^2}{(a_n+b_n)}$, $d_n := \frac{p_n-q_n}{2}$ and

$t_n = \frac{c_n}{2(1-\hat{p}_n)}$.

Further, for any two labeling of the nodes σ and τ, we define their overlap to be

$$\text{ov}(\sigma, \tau) := \frac{1}{n}\left(\sum_{i=1}^{n} \sigma_i \tau_i - \frac{1}{n}\left(\sum_{i=1}^{n} \sigma_i\right)\left(\sum_{i=1}^{n} \tau_i\right)\right). \tag{2.3.1}$$

Now we define mutual contiguity of two sequences of measures as follows:

Definition 8. Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures, such that for each n, $\mathbb{P}_n$ and $\mathbb{Q}_n$ both are defined on the same measurable space $(\Omega_n, \mathcal{F}_n)$. We then say that the sequences are mutually contiguous if for every sequence of measurable sets $A_n \subset \Omega_n$,

$$\mathbb{P}_n(A_n) \to 0 \Leftrightarrow \mathbb{Q}_n(A_n) \to 0.$$

Two sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are called asymptotically mutually singular if there exists a sequence of measurable sets A_n such that $\mathbb{P}_n(A_n) \to 1$ and $\mathbb{Q}_n(A_n^c) \to 1$ as $n \to \infty$.

We are now ready to state the main results of the paper.

Theorem 2. *i)If $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0, 1)$ and $c_n \to c < 2(1-p)$, then the probability measures $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous. So there does not exist an estimator (A_n, B_n) for (a_n, b_n) such that $|A_n - a_n| + |B_n - b_n| = o_p(a_n - b_n)$.*
ii)If $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0, 1)$ and $c_n \to c > 2(1-p)$, then the probability measures $\mathbb{P}_n$ and $\mathbb{P}'_n$ are asymptotically mutually singular. Further there exists an estimator (A_n, B_n) for (a_n, b_n) such that $|A_n - a_n| + |B_n - b_n| = o_p(a_n - b_n)$.

Theorem 3. *If $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0, 1)$ and $c_n \to c < 2(1-p)$, then there is no reconstruction algorithm which performs better than the random guessing i.e. for any estimate of the labeling $\{\hat{\sigma}_i\}_{i=1}^{n}$ we have*

$$\text{ov}(\sigma, \hat{\sigma}) \xrightarrow{P} 0. \tag{2.3.2}$$

2.4. A result on contiguity

In this section we provide a very brief description of contiguity of probability measures. We suggest the reader to have a look at the discussion about contiguity of measures in Janson (1995) for further details. In this section we state several propositions and apart from Proposition 9 and Proposition 8, all the proofs can be found in Janson (1995).

Definition 8 of contiguity might appear a little abstract. However the following reformulation is perhaps more useful to understand the contiguity concept.

Proposition 6. *Two sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if and only if for every $\varepsilon > 0$ there exist $n(\varepsilon)$ and $K(\varepsilon)$ such that for all $n > n(\varepsilon)$ there exists a set $B_n \in \mathcal{F}_n$ with $\mathbb{P}_n(B_n^c), \mathbb{Q}_n(B_n^c) \leq \varepsilon$ such that*

$$K(\varepsilon)^{-1} \leq \frac{\mathbb{Q}_n(A_n)}{\mathbb{P}_n(A_n)} \leq K(\varepsilon). \;\; \forall A_n \subset B_n.$$

Although Proposition 6 gives an equivalent condition, verifying this condition is often difficult. However under the assumption of convergence of $\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$, one gets the following simplified result.

Proposition 7. *Suppose that $L_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$, regarded as a random variable on $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, converges in distribution to some random variable L as $n \to \infty$. Then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are contiguous if and only if $L > 0$ a.s. and* $\mathrm{E}[L] = 1$.

We now introduce the concept of Wasserstein's metric which will be used in the proof of Proposition 9.

Definition 9. Let F and G be two distribution functions with finite p th moment. Then the Wasserstein distance W_p between F and G is defined to be

$$W_p(F, G) = \left[\inf_{X \sim F, Y \sim G} \mathrm{E}|X - Y|^p\right]^{\frac{1}{p}}.$$

Here X and Y are random variables having distribution functions F and G respectively.

In particular, the following result will be useful in our proof:

Proposition 8. *Let F_n be a sequence of distribution functions and F be a distribution function. Then F_n converge to F in distribution and $\int x^2 dF_n(x) \to \int x^2 dF(x)$ if $W_2(F_n, F) \to 0$.*

The proof of Proposition 8 is well known. One might look at Mallows (1972) for a reference.

With Proposition 7 in hand, we now state the most important result in this section. This result will be used to prove Theorem 2. Although, Proposition 9 is written in a complete different notation, one can check that it is analogous to Theorem 1 in Janson (1995).

Proposition 9. *Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures such that for each n, both of them are defined on $(\Omega_n, \mathcal{F}_n)$. Suppose that for each $i \geq 3$, $X_{n,i}$ are random variables defined on $(\Omega_n, \mathcal{F}_n)$. Then the probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if the following conditions hold:*

i) $\mathbb{P}_n \ll \mathbb{Q}_n$ *and* $\mathbb{Q}_n \ll \mathbb{P}_n$ *for each* n.

ii) For any fixed $k \geq 3$, *one has* $(X_{n,3}, \ldots, X_{n,k})\,|\mathbb{P}_n \xrightarrow{d} (Z_3, \ldots, Z_k)$ *and* $(X_{n,3}, \ldots, X_{n,k})\,|\mathbb{Q}_n \xrightarrow{d} (Z'_3, \ldots, Z'_k)$.

iii) $Z_i \sim N(0, 2i)$ *and* $Z'_i \sim N(t^{\frac{i}{2}}, 2i)$ *are sequences of independent random variables. Here* $|t| < 1$.

iv)

$$\mathrm{E}_{\mathbb{P}_n}\left[\left(\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right)^2\right] \to \exp\left\{-\frac{t}{2} - \frac{t^2}{4}\right\}\frac{1}{\sqrt{1-t}}. \tag{2.4.1}$$

Further,

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}|\mathbb{P}_n \xrightarrow{d} \exp\left\{\sum_{i=3}^{\infty} \frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}. \tag{2.4.2}$$

Proof. In this proof for simplicity we denote $\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$ by Y_n. We break the proof into two steps.

Step 1. In this step we prove the random variable in the right hand side of (**??**) is almost

surely positive and has mean 1. Let us define

$$W := \exp\left\{\sum_{i=3}^{\infty} \frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}$$

and

$$W^{(m)} := \exp\left\{\sum_{i=3}^{m} \frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}.$$

As $Z_i \sim N(0, 2i)$,

$$\mathrm{E}\left[\exp\left\{\frac{2t^{\frac{i}{2}} Z_i - t^i}{4i}\right\}\right] = \exp\left\{\frac{4t^i \times 2i}{2 \times 16i^2} - \frac{t^i}{4i}\right\} = 1.$$

So $\{W^{(m)}\}_{m=3}^{\infty}$ is a martingale sequence and

$$\mathrm{E}\left[W^{(m)2}\right] = \prod_{i=3}^{m} \exp\left\{\frac{t^i}{2i}\right\} = \exp\left\{\sum_{i=3}^{m} \frac{t^i}{2i}\right\}.$$

Now

$$\sum_{i=3}^{\infty} \frac{t^i}{2i} = \frac{1}{2}\left(-\log(1-t) - t - \frac{t^2}{2}\right) \quad \forall \ \ |t| < 1.$$

So $W^{(m)}$ is a L^2 bounded martingale. Hence, W is a well defined random variable,

$$\mathrm{E}[W^2] = \exp\left\{-\frac{t}{2} - \frac{t^2}{4}\right\} \frac{1}{\sqrt{1-t}}$$

and $\mathrm{E}[W] = 1$.

Now observe that $Z_i \stackrel{d}{=} -Z_i$ for each i and whenever $|t| < 1$, the series $\sum_{i=3}^{\infty} \frac{t^i}{4i}$ converges. So

$$W^{-1} \stackrel{d}{=} \exp\left\{\sum_{i=3}^{\infty} \frac{2t^{\frac{i}{2}} Z_i + t^i}{4i}\right\}.$$

However, $E[W^{-1}] = \exp\left\{\sum_{i=3}^{\infty} \frac{t^i}{2i}\right\} < \infty$ implies $W > 0$ a.s.

Step 2. Now we come to the harder task of proving $Y_n \xrightarrow{d} W$. Since

$$\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[(Y_n)^2\right] < \infty$$

from condition iv), the sequence Y_n is tight. Hence from Prokhorov's theorem there is a sub sequence $\{n_k\}_{k=1}^{\infty}$ such that Y_{n_k} converge in distribution to some random variable $W(\{n_k\})$. We shall prove that the distribution of $W(\{n_k\})$ does not depend on the sub sequence $\{n_k\}$. In particular, $W(\{n_k\}) \stackrel{d}{=} W$.

Since Y_{n_k} converges in distribution to $W(\{n_k\})$, for any further sub sequence $\{n_{k_l}\}$ of $\{n_k\}$, $Y_{n_{k_l}}$ also converges in distribution to $W(\{n_k\})$.

Given $\varepsilon > 0$ take m big enough such that

$$\exp\left\{\sum_{i=3}^{\infty} \frac{t^i}{2i}\right\} - \exp\left\{\sum_{i=3}^{m} \frac{t^i}{2i}\right\} < \varepsilon.$$

For this m, look at the joint distribution of $(Y_{n_k}, X_{n_k,3}, \ldots, X_{n_k,m})$. This sequence of $m-1$ dimensional random vectors with respect to $\mathbb{P}_{n_k}$ is also tight from condition ii). So it has a further sub sequence such that

$$(Y_{n_{k_l}}, X_{n_{k_l},3}, \ldots, X_{n_{k_l},m})|\mathbb{P}_{n_{k_l}} \xrightarrow{d} (W(\{n_k\}), Z_3 \ldots, Z_m)\,.$$

Here we have used condition ii) for the convergence of $(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m})|\mathbb{P}_{n_{k_l}}$.

The most important part of this proof is to show, we can define the random variables $W^{(m)}$ and $W(\{n_k\})$ in such a way that there exist suitable σ algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ such that $W^{(m)} \in \mathcal{F}_1$ and $W(\{n_k\}) \in \mathcal{F}_2$ and $\mathrm{E}\left[W(\{n_k\}) \mid \mathcal{F}_1\right] = W^{(m)}$.

From condition iv) we have $\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[Y_n^2\right] < \infty$. As a consequence, the sequence the sequence $Y_{n_{k_l}}$ is uniformly integrable. This together with condition i) will give us

$$1 = \mathrm{E}_{\mathbb{P}_{n_{k_l}}}\left[Y_{n_{k_l}}\right] \to \mathrm{E}[W(\{n_k\})] = 1. \tag{2.4.3}$$

Now take any positive bounded continuous function $f : \mathbb{R}^m \to \mathbb{R}$. By Fatou's lemma

$$\liminf \mathrm{E}_{\mathbb{P}_{n_{k_l}}} \left[f \left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m} \right) Y_{n_{k_l}} \right] \geq \mathrm{E} \left[f \left(Z_3, \ldots, Z_m \right) W(\{n_k\}) \right]. \tag{2.4.4}$$

However for any constant ξ we have

$$\xi = \xi \mathrm{E}_{\mathbb{P}_{n_{k_l}}} \left[Y_{n_{k_l}} \right] \to \xi \mathrm{E}[W(\{n_k\})] = \xi$$

from (2.4.3).

So (2.4.4) holds for any bounded continuous function f. On the other hand replacing f by $-f$ we have

$$\lim \mathrm{E}_{\mathbb{P}_{n_{k_l}}} \left[f \left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m} \right) Y_{n_{k_l}} \right] = \mathrm{E} \left[f \left(Z_3, \ldots, Z_m \right) W(\{n_k\}) \right]. \tag{2.4.5}$$

Now applying condition ii) we have

$$\int f \left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m} \right) Y_{n_{k_l}} d\mathbb{P}_{n_{k_l}} = \int f \left(X_{n_{k_l},3}, \ldots, X_{n_{k_l},m} \right) d\mathbb{Q}_{n_{k_l}} \to \int f(Z'_3, \ldots, Z'_m) dQ. \tag{2.4.6}$$

Here Q is the measure induced by $(Z'_3, \ldots, Z'_m)$. In particular, one can take the measure Q such that $(Z_3, \ldots, Z_m)$ themselves are distributed as $(Z'_3, \ldots, Z'_m)$ under the measure Q. This is true due to the following observation.

$$\int f(Z'_3, \ldots, Z'_m) dQ = \mathrm{E} \left[f(Z_3, \ldots, Z_m) W^{(m)} \right]$$

for any bounded continuous function f. Since f is any bounded continuous function, we have

$$\int_A dQ = \mathrm{E} \left[\mathbb{I}_A W^{(m)} \right]$$

for any $A \in \sigma(Z_3, \ldots, Z_m)$. Here for any set A, $\mathbb{I}_A$ denotes the indicator function taking value one on A.

Now looking back into (2.4.5), we have for any $A \in \sigma(Z_3, \ldots, Z_m)$,

$$\mathrm{E}\left[\mathbb{I}_A W^{(m)}\right] = \mathrm{E}\left[\mathbb{I}_A W(\{n_k\})\right].$$

Since $W^{(m)}$ is $\sigma(Z_3, \ldots, Z_m)$ measurable, we have $W^{(m)} = \mathrm{E}\left[W(\{n_k\}) \mid \sigma(Z_3, \ldots, Z_m)\right]$

From Fatou's lemma

$$\mathrm{E}[W(\{n_k\})^2] \leq \liminf_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}[Y_n^2] = \exp\left\{\sum_{i=3}^{\infty} \frac{t^i}{2i}\right\}.$$

As a consequence, we have

$$0 \leq \mathrm{E}|W(\{n_k\}) - W^{(m)}|^2 = \mathrm{E}[W(\{n_k\})^2] - \mathrm{E}[W^{(m)2}] < \varepsilon.$$

So $W_2(F^{W^{(m)}}, F^{W(\{n_k\})}) < \sqrt{\varepsilon}$. Here $F^{W^{(m)}}$ and $F^{W(\{n_k\})}$ denote the distribution functions corresponding to $W^{(m)}$ and $W(\{n_k\})$ respectively. As a consequence, $W_2(F^{W^{(m)}}, F^{W(\{n_k\})}) \to 0$ as $m \to \infty$. Hence by Proposition **??**, $W^{(m)} \xrightarrow{d} W(\{n_k\})$.

On the other hand, we have already proved $W^{(m)}$ converge to W in L^2. So $W(\{n_k\}) \stackrel{d}{=} W$.

In Step 1 and Step 2 we verified all the conditions required to use Proposition **??**. Now using Proposition **??** the proof of Proposition **??** is complete. □

Remark 2. One might observe that the second part in assumption *ii*) of Proposition 9 is slightly weaker than (A2) in Theorem 1 of Janson (1995). For our purpose this is sufficient since we use the fact that $Y_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$. However, in Theorem 1 of Janson (1995) Y_n can be any random variable.

2.5. Signed cycles and their asymptotic distributions

We have discussed in the introduction that the proof of Mossel et al. (2015) crucially used the fact that the asymptotic distribution of short cycles turn out to be Poisson. However, in the denser case one does not get a Poisson limit for the short cycles. So their proof does not work in the denser case. Here we consider instead the "signed cycles" defined as follows:

Definition 10. For a random graph G the signed cycle of length k is defined to be:

$$C_{n,k}(G) = \left(\frac{1}{\sqrt{np_{n,\text{av}}(1-p_{n,\text{av}})}}\right)^k \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1} - p_{n,\text{av}})\ldots(x_{i_{k-1}i_0} - p_{n,\text{av}})$$

where $i_0, i_1, \ldots, i_{k-1}$ are all distinct and $p_{n,\text{av}}$ is the average connection probability i.e. $p_{n,\text{av}} = \frac{1}{n(n-1)}\sum_{i\neq j}\text{E}[x_{i,j}]$. Observe that for $\mathcal{G}(n,p_n,q_n)$, $p_{n,\text{av}}$ is equal to $\hat{p}_n$.

One should note that when $k=3$ a similar kind of random variable was called "signed triangle" in Bubeck et al. (2014)

It is intuitive that one might expect asymptotic normal distribution for $C_{n,k}$'s when $n\to\infty$ and $\hat{p}_n$ is sufficiently large. Our next result formalizes this intuition.

Proposition 1. *i) When $G\sim\mathbb{P}'_n$, $n(p_n+q_n)\to\infty$ and*
$3\le k_1<\ldots<k_l = o(\log(\hat{p}_n n))$,

$$\left(\frac{C_{n,k_1}(G)}{\sqrt{2k_1}},\ldots,\frac{C_{n,k_l}(G)}{\sqrt{2k_l}}\right)\stackrel{d}{\to} N_l(0,I_l). \tag{2.5.1}$$

ii) When $G\sim\mathbb{P}_n$, $np_n\to\infty$, $c_n\to c\in(0,\infty)$ and
$3\le k_1<\ldots<k_l = o\left(\min(\log(\hat{p}_n n),\sqrt{\log(n)})\right)$,

$$\left(\frac{C_{n,k_1}(G)-\mu_1}{\sqrt{2k_1}},\ldots,\frac{C_{n,k_l}(G)-\mu_l}{\sqrt{2k_l}}\right)\stackrel{d}{\to} N_l(0,I_l) \tag{2.5.2}$$

where $\mu_i = \left(\sqrt{\frac{c_n}{2(1-\hat{p}_n)}}\right)^{k_i}$ for $1\le i\le m$.

The proof of Proposition 1 is inspired from the remarkable paper by Anderson and Zeitouni (2006). However, the model in this case is simpler which makes the proof less cumbersome. The fundamental idea is to prove that the signed cycles converge in distribution by using the method of moments and the limiting random variables satisfy the Wick's formula. At first we state the method of moments.

Lemma 1. *Let* $(Y_{n,1}, \ldots, Y_{n,l})$ *be a sequence of random vectors of* l *dimension. Then* $(Y_{n,1}, \ldots, Y_{n,l}) \xrightarrow{d} (Z_1, \ldots, Z_l)$ *if the following conditions are satisfied:*

i)

$$\lim_{n\to\infty} \mathrm{E}[X_{n,1} \ldots X_{n,m}] \tag{2.5.3}$$

exists for any fixed m *and* $X_{n,i} \in \{Y_{n,1}, \ldots, Y_{n,l}\}$ *for* $1 \le i \le m$.

ii) (Carleman's Condition)Carleman (1926)

$$\sum_{h=1}^{\infty} \Big(\lim_{n\to\infty} \mathrm{E}[X_{n,i}^{2h}]\Big)^{-\frac{1}{2h}} = \infty \quad \forall\ 1 \le i \le l.$$

Further,

$$\lim_{n\to\infty} \mathrm{E}[X_{n,1} \ldots X_{n,m}] = \mathrm{E}[X_1 \ldots X_m].$$

Here $X_{n,i} \in \{Y_{n,1}, \ldots, Y_{n,l}\}$ *for* $1 \le i \le m$ *and* X_i *is the in distribution limit of* $X_{n,i}$. *In particular, if* $X_{n,i} = Y_{n,j}$ *for some* $j \in \{1, \ldots, l\}$ *then* $X_i = Z_j$.

The method of moments is very well known and much useful in probability theory. We omit its proof.

Now we state the Wick's formula for Gaussian random variables which was first proved by Isserlis (1918) and later on introduced by Wick (1950) in the physics literature in 1950.

Lemma 2. *(Wick's formula)Wick (1950) Let* $(Y_1, \ldots, Y_l)$ *be a multivariate mean* 0 *random vector of dimension* l *with covariance matrix* Σ*(possibly singular). Then* $((Y_1, \ldots, Y_l))$ *is jointly Gaussian if and only if for any integer* m *and* $X_i \in \{Y_1, \ldots, Y_l\}$ *for* $1 \le i \le m$

$$\mathrm{E}[X_1 \ldots X_m] = \begin{cases} \sum_\eta \prod_{i=1}^{\frac{m}{2}} \mathrm{E}[X_{\eta(i,1)} X_{\eta(i,2)}] & \text{for } m \text{ even} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{2.5.4}$$

Here η *is a partition of* $\{1, \ldots, m\}$ *into* $\frac{m}{2}$ *blocks such that each block contains exactly* 2 *elements and $\eta(i,j)$ denotes the j th element of the i th block of η for $j = 1, 2$.*

The proof of the aforesaid Lemma is omitted. However, it is good to note that the random variables $Y_1, \ldots, Y_l$ may also be the same. In particular, taking $Y_1 = \cdots = Y_l$, Lemma 2 also provides a description of the moments of Gaussian random variables. With Lemma 1 and 2 in hand, we now jump into the proof of Proposition 1.

Proof of Proposition 1

At first we introduce some notations and some terminologies. We denote a word w to be an ordered sequence of integers (to be called letters) $(i_0, \ldots, i_{k-1}, i_k)$ such that $i_0 = i_k$ and all the numbers i_j for $0 \le j \le k-1$ are distinct. For a word $w = (i_0, \ldots, i_{k-1}, i_k)$, its length $l(w)$ is $k+1$. The graph induced by a word w is denoted by G_w and defined as follows. One treats the letters $(i_0, \ldots, i_k)$ as nodes and puts an edge between the nodes $(i_j, i_{j+1})_{0 \le j \le k-1}$. Note that for a word w of length $k+1$, $G_w = (V_w, E_w)$ is just a k cycle. For a word $w = (i_0, \ldots, i_k)$ its mirror image is defined by $\tilde{w} = (i_0, i_{k-1}, i_{k-2}, \ldots, i_1, i_0)$. Further for a cyclic permutation τ of the set $\{0, 1, \ldots, k-1\}$, we define $w^\tau := (i_{\tau(0)}, \ldots, i_{\tau(k-1)}, i_{\tau(0)})$. Finally two words w and x are called paired if there is a cyclic permutation τ such that either $x^\tau = w$ or $\tilde{x}^\tau = w$. An ordered tuple of m words, $(w_1, \ldots, w_m)$ will be called a sentence. For any sentence $a = (w_1, \ldots w_m)$, $G_a = (V_a, E_a)$ is the graph with $V_a = \cup_{i=1}^m V_{w_i}$ and $E_a = \cup_{i=1}^m E_{w_i}$.

Proof of part i) We complete the proof of this part in two steps. In the first step the asymptotic variances of $(C_{n,k_1}(G), \ldots, C_{n,k_l}(G))$ will be calculated and the second step will be dedicated towards proving the asymptotic normality and independence of $(C_{n,k_1}(G), \ldots, C_{n,k_l}(G))$.

Step 1: Observe that when $G \sim \mathbb{P}'_n$, the distribution of $C_{n,k_1}(G), \ldots, C_{n,k_l}(G)$ is trivially independent of the labels σ_i and $\mathrm{E}[C_{n,k}(G)] = 0$ for any k. Now we prove that $\mathrm{Var}(C_{n,k}(G)) \sim 2k$ for any $k = o(\sqrt{n})$. Let for any word $w = (i_0, \ldots, i_k)$, $X_w := \prod_{j=0}^{k-1} \left(x_{i_j, i_{j+1}} - \hat{p}_n \right)$. Now observe that

$$\begin{aligned}\mathrm{Var}(C_{n,k}) &= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k \mathrm{E}\left[(\sum_w X_w)^2\right] \\ &= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k \mathrm{E}\left[\sum_{w,x} X_w X_x\right].\end{aligned} \tag{2.5.5}$$

Since both X_w and X_x are product of independent mean 0 random variables each coming exactly once, $\mathrm{E}[X_w X_x] \neq 0$ if and only if all the edges in G_w are repeated in G_x. Observe that since G_w and G_x are cycles of length k, this is satisfied if and only if w and x are paired. There are k many cyclic permutations τ of the set $\{0, \ldots, k-1\}$ and for a given w and τ, there are only two possible choices of x such that w and x are paired. These choices are obtained when $x^\tau = w$ and $\tilde{x}^\tau = w$. As a consequence for any word w, exactly $2k$ words are paired with it. Now observe that when w and x are paired, $X_w X_x$ is a product of k random variables each appearing exactly twice. As a consequence, $\mathrm{E}[X_w X_x] = (\hat{p}_n(1-\hat{p}_n))^k$. Also the total number of words is given by $n(n-1)\ldots(n-k+1)$ for the choices of $i_0, \ldots, i_{k-1}$. It is well known that

$$\frac{n(n-1)\ldots(n-k+1)}{n^k} \to 1$$

whenever $k = o(\sqrt{n})$. So

$$\mathrm{Var}(C_{n,k}) = 2k\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k n(n-1)\ldots(n-k+1)\left(\hat{p}_n(1-\hat{p}_n)\right)^k \sim 2k \tag{2.5.6}$$

as long as $k = o(\sqrt{n})$. This completes **Step 1** of the proof.

Step 2: Now we claim that in order to complete **Step 2**, is enough to prove the following two limits.

$$\lim_{n\to\infty} \mathrm{E}\left[C_{n,k_1}(G)C_{n,k_2}(G)\right] \to 0 \tag{2.5.7}$$

whenever $k_1 \neq k_2$ and there exists random variables $Z_1, \ldots, Z_l$ such that for any fixed m

$$\lim_{n\to\infty} \mathrm{E}[X_{n,1}\ldots X_{n,m}] \to \begin{cases} \sum_\eta \prod_{i=1}^{\frac{m}{2}} \mathrm{E}[Z_{\eta(i,1)}Z_{\eta(i,2)}] & \text{for } m \text{ even} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{2.5.8}$$

where $X_{n,i} \in \{\frac{C_{n,k_1}(G)}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l}(G)}{\sqrt{2k_l}}\}$.

First observe that (2.5.8) will simultaneously imply part $i)$ and $ii)$ of Lemma 1. Implication of $i)$ is obvious. However, for $ii)$ one can take $X_{n,i}$'s to be all equal and from Wick's formula (Lemma 2) the limiting distribution of $X_{n,i}$'s are normal. It is well known that normal random variables satisfy Carleman's condition. On the other hand (2.5.8) also implies that the limit of $(\frac{C_{n,k_1}(G)}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l}(G)}{\sqrt{2k_l}})$ is jointly normal. Hence applying (2.5.7), one gets the asymptotic independence.

We first prove (2.5.7). Observe that

$$\mathrm{E}\left[C_{n,k_1}(G)C_{n,k_2}(G)\right] = \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k_1+k_2}{2}} \mathrm{E}\left[\sum_{w,x} X_w X_x\right].$$

However, here $l(w) = k_1 + 1$ and $l(x) = k_2 + 1$. So $\mathrm{E}\left[\sum_{w,x} X_w X_x\right] = 0$. As a consequence, (2.5.7) holds.

Now we prove (2.5.8). Let l_i be the length of any word corresponding to $X_{n,i}$. Observe that $l_i \in \{k_1+1, \ldots, k_l+1\}$ for any i. At first we expand the left hand side of (2.5.8).

$$\mathrm{E}[X_{n,1}\ldots X_{n,m}] = \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{w_1,\ldots,w_m} E\left[X_{w_1}\ldots X_{w_m}\right]. \tag{2.5.9}$$

Here the graphs $G_{w_1}, \ldots, G_{w_m}$ are cycles of length $l_1 - 1, \ldots, l_m - 1$ respectively. So in order to have $E\left[X_{w_1}\ldots X_{w_m}\right] \neq 0$, we need each of the edges in $G_{w_1}, \ldots, G_{w_m}$ to be traversed more than once. The sentence $a := (w_1, \ldots, w_m)$ formed by such $(w_1, \ldots, w_m)$ will be called a weak CLT sentence. Given a weak CLT sentence a, we introduce a partition $\eta(a)$, of $\{1, \ldots, m\}$ in the following way. If i, j are in same block of the partition $\eta(a)$, then G_{w_i} G_{w_j} have at least one edge in common.

As a consequence, we can further write the left hand side of (2.5.9) in the following way.

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\ldots,w_m \ | \ \eta=\eta(w_1,\ldots,w_m)} E\left[X_{w_1} \ldots X_{w_m}\right]. \tag{2.5.10}$$

Observe that each block in η should have at least 2 elements. Otherwise, in this case $E\left[X_{w_1} \ldots X_{w_m}\right] = 0$. As a consequence, the number of blocks in $\eta \leq [\frac{m}{2}]$.

Now we prove that if the number of blocks in $\eta < \frac{m}{2}$, then

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\ldots,w_m \ | \ \eta=\eta(w_1,\ldots,w_m)} E\left[X_{w_1} \ldots X_{w_m}\right] \to 0.$$

If $\eta(w_1, \ldots, w_m)$ have strictly less than $\frac{m}{2}$ blocks, then a has strictly less than $\frac{m}{2}$ connected components. From Proposition 4.9 and Lemma 4.10 of Anderson and Zeitouni (2006) it follows that in this case $\#V_a < \sum_{i=1}^{m} \frac{l_i-1}{2}$. However each connected component is formed by a union of several cycles so $V_a \leq E_a$. Now the following lemma gives a bound on the number of weak CLT sentences having strictly less than $\frac{m}{2}$ connected components.

Lemma 3. *Let $\mathcal{A}$ be the set of weak CLT sentences such that for each $a \in \mathcal{A}$, $\#V_a = t$. Then*

$$\#\mathcal{A} \leq 2^{\sum_i l_i} \left(C_1 \sum_i l_i\right)^{C_2 m} \left(\sum_i l_i\right)^{3(\sum_i l_i - 2t)} n^t. \tag{2.5.11}$$

The proof of Lemma 3 is rather technical and requires some amount of random matrix theory. So we defer its proof to the appendix. However, assuming Lemma 3, we have

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i(l_i-1)}{2}} \sum_{a\,:\,V_a<\sum_{i=1}^m \frac{l_i-1}{2}} E\left[X_{w_1}\dots X_{w_m}\right]$$
$$\leq \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i(l_i-1)}{2}} \sum_{t<\frac{\sum_i(l_i-1)}{2}} \sum_{e=t}^{\sum_i\frac{(l_i-1)}{2}} 2^{\sum_i l_i}\left(C_1\sum_i l_i\right)^{C_2 m}\left(\sum_i l_i\right)^{3(\sum_i l_i-2t)} n^t\hat{p}_n^e. \tag{2.5.12}$$

Now observe that $\sum_{e=t}^{\infty}\hat{p}_n^e \leq \frac{1}{1-\hat{p}_n}\hat{p}_n^t$. As we consider $p<1$, we have for large enough n, $\frac{1}{1-\hat{p}_n}\leq D$ for some deterministic constant D. Plugging in this estimate in (2.5.12) we have the first expression in (2.5.12) is lesser or equal to

$$D\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i(l_i-1)}{2}} \sum_{t<\frac{\sum_i(l_i-1)}{2}} 2^{\sum_i l_i}\left(C_1\sum_i l_i\right)^{C_2 m}\left(\sum_i l_i\right)^{3m}\left(\sum_i l_i\right)^{3((\sum_i l_i-1)-2t)} n^t\hat{p}_n^t.$$
$$\leq D\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{\sum_i(l_i-1)} 2^m C_1^{C_2 m}\left(\sum_i l_i\right)^{(C_2+3)m} \underbrace{\sum_{t<\frac{\sum_i(l_i-1)}{2}}\left(\frac{(\sum_i l_i)^6}{n\hat{p}_n}\right)^{\sum_i\frac{l_i-1}{2}-t}}_{T_1((say))}. \tag{2.5.13}$$

Observe that T_1 is just a geometric series. When $k_l = o(\log(\hat{p}_n n))$ we have,

$$\left(\frac{(\sum_i l_i)^6}{n\hat{p}_n}\right)\leq\frac{(mk_l)^6}{n\hat{p}_n}\to 0.$$

Now, the lowest value of $\sum_{i=1}^m(l_i-1)-2t$ is 1. As the geometric series $\sum_{j=1}^{\infty}\kappa^j$, for $\kappa<1$ is comparable to its first term, we can give the following final bound to (2.5.12),

$$C_3\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{\sum_i(l_i-1)} 2^m C_1^{C_2 m}\left(\sum_i l_i\right)^{(C_2+3)m}\frac{(\sum_i l_i)^3}{\sqrt{n\hat{p}_n}}. \tag{2.5.14}$$

Here C_3 is a universal constant. Observe that the dominant term in the numerator of

(2.5.14) is

$$\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{\sum_i (l_i-1)} \leq \left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right)^{m(k_l-1)}.$$

However from our assumption $m(k_l-1)\log\left(\frac{2}{\sqrt{(1-\hat{p}_n)}}\right) - \frac{1}{2}\log(n\hat{p}_n) \to -\infty$. As a consequence, the first expression in (2.5.12) goes to 0.

Once this is proved all the other partitions left are pair partitions i.e. it has exactly $\frac{m}{2}$ many blocks. In particular, m is even. We now fix a partition η of this kind. Let for any $i \in \{1, \ldots, \frac{m}{2}\}$, $\eta(i,1) < \eta(i,2)$ be the elements in the i th block. Observe now that fixing a pair partition η and $(w_1, \ldots, w_m)$ such that $\eta(w_1, \ldots, w_m) = \eta$, the random variables $X_{w_{\eta(i_1,j)}}$ and $X_{w_{\eta(i_2,j)}}$ are independent when ever $i_1 \neq i_2$ for any $j \in \{1,2\}$. As a consequence, we now can rewrite (2.5.10) as follows:

$$\begin{aligned}
&\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\ldots,w_m \ \mid\ \eta=\eta(w_1,\ldots,w_m)} E\left[X_{w_1}\ldots X_{w_m}\right] \\
&= o(1) + \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i (l_i-1)}{2}} \sum_{\eta \ \mid\ \eta \text{ pair parition}} \sum_{w_1,\ldots,w_m \ \mid\ \eta=\eta(w_1,\ldots,w_m)} \prod_{i=1}^{\frac{m}{2}} \mathrm{E}\left[X_{w_{\eta(i,1)}} X_{w_{\eta(i,2)}}\right]
\end{aligned} \tag{2.5.15}$$

Now observe that whenever $\prod_{i=1}^{\frac{m}{2}} \mathrm{E}\left[X_{w_{\eta(i,1)}} X_{w_{\eta(i,2)}}\right] \neq 0$, we have $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$ are paired. In particular $l(w_{\eta(i,1)}) = l(w_{\eta(i,2)})$ and there are $(1+o(1))(2(l_{\eta(i,1)}-1))n^{l_{\eta(i,1)}-1}$ many such choices of $(w_{\eta(i,1)}, w_{\eta(i,2)})$ for every i. Here $l_{\eta(i,1)}$ is the common length of the words $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$. On the other hand, in this case $\mathrm{E}\left[X_{w_{\eta(i,1)}} X_{w_{\eta(i,2)}}\right] = (\hat{p}_n(1-\hat{p}_n))^{l_{\eta(i,1)}-1}$.

Hence, we get the following final reduction to (2.5.15):

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i(l_i-1)}{2}} \sum_{\eta} \sum_{w_1,\dots,w_m \mid \eta=\eta(w_1,\dots,w_m)} E\left[X_{w_1}\dots X_{w_m}\right]$$

$$= o(1) + (1+o(1))$$
$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{\sum_i(l_i-1)}{2}} \tag{2.5.16}$$
$$\sum_{\eta \mid \eta \text{ pair parition}} \prod_{i=1}^{\frac{m}{2}} 2(l_{\eta(i,1)}-1)\mathbb{I}_{l_{\eta(i,1)}=l_{\eta(i,2)}} n^{\frac{\sum_i(l_i-1)}{2}} \left(\hat{p}_n(1-\hat{p}_n)\right)^{\frac{\sum_i(l_i-1)}{2}}$$
$$= o(1) + (1+o(1)) \sum_{\eta \mid \eta \text{ pair parition}} \prod_{i=1}^{\frac{m}{2}} 2(l_{\eta(i,1)}-1)\mathbb{I}_{l_{\eta(i,1)}=l_{\eta(i,2)}}.$$

This completes the proof. □

Proof of part ii) We now give a proof of part ii) of Proposition 1. Recall that $d_n = \frac{p_n-q_n}{2}$. We have

$$C_{n,k}(G) = \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\dots,i_{k-1}} (x_{i_0,i_1}-\hat{p}_n)\dots(x_{i_{k-1}i_0}-\hat{p}_n)$$
$$= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\dots,i_{k-1}} (x_{i_0,i_1}-p_{i_0,i_1}+p_{i_0,i_1}-\hat{p}_n)\dots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}+p_{i_{k-1},i_k}-\hat{p}_n)$$
$$= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\dots,i_{k-1}} (x_{i_0,i_1}-p_{i_0,i_1}+\sigma_{i_0}\sigma_{i_1}d_n)\dots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}+\sigma_{i_{k-1}}\sigma_{i_k}d_n)$$
$$= \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\dots,i_{k-1}} \left[(x_{i_0,i_1}-p_{i_0,i_1})\dots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}) + d_n^k \prod_{j=0}^{k-1}\sigma_{i_j}\sigma_{i_{j+1}}\right] + V_{n,k} \tag{2.5.17}$$

where $p_{i,j} = p_n$ if $\sigma_i = \sigma_j$ and q_n otherwise. Here $V_{n,k}$ is obtained by taking the sum of all the remaining terms in the expansion of

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\dots,i_{k-1}} (x_{i_0,i_1}-p_{i_0,i_1}+\sigma_{i_0}\sigma_{i_1}d_n)\dots(x_{i_{k-1}i_0}-p_{i_{k-1},i_k}+\sigma_{i_{k-1}}\sigma_{i_k}d_n)$$

apart from

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1} - p_{i_0,i_1}) \ldots (x_{i_{k-1}i_0} - p_{i_{k-1},i_k})$$

and

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} d_n^k \prod_{j=0}^{k-1} \sigma_{i_j}\sigma_{i_{j+1}}.$$

At first we prove that

$$\prod_{j=0}^{k-1} \sigma_{i_j}\sigma_{i_{j+1}} = 1 \tag{2.5.18}$$

irrespective of the values of σ_i's. The proof of this is straight forward since $i_0 = i_k$ we have

$$\prod_{j=0}^{k-1} \sigma_{i_j}\sigma_{i_{j+1}} = \prod_{j=0}^{k-1} \sigma_j^2 = 1.$$

As $d_n = \sqrt{\frac{c_n\hat{p}_n}{2n}}$, we have

$$\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \sum_{i_0,i_1,\ldots,i_{k-1}} d_n^k = (1+o(1))\frac{d_n^k n^k}{(n\hat{p}_n(1-\hat{p}_n))^{\frac{k}{2}}} = (1+o(1))\left(\sqrt{\frac{c_n}{2(1-\hat{p}_n)}}\right)^k.$$

This explains the mean term. The proof of asymptotic normality and independence of

$$D_{n,k}(G) := \left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} \left[\sum_{i_0,i_1,\ldots,i_{k-1}} (x_{i_0,i_1} - p_{i_0,i_1}) \ldots (x_{i_{k-1}i_0} - p_{i_{k-1},i_k})\right]$$

is exactly same as part i). We only note that here the variance is also $2k$. To see this, we have

$$d_n = \sqrt{\frac{c_n\hat{p}_n}{2n}}$$

and whenever, $k = o(\log(\hat{p}_n n))$ both

$$\lim_{n\to\infty} \left(\frac{(\hat{p}_n + d_n)(1-\hat{p}_n - d_n)}{\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}} = 1 \tag{2.5.19}$$

and

$$\lim_{n\to\infty}\left(\frac{(\hat{p}_n-d_n)(1-\hat{p}_n+d_n)}{\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}}=1. \tag{2.5.20}$$

It is easy to see that $\mathrm{Var}\left(\frac{D_{n,k}(G)}{\sqrt{2k}}\right)$ lies between the left hand side of (2.5.19) and (2.5.20). As a consequence, $\mathrm{Var}\left(\frac{D_{n,k}(G)}{\sqrt{2k}}\right)\to 1$.

It is easy to observe that $\mathrm{E}\left[V_{n,k}\right]$ is always 0. Now our final task is to prove $\mathrm{Var}(V_{n,k})\to 0$. This will prove that $V_{n,k}\xrightarrow{p}0$ and the proof will be completed.

Let us fix a word w and let $\emptyset\subsetneq E_f\subsetneq E_w$ be any subset. Then

$$V_{n,k}=\sum_w V_{n,k,w}$$

where

$$V_{n,k,w}:=\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^{\frac{k}{2}}\sum_{\emptyset\subsetneq E_f\subsetneq E_w}\prod_{e\in E_f}\sigma_e d_n\prod_{e\in E\backslash E_f}(x_e-p_e).$$

Here for any edge i,j, $x_e=x_{i,j}$, $p_e=p_{i,j}$ and $\sigma_e=\sigma_i\sigma_j$. Now

$$\mathrm{Var}(V_{n,k})=\sum_{w,x}\mathrm{Cov}(V_{n,k,w},V_{n,k,x}).$$

We now find an upper bound of $\mathrm{Cov}(V_{n,k,w},V_{n,k,x})$.

At first fix any word w and the set $\emptyset\subsetneq E_f\subsetneq E_w$ and consider all the words x such that $E_w\cap E_x=E_w\backslash E_f$. As every edge in G_w and G_x appear exactly once,

$$\begin{aligned}
\mathrm{Cov}(V_{n,k,w},V_{n,k,x}) &= \sum_{E_w\backslash E'\subset E_w\backslash E_f}\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k\prod_{e\in E'}(\pm d_n^2)\mathrm{E}\prod_{e\in E_w\backslash E'}(x_e-p_e)^2\\
&= \sum_{E_w\backslash E'\subset E_w\backslash E_f}\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k(\pm d_n^{2\#E'})(1+o(1))\left(\hat{p}_n(1-\hat{p}_n)\right)^{k-\#E'}\\
&\le \sum_{E_w\backslash E'\subset E_w\backslash E_f}(1+o(1))\left(\frac{1}{n\hat{p}_n(1-\hat{p}_n)}\right)^k\left(\frac{c_n}{2}\right)^{\#E'}\left(\frac{\hat{p}_n}{n}\right)^{\#E'}\hat{p}_n^{k-\#E'}\\
&\le (C)^k\frac{1}{n^{k+\#E_f}}
\end{aligned} \tag{2.5.21}$$

where C is some known constant. The last inequality holds since $\#E'\ge\#E_f$ and

$\#(E_w \backslash E' \subset E_w \backslash E_f) \leq 2^k$.

Observe that the graph corresponding to the edges $E_w \backslash E_f$ is a disjoint collection of straight lines. Let the number of such straight lines be ζ. Obviously $\zeta \leq \#(E_w \backslash E_f)$. The number of ways these ζ components can be placed in x is bounded by $k^\zeta \leq k^{\#(E_w \backslash E_f)}$ and all other nodes in x can be chosen freely. So there are at most $n^{k - \#V_{E_w \backslash E_f}} k^{\#(E_w \backslash E_f)}$ choices of such x. Here $V_{E_w \backslash E_f}$ is the set of vertices of the graph corresponding to $(E_w \backslash E_f)$. Observe that, whenever $k > \#E_f > 0$, $E_w \backslash E_f$ is a forest so

$$\#V_{E_w \backslash E_f} \geq \#(E_w \backslash E_f) + 1 \Leftrightarrow k - \#V_{E_w \backslash E_f} \leq \#E_f - 1.$$

As a consequence,

$$\sum_{x \ | \ E_w \cap E_x = E_w \backslash E_f} \mathrm{Cov}(V_{n,k,w}, V_{n,k,x}) \leq (C)^k \frac{1}{n^{k + \#E_f}} n^{\#E_f - 1} k^{\#(E_w \backslash E_f)} \leq (C)^k \frac{1}{n^{k+1}} k^k. \tag{2.5.22}$$

The right hand side of (2.5.22) does not depend on E_f and there are at most 2^k nonempty subsets E_f of E^w. So

$$\sum_x \mathrm{Cov}(V_{n,k,w}, V_{n,k,x}) \leq (2C)^k k^k \frac{1}{n^{k+1}}.$$

Finally there are at most n^k many w. So

$$\sum_w \sum_x \mathrm{Cov}(V_{n,k,w}, V_{n,k,x}) \leq (2C)^k k^k \frac{1}{n}. \tag{2.5.23}$$

Now we use the fact $k = o(\sqrt{\log(n)})$. In this case

$$k \log(2C) + k \log(k) \leq \sqrt{log(n)} \log(\sqrt{\log n}) = o(log(n)) \Leftrightarrow (2C)^k k^k = o(n).$$

This concludes the proof. □

2.6. Calculation of second moment and completion of the proof of Theorem 2

With Propositions 9 and 1 in hand the rest of the proof of Theorem 2 should be very straight forward. We at first prove that $\lim_{n\to\infty} \mathbb{E}\left(\frac{d\mathbb{P}_n}{d\mathbb{P}'_n}\right)^2$ is the right hand side of (2.4.1) with $t = \frac{c}{2(1-p)}$ whenever $\frac{a_n}{n} \to p \in [0, 1)$.

Lemma 4. *Let $Y_n := \frac{d\mathbb{P}_n}{d\mathbb{P}'_n}$. Whenever $p_n \to p \in [0, 1)$, we have*

$$\mathrm{E}_{\mathbb{P}'_n}[Y_n^2] \to \exp\left\{-\frac{t}{2} - \frac{t^2}{4}\right\} \frac{1}{\sqrt{1-t}}, \quad t = \frac{c}{2(1-p)} < 1.$$

Proof. The proof of Lemma 4 is similar to the proof of Lemma 5.4. in Mossel et al. (2015). The notations used in this proof are slightly different from that of Lemma 5.4 in Mossel et al. (2015) for understanding case when p is not necessarily 0.

At first we introduce some notations. Given a labeled graph (G, σ) we define

$$W_{uv} = W_{uv}(G, \sigma) = \begin{cases} \frac{p_n}{\hat{p}_n} & \text{if } \sigma_u\sigma_v = 1 \text{ and } (u, v) \in E \\ \frac{q_n}{\hat{p}_n} & \text{if } \sigma_u\sigma_v = -1 \text{ and } (u, v) \in E \\ \frac{1-p_n}{1-\hat{p}_n} & \text{if } \sigma_u\sigma_v = 1 \text{ and } (u, v) \notin E \\ \frac{1-q_n}{1-\hat{p}_n} & \text{if } \sigma_u\sigma_v = -1 \text{ and } (u, v) \notin E \end{cases} \tag{2.6.1}$$

and define V_{uv} by the same formula, but with σ replaced by τ. Now

$$Y_n = \frac{1}{2^n} \sum_{\sigma\in\{1,-1\}^n} \prod_{(u,v)} W_{uv}$$

and

$$Y_n^2 = \frac{1}{2^{2n}} \sum_{\sigma,\tau} \prod_{(u,v)} W_{uv}V_{uv}.$$

Since $\{W_{uv}\}$ are independent given σ, it follows that

$$\mathrm{E}_{\mathbb{P}'_n}(Y_n^2) = \frac{1}{2^{2n}} \sum_{\sigma,\tau} \prod_{(u,v)} \mathrm{E}_{\mathbb{P}'_n}\left(W_{uv}V_{uv}\right).$$

Now we consider the following cases:

1. $\sigma_u\sigma_v = 1$ and $\tau_u\tau_v = 1$.

2. $\sigma_u\sigma_v = -1$ and $\tau_u\tau_v = -1$.

3. $\sigma_u\sigma_v = 1$ and $\tau_u\tau_v = -1$.

4. $\sigma_u\sigma_v = -1$ and $\tau_u\tau_v = 1$.

Let $t = \frac{c}{2(1-p)}$. We at first calculate $\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv})$ for cases 1 and 3.

Case 1:

$$\begin{aligned}
\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv}) &= \left(\frac{p_n}{\hat{p}_n}\right)^2 \hat{p}_n + \left(\frac{1-p_n}{1-\hat{p}_n}\right)^2 (1-\hat{p}_n). \\
&= \frac{p_n^2}{\hat{p}_n} + \frac{(1-p_n)^2}{1-\hat{p}_n} \\
&= \frac{(\hat{p}_n+d_n)^2}{\hat{p}_n} + \frac{(1-\hat{p}_n-d_n)^2}{1-\hat{p}_n} \\
&= 1 + d_n^2\left(\frac{1}{\hat{p}_n} + \frac{1}{1-\hat{p}_n}\right) = 1 + \frac{d_n^2}{\hat{p}_n(1-\hat{p}_n)} = 1 + \frac{c_n}{2n(1-\hat{p}_n)} \\
&= 1 + \frac{t_n}{n}
\end{aligned} \tag{2.6.2}$$

where $d_n = \frac{p_n - q_n}{2}$ and $t_n = \frac{c_n}{2(1-\hat{p}_n)} = (1+o(1))t$ as before.

Case 3:

$$\begin{aligned}
\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv}) &= \left(\frac{p_n}{\hat{p}_n}\cdot\frac{q_n}{\hat{p}_n}\right) \hat{p}_n + \left(\frac{1-p_n}{1-\hat{p}_n}\cdot\frac{1-q_n}{1-\hat{p}_n}\right) (1-\hat{p}_n). \\
&= \frac{p_nq_n}{\hat{p}_n} + \frac{(1-p_n)(1-q_n)}{1-\hat{p}_n} \\
&= \frac{(\hat{p}_n+d_n)(\hat{p}_n-d_n)}{\hat{p}_n} + \frac{(1-\hat{p}_n-d_n)(1-\hat{p}_n+d_n)}{1-\hat{p}_n} \\
&= 1 - d_n^2\left(\frac{1}{\hat{p}_n} + \frac{1}{1-\hat{p}_n}\right) = 1 - \frac{d_n^2}{\hat{p}_n(1-\hat{p}_n)} = 1 - \frac{t_n}{n}
\end{aligned} \tag{2.6.3}$$

It is easy to observe that $\mathrm{E}_{\mathbb{P}'_n}(W_{uv}V_{uv}) = 1+\frac{t_n}{n}$ and $1-\frac{t_n}{n}$ for Case 2 and Case 4 respectively.

We now introduce another parameter $\rho = \rho(\sigma,\tau) = \frac{1}{n}\sum_i \sigma_i\tau_i$. Let $S_\pm$ be the number of

$\{u, v\}$ such that $\sigma_u \sigma_v \tau_u \tau_v = \pm 1$ respectively. It is easy to observe that

$$\rho^2 = \frac{1}{n} + \frac{2}{n^2}(S_+ - S_-) \tag{2.6.4}$$

and

$$1 - \frac{1}{n} = \frac{2}{n^2}(S_+ + S_-). \tag{2.6.5}$$

So

$$S_+ = (1+\rho^2)\frac{n^2}{4} - \frac{n}{2}, \quad S_- = (1-\rho^2)\frac{n^2}{4}. \tag{2.6.6}$$

Now

$$\begin{aligned} \mathrm{E}_{\mathbb{P}'_n}(Y_n^2) &= \frac{1}{2^{2n}} \sum_{\sigma,\tau} \left(1 + \frac{t_n}{n}\right)^{S_+} \left(1 - \frac{t_n}{n}\right)^{S_-} \\ &= \frac{1}{2^{2n}} \sum_{\sigma,\tau} \left(1 + \frac{t_n}{n}\right)^{(1+\rho^2)\frac{n^2}{4} - \frac{n}{2}} \left(1 - \frac{t_n}{n}\right)^{(1-\rho^2)\frac{n^2}{4}}. \end{aligned} \tag{2.6.7}$$

Observe that $t_n = (1+o(1))t$ is a bounded sequence. It is easy to check by taking logarithm and Taylor expansion that for any bounded sequence x_n,

$$\left(1 + \frac{x_n}{n}\right)^{n^2} = (1+o(1)) \exp\left\{n x_n - \frac{1}{2} x_n^2\right\}.$$

So we can write the right hand side of (2.6.7) as

$$\begin{aligned} &(1+o(1))\frac{1}{2^{2n}} \sum_{\sigma,\tau} e^{-\frac{t_n}{2}} \exp\left[\left(n t_n - \frac{t_n^2}{2}\right)\left(\frac{1+\rho^2}{4}\right)\right] \times \exp\left[\left(-n t_n - \frac{t_n^2}{2}\right)\left(\frac{1-\rho^2}{4}\right)\right] \\ =&(1+o(1))\frac{1}{2^{2n}} \sum_{\sigma,\tau} e^{-\frac{t_n}{2} - \frac{t_n^2}{4}} \exp\left[\frac{n t_n \rho^2}{2}\right] \\ =&(1+o(1)) e^{-\frac{t_n}{2} - \frac{t_n^2}{4}} \frac{1}{2^{2n}} \sum_{\sigma,\tau} \exp\left[\frac{(1+o(1)) t n \rho^2}{2}\right] \end{aligned} \tag{2.6.8}$$

From Lemma 5.5 in Mossel et al. (2015)

$$\frac{1}{2^{2n}}\sum_{\sigma,\tau}\exp\left[\frac{(1+o(1))nt\rho^2}{2}\right]\to\frac{1}{\sqrt{1-t}}.$$

So the right hand side of (2.6.8) converges to

$$\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}\frac{1}{\sqrt{1-t}}$$

as required. □

Proof of Theorem 2:

Proof of part i) We take $X_{n,i} = C_{n,i}(G)$.

At first observe that when $p_n \to p \in [0,1)$ for any fixed i, $\mu_i := \left(\sqrt{\frac{c_n}{2(1-\hat{p}_n)}}\right)^i$ converges to $\left(\sqrt{\frac{c}{2(1-p)}}\right)^i$ as $n \to \infty$.

From Proposition 1 and Lemma 1 we see that $C_{n,i}(G)$'s satisfy all the required conditions for Proposition 9. Hence $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous.

It is easy to see that the estimate $\hat{d}_n := \frac{1}{n-1}\sum_{i\neq j} x_{i,j}$ has mean $\frac{a_n+b_n}{2}$ and variance $O(\frac{a_n+b_n}{n})$. So

$$\hat{d}_n - \frac{a_n+b_n}{2} = o_p(\sqrt{a_n+b_n}) = o_p(a_n-b_n)$$

Suppose under $\mathbb{P}_n$ there exist estimators A_n of a_n and B_n of b_n such that

$$|A_n-a_n|+|B_n-b_n| = o_p(a_n-b_n).$$

Then $2(\hat{d}_n - B_n)-(a_n-b_n) = o_p(a_n-b_n)$ i.e.

$$\frac{2(\hat{d}_n-B_n)}{a_n-b_n}|\mathbb{P}_n \xrightarrow{P} 1.$$

However, from the fact that $\mathbb{P}_n$ and $\mathbb{P}'_n$ are mutually contiguous we also have

$$\frac{2(\hat{d}_n - B_n)}{a_n - b_n} | \mathbb{P}'_n \xrightarrow{P} 1$$

which is impossible.

Proof of part ii) It is easy to observe that $\mathbb{P}_n$ and $\mathbb{P}'_n$ are asymptotically singular as for any $k_n \to \infty$, $\frac{\mu_{k_n}}{\sqrt{2k_n}} \to \infty$. Now we construct estimators for a_n and b_n. Let us define

$$\hat{f}_{n,k_n} = \begin{cases} \left(\sqrt{2k_n}\tilde{C}_{n,k_n}(G)\right)^{\frac{1}{k_n}} & \text{if } \ C_{n,k_n}(G) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

where $\tilde{C}_{n,k_n}(G)$ is the sample version of C_{n,k_n} where $\hat{p}_n$ is replaced by $\hat{d}_n/n$. One can show that $C_{n,k_n}(G)$ and $\tilde{C}_{n,k_n}(G)$ has the same limiting distribution under $\mathbb{P}_n$ and $\mathbb{P}'_n$. Hence $\mathbb{P}_n$ $\hat{f}_{n,k_n} \xrightarrow{P} \sqrt{\frac{c}{2(1-p)}}$ as $k_n \to \infty$. It is easy to see that under $\mathbb{P}_n$

$$\begin{aligned} &\frac{\hat{d}_n - \frac{(a_n+b_n)}{2}}{\sqrt{a_n+b_n}} \xrightarrow{P} 0 \Rightarrow \frac{\hat{d}_n - \frac{(a_n+b_n)}{2}}{a_n+b_n} \xrightarrow{P} 0 \Rightarrow \sqrt{\frac{\hat{d}_n}{\frac{a_n+b_n}{2}}} \xrightarrow{P} 1. \\ \Rightarrow &\sqrt{\hat{d}_n} - \sqrt{\frac{a_n+b_n}{2}} = o_p(\sqrt{a_n+b_n}) = o_p(a_n - b_n) \end{aligned} \tag{2.6.9}$$

As under both $\mathbb{P}_n$ and $\mathbb{P}'_n$, $\hat{d}_n/n \to p$,

$$\sqrt{\frac{\hat{d}_n(1-\hat{d}_n/n)}{\frac{a_n+b_n}{2}(1-p)}} \xrightarrow{P} 1.$$

So

$$\sqrt{\hat{d}_n(1-\hat{d}_n/n)} - \sqrt{\frac{a_n+b_n}{2}(1-p)} = o_p(a_n-b_n) \ \forall \ p \in [0,1).$$

So $\sqrt{\hat{d}_n(1-\hat{d}_n/n)}\hat{f}_{n,k_n} - \frac{a_n-b_n}{2} = o_p(a_n-b_n)$ under $\mathbb{P}_n$. As a consequence, the estimators $\hat{A} = \hat{d}_n + \sqrt{\hat{d}_n(1-\hat{d}_n/n)}\hat{f}_{n,k_n}$ and $\hat{B} = \hat{d}_n - \sqrt{\hat{d}_n(1-\hat{d}_n/n)}\hat{f}_{n,k_n}$ have the required property. This concludes the proof. □

We end the discussion of this section by the following remark on the computation of the signed cycles.

Remark 3. In general the direct computation of the random variables $C_{n,k}(G)$'s take at least $O(n^k)$ amount of time. So it might appear that the statistics $C_{n,k}(G)$'s are not useful for any practical purpose. Fortunately, this is not the case. It was proved in Banerjee and Ma (2017a) that whenever k is odd, the difference between $C_{n,k}(G)$ and $\sum_{i=1}^{n} P_k(\lambda_i)$ converges in probability to 0 for any $k = o\left(\min(\log(\hat{p}_n n), \sqrt{\log(n)})\right)$. Here $\{\lambda_i\}_{1\leq i\leq n}$ are the eigenvalues of the centered adjacency matrix of the graph and $P_k(\cdot)$ is the Chebyshev polynomial of degree k (look at (2.7)-(2.8) in Banerjee and Ma (2017a) for definition). The case when k is even is more complicated. In this case one can prove $C_{n,2k}(G) - \sum_{i=1}^{n} P_{2k}(\lambda_i) - E_{2k} \stackrel{P}{\to} 0$ where E_{2k} is an additional error term. One can prove that $\mathrm{Var}\left[E_{2k}\right]$ converges to 0 and find the asymptotic value of $\mathrm{E}[E_{2k}]$ explicitly under additional growth conditions on $\hat{p}_n$. As a consequence, the signed cycles of growing orders can be computed by the spectral decomposition of the centered adjacency matrix of the graph. It is well known that this has $O(n^3\log(n))$ time complexity. One might check Banerjee and Ma (2017a) for details.

2.7. Proof of non reconstructability

In this section we provide a proof of the non-reconstruction results stated in Theorem 3. Our proof technique relies on fine analysis of conditional probabilities. Technically, this proof is closely related to the non-reconstruction proof in section 6.2 of Banks et al. (2016) rather than the original proof given in Mossel et al. (2015). At first we prove one Proposition and one Lemma which will be crucial for our proof.

Proposition 10. *Suppose $a_n, b_n \to \infty$, $\frac{a_n}{n} \to p \in [0,1)$, $c_n \to c$ and $c < 2(1-p)$. Then for any fixed r and any two configurations $(\sigma_1^{(1)}, \ldots, \sigma_r^{(1)})$, $(\sigma_1^{(2)}, \ldots, \sigma_r^{(2)})$*

$$\mathrm{TV}\left(\mathbb{P}_n(G|(\sigma_1^{(1)}, \ldots, \sigma_r^{(1)})), \mathbb{P}_n(G|(\sigma_1^{(2)}, \ldots, \sigma_r^{(2)}))\right) = o(1)$$

Here $\mathrm{TV}(\mu_1, \mu_2)$ is the total variation distance between two probability measures μ_1 and μ_2.

Proof. We know that

$$\begin{aligned}
&\text{TV}\left(\mathbb{P}_n(G|\sigma_u^{(1)}\quad u\in[r]),\mathbb{P}_n(G|\sigma_u^{(2)}\quad u\in[r])\right)\\
&=\sum_G\left|(\mathbb{P}_n(G|\sigma_u^{(1)}\quad u\in[r])-\mathbb{P}_n(G|\sigma_u^{(2)}\quad u\in[r])\right|\\
&=\sum_G\left|(\mathbb{P}_n(G|\sigma_u^{(1)}\quad u\in[r])-\mathbb{P}_n(G|\sigma_u^{(2)}\quad u\in[r])\right|\frac{\sqrt{\mathbb{P}'_n(G)}}{\sqrt{\mathbb{P}'_n(G)}}\\
&\leq\left(\sum_G\mathbb{P}'_n(G)\right)^{\frac{1}{2}}\left(\sum_G\frac{\left(\mathbb{P}_n(G|\sigma_u^{(1)}\quad u\in[r])-\mathbb{P}_n(G|\sigma_u^{(2)}\quad u\in[r]\right)^2}{\mathbb{P}'_n(G)}\right)^{\frac{1}{2}}\\
&=\left(\sum_G\frac{\left(\sum_{\tilde{\sigma}}\mathbb{P}_n(\tilde{\sigma})\left(\mathbb{P}_n(G|\sigma^{(1)},\tilde{\sigma})-\mathbb{P}_n(G|\sigma^{(2)},\tilde{\sigma})\right)^2}{\mathbb{P}'_n(G)}\right)^{\frac{1}{2}}.
\end{aligned}\tag{2.7.1}$$

Here $\sigma^{(1)} := \left\{(\sigma_1^{(1)},\ldots,\sigma_r^{(1)}\right\}$, $\sigma^{(2)} := \left\{(\sigma_1^{(2)},\ldots,\sigma_r^{(2)})\right\}$ and $\tilde{\sigma}$ is any configuration on $\{r+1,\ldots,n\}$.

Now observe that

$$\begin{aligned}
&\left(\sum_{\tilde{\sigma}}\mathbb{P}_n(\tilde{\sigma})\left(\mathbb{P}_n(G|\sigma^{(1)},\tilde{\sigma})-\mathbb{P}_n(G|\sigma^{(2)},\tilde{\sigma}\right)\right)^2\\
&=\sum_{\tilde{\sigma},\tilde{\tau}}\mathbb{P}_n(\tilde{\sigma})\mathbb{P}_n(\tilde{\tau})\left(\mathbb{P}_n(G|\sigma^{(1)},\tilde{\sigma})\mathbb{P}_n(G|\sigma^{(1)},\tilde{\tau})+\mathbb{P}_n(G|\sigma^{(2)},\tilde{\sigma})\mathbb{P}_n(G|\sigma^{(2)},\tilde{\tau})\right.\\
&\left.-\mathbb{P}_n(G|\sigma^{(1)},\tilde{\sigma})\mathbb{P}_n(G|\sigma^{(2)},\tilde{\tau})-\mathbb{P}_n(G|\sigma^{(2)},\tilde{\sigma})\mathbb{P}_n(G|\sigma^{(1)},\tilde{\tau})\right).
\end{aligned}\tag{2.7.2}$$

We shall prove that the value of

$$\sum_G\sum_{\tilde{\sigma},\tilde{\tau}}\mathbb{P}_n(\tilde{\sigma})\mathbb{P}_n(\tilde{\tau})\frac{\mathbb{P}_n(G|\sigma^{(1)},\tilde{\sigma})\mathbb{P}_n(G|\sigma^{(2)},\tilde{\tau})}{\mathbb{P}'_n(G)}\tag{2.7.3}$$

does not depend on $\sigma^{(1)}$ and $\sigma^{(2)}$ upto $o(1)$ terms. This will prove that the final expression in (2.7.1) goes to 0. As a consequence, the proof of Proposition 10 will be complete.

At first we recall the definition of $W_{uv}(G, \sigma)$ from (2.6.1). It is easy to observe that

$$\begin{aligned}
&\sum_G \sum_{\tilde\sigma,\tilde\tau} \frac{\mathbb{P}_n(\tilde\sigma)\mathbb{P}_n(\tilde\tau)\left(\mathbb{P}_n(G|\sigma^{(1)},\tilde\sigma)\mathbb{P}_n(G|\sigma^{(2)},\tilde\tau)\right)}{\mathbb{P}'_n(G)} \\
&= \sum_{\tilde\sigma,\tilde\tau} \frac{1}{2^{2(n-r)}} \sum_G \left(\prod_{uv} W(G,\sigma^{(1)},\tilde\sigma) W(G,\sigma^{(2)},\tilde\tau)\right) \mathbb{P}'_n(G) \qquad (2.7.4)\\
&= \frac{1}{2^{2(n-r)}} \sum_{\tilde\sigma,\tilde\tau} \prod_{u,v} \mathrm{E}_{\mathbb{P}'_n}(W(G,\sigma^{(1)},\tilde\sigma) W(G,\sigma^{(2)},\tilde\tau)).
\end{aligned}$$

Observe that the sum in the final expression of (2.7.4) is taken over $(\tilde\sigma, \tilde\tau)$ so the configurations in $\sigma^{(1)}$ and $\sigma^{(2)}$ remain unchanged.

Now let us introduce the following parameters

$$\begin{aligned}
\rho^{\text{fix}} &:= \frac{1}{r}\sum_{i=1}^{r} \sigma_i^{(1)}\sigma_i^{(2)} \\
S_{\pm}^{\text{fix}} &:= \sum_{u,v\in[r]} I_{\{\sigma_u^{(1)}\sigma_v^{(1)}\sigma_u^{(2)}\sigma_v^{(2)}=\pm 1\}}
\end{aligned} \qquad (2.7.5)$$

where I_A denotes the indicator variable corresponding to set A. We similarly define

$$\begin{aligned}
\rho(\tilde\sigma,\tilde\tau) &:= \frac{1}{n-r}\sum_{i=r+1}^{n} \tilde\sigma_i\tilde\tau_i \\
S_{\pm}(\tilde\sigma,\tilde\tau) &:= \sum_{u,v\notin[r]} I_{\{\tilde\sigma_u\tilde\sigma_v\tilde\tau_u\tilde\tau_v=\pm 1\}}.
\end{aligned} \qquad (2.7.6)$$

Finally for each $u \in [r]$ define

$$S_{u,\pm}(\tilde\sigma,\tilde\tau) = \#\{v \notin [r] \ : \ \tilde\sigma_v\tilde\tau_v = \pm\sigma_u^{(1)}\sigma_u^{(2)}\}. \qquad (2.7.7)$$

By using arguments similar to the proof of Lemma 4 one can show that the right hand side of the final expression of (2.7.4) further simplifies to

$$\begin{aligned}
&= \left(1+\frac{t_n}{n}\right)^{S_+^{\text{fix}}}\left(1-\frac{t_n}{n}\right)^{S_-^{\text{fix}}}\frac{1}{2^{2(n-r)}}\sum_{\tilde\sigma,\tilde\tau}\left(1+\frac{t_n}{n}\right)^{S_+(\tilde\sigma,\tilde\tau)}\left(1-\frac{t_n}{n}\right)^{S_-(\tilde\sigma,\tilde\tau)} \\
&\prod_{u\in[r]}\left(1+\frac{t_n}{n}\right)^{S_{u,+}(\tilde\sigma,\tilde\tau)}\times\left(1-\frac{t_n}{n}\right)^{S_{u,-}(\tilde\sigma,\tilde\tau)} \\
&= \left(1+\frac{t_n}{n}\right)^{S_+^{\text{fix}}}\left(1-\frac{t_n}{n}\right)^{S_-^{\text{fix}}}\frac{1}{2^{2(n-r)}}\sum_{\tilde\sigma,\tilde\tau}\left(1+\frac{t_n}{n}\right)^{\left(1+\rho(\tilde\sigma,\tilde\tau)^2\right)\frac{(n-r)^2}{4}-\frac{n-r}{2}} \\
&\left(1-\frac{t_n}{n}\right)^{\left(1-\rho(\tilde\sigma,\tilde\tau)^2\right)\frac{(n-r)^2}{4}}\prod_{u\in[r]}\left(1+\frac{t_n}{n}\right)^{n\frac{S_{u,+}(\tilde\sigma,\tilde\tau)}{n}}\left(1-\frac{t_n}{n}\right)^{n\frac{S_{u,-}(\tilde\sigma,\tilde\tau)}{n}}.
\end{aligned} \tag{2.7.8}$$

It is easy to see that for any fixed $u\in[r]$ and $\sigma_u^{(1)}$, $\sigma_u^{(2)}$ when $\tilde\sigma$ and $\tilde\tau$ are chosen independently and uniformly over $\{\pm1\}$ for each vertex $v\notin[r]$, both $\frac{S_{u,\pm}(\tilde\sigma,\tilde\tau)}{n}\xrightarrow{a.s.}\frac{1}{2}$. On the other hand $|S_{u,\pm}|\le n$. So both the quantities

$$\prod_{u\in[r]}\left(1+\frac{t_n}{n}\right)^{n\frac{S_{u,+}(\tilde\sigma,\tilde\tau)}{n}}$$

and

$$\prod_{u\in[r]}\left(1-\frac{t_n}{n}\right)^{n\frac{S_{u,-}(\tilde\sigma,\tilde\tau)}{n}}$$

are uniformly bounded over $\tilde\sigma$, $\tilde\tau$ and converge almost surely to $\exp\left(\frac{tr}{2}\right)$ and $\exp\left(-\frac{tr}{2}\right)$ under uniform independent assignment.

Now S_+^{fix} and S_-^{fix} are both bounded by r^2 also $t_n=(1+o(1))t$. So

$$\left(1+\frac{t_n}{n}\right)^{S_+^{\text{fix}}}\left(1-\frac{t_n}{n}\right)^{S_-^{\text{fix}}}=(1+o(1)).$$

On the other hand one can repeat the arguments in the proof of Lemma 4 to conclude that

$$\sum_{\tilde\sigma,\tilde\tau}\left(1+\frac{t_n}{n}\right)^{\left(1+\rho(\tilde\sigma,\tilde\tau)^2\right)\frac{(n-r)^2}{4}-\frac{n-r}{2}}\left(1-\frac{t_n}{n}\right)^{\left(1-\rho(\tilde\sigma,\tilde\tau)^2\right)\frac{(n-r)^2}{4}}\to\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}.$$

Combining all the arguments one gets the first expression in (2.7.8) converges to

$$\begin{aligned}&\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}\exp\left(\frac{tr}{2}\right)\exp\left(-\frac{tr}{2}\right)\\&=\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}.\end{aligned}$$

As a result

$$\sum_G\sum_{\tilde\sigma,\tilde\tau}\mathbb{P}_n(\tilde\sigma)\mathbb{P}_n(\tilde\tau)\frac{\mathbb{P}_n(G|\sigma^{(1)},\tilde\sigma)\mathbb{P}_n(G|\sigma^{(2)},\tilde\tau)}{\mathbb{P}'_n(G)}=(1+o(1))\frac{1}{\sqrt{1-t}}\exp\left\{-\frac{t}{2}-\frac{t^2}{4}\right\}$$

irrespective of the value of $\sigma^{(1)}$ and $\sigma^{(2)}$. So the final expression in (2.7.1) goes to 0. Hence the proof is complete. □

We now prove the following easy consequence of Proposition 10 which states that the posterior distribution of a single label is essentially unchanged if we know a bounded number of other labels.

Lemma 5. *Suppose S is a set of finite cardinality r, $u\notin S$ be a fixed node and π gives probability $\frac{1}{2}$ to both ±1. Then under the conditions of Proposition 10*

$$\mathrm{E}\left[\mathrm{TV}(\mathbb{P}_n(\sigma_u|G,\sigma_S),\pi)|\sigma_S\right]=o(1).$$

Proof. Observe that $\mathbb{P}_n(\sigma_u=i)=\pi(i)$ from the model assumption. So

$$\begin{aligned}\mathrm{E}\left[\mathrm{TV}(\mathbb{P}_n(\sigma_u|G,\sigma_S),\pi)|\sigma_S\right]&=\sum_G\sum_{i=\pm1}\left|\mathbb{P}_n\left(\sigma_u=i|G,\sigma_S\right)-\mathbb{P}_n(\sigma_u=i)\right|\mathbb{P}_n(G|\sigma_S)\\&=\sum_{i=\pm1}\mathbb{P}_n(\sigma_u=i)\sum_G\left|\frac{\mathbb{P}_n\left(\sigma_u=i|G,\sigma_S\right)}{\mathbb{P}_n(\sigma_u=i)}-1\right|\mathbb{P}_n(G|\sigma_S)\\&=\sum_{i=\pm1}\mathbb{P}_n(\sigma_u=i)\sum_G\left|\frac{\mathbb{P}_n\left(\sigma_u=i\cap G\cap\sigma_S\right)\mathbb{P}_n(\sigma_S)}{\mathbb{P}_n(\sigma_u=i\cap\sigma_S)\mathbb{P}_n(G\cap\sigma_S)}-1\right|\mathbb{P}_n(G|\sigma_S)\\&=\sum_{i=\pm1}\mathbb{P}_n(\sigma_u=i)\sum_G\left|\frac{\mathbb{P}_n(G|\sigma_S,\sigma_u=i)}{\mathbb{P}_n(G|\sigma_S)}-1\right|\mathbb{P}_n(G|\sigma_S)\end{aligned}\tag{2.7.9}$$

Observe that

$$\mathbb{P}_n(G|\sigma_S) = \frac{1}{2}\left(\mathbb{P}_n(G|\sigma_S, \sigma_u = 1) + \mathbb{P}_n(G|\sigma_S, \sigma_u = -1)\right).$$

As a consequence, the final expression of the right hand side of (2.7.9) becomes

$$\frac{1}{2}\sum_{i=\pm 1}\mathbb{P}_n(\sigma_u = i)\mathrm{TV}\left(\mathbb{P}_n(G|\sigma_S, \sigma_u = i), \mathbb{P}_n(G|\sigma_S, \sigma_u = -i)\right).$$

So the proof is complete by applying Proposition 10. □

With Proposition 10 and Lemma 5 in hand, we now give a proof of Theorem 3.

Proof of Theorem 3:

Let $\hat{\sigma}$ be any estimate of the labeling of the nodes, σ be the true labeling and $f : \{1, 2\} \to \{\pm 1\}$ be the function such that $f(1) = 1$ and $f(2) = -1$.

It is elementary to check that

$$\frac{1}{2}\mathrm{ov}(\sigma, \hat{\sigma}) = \frac{1}{n}\left[N_{11} + N_{22} - \frac{1}{n}(N_{1\cdot}N_{\cdot 1}) - \frac{1}{n}(N_{2\cdot}N_{\cdot 2})\right]. \tag{2.7.10}$$

Here

$$\begin{aligned} N_{ij} &= \left|\sigma^{-1}\{f(i)\} \cap \hat{\sigma}^{-1}\{f(j)\}\right| \\ N_{i\cdot} &= \left|\sigma^{-1}\{f(i)\}\right| \\ N_{\cdot j} &= \left|\hat{\sigma}^{-1}\{f(j)\}\right|. \end{aligned} \tag{2.7.11}$$

So it is sufficient to prove that

$$\frac{1}{n^2}\mathrm{E}_{\mathbb{P}_n}\left[N_{ii} - \frac{1}{n}N_{i\cdot}N_{\cdot i}\right]^2 = \frac{1}{n^2}\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}^2 - \frac{2}{n}N_{ii}N_{i\cdot}N_{\cdot i} + \frac{1}{n^2}N_{i\cdot}^2N_{\cdot i}^2\right] \to 0 \quad i \in \{1, 2\}.$$

Now

$$
\begin{aligned}
\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}^2\right] &= \mathrm{E}_{\mathbb{P}_n}\left[\sum_{u,v} I_{\{\sigma_u=f(i)\}} I_{\{\sigma_v=f(i)\}} I_{\{\hat{\sigma}_u=f(i)\}} I_{\{\hat{\sigma}_v=f(i)\}}\right] \\
&= \mathrm{E}_{\mathbb{P}_n}\left[\mathrm{E}\left[\sum_{u,v} I_{\{\sigma_u=f(i)\}} I_{\{\sigma_v=f(i)\}} I_{\{\hat{\sigma}_u=f(i)\}} I_{\{\hat{\sigma}_v=f(i)\}}\right] |G\right] \qquad (2.7.12) \\
&= \mathrm{E}_{\mathbb{P}_n}\left[\mathrm{E}\left[\sum_{u,v} I_{\{\sigma_u=f(i)\}} I_{\{\sigma_v=f(i)\}}\right] I_{\{\hat{\sigma}_u=f(i)\}} I_{\{\hat{\sigma}_v=f(i)\}} |G\right]
\end{aligned}
$$

The last step follows from the fact that $\hat{\sigma}$ is a function of G. Now

$$
\begin{aligned}
\mathrm{E}\left[I_{\{\sigma_u=f(i)\}} I_{\{\sigma_v=f(i)\}}|G\right] &= \mathrm{E}\left[I_{\{\sigma_u=f(i)\}}|G,\sigma_v=f(i)\right] \mathbb{P}_n\left(\sigma_v=f(i)|G\right) \\
&= (\pi(f(i))+o(1))\mathbb{P}_n(G|\sigma_v=f(i))\frac{\mathbb{P}_n(\sigma_v=f(i))}{\mathbb{P}_n(G)} \\
&= (\pi^2(f(i))+o(1))\frac{\mathbb{P}_n(G|\sigma_v=f(i))}{\mathbb{P}_n(G)}
\end{aligned}
$$

Here the second step follows from Lemma 5. Now,

$$
\begin{aligned}
&\left|\mathrm{E}_{\mathbb{P}_n}\left[\mathrm{E}\sum_{u,v}\left(I_{\{\sigma_u=f(i)\}} I_{\{\sigma_v=f(i)\}} - \pi^2(f(i))\right) I_{\{\hat{\sigma}_u=f(i)\}} I_{\{\hat{\sigma}_v=f(i)\}} |G\right]\right| \\
&\le \mathrm{E}_{\mathbb{P}_n}\left[\sum_{u,v}\left|\mathrm{E}\left[\left(I_{\{\sigma_u=f(i)\}} I_{\{\sigma_v=f(i)\}} - \pi^2(f(i))\right) I_{\{\hat{\sigma}_u=f(i)\}} I_{\{\hat{\sigma}_v=f(i)\}} |G\right]\right|\right] \\
&= \mathrm{E}_{\mathbb{P}_n}\left[\sum_{u,v}\left|\pi^2(f(i)) I_{\{\hat{\sigma}_u=f(i)\}} I_{\{\hat{\sigma}_v=f(i)\}}\left(\frac{\mathbb{P}_n(G|\sigma_v=f(i))}{\mathbb{P}_n(G)}-1\right)+o(1)\right|\right] \qquad (2.7.13) \\
&\le \sum_{u,v}\sum_{G}\left|\mathbb{P}_n(G|\sigma_v=f(i))-\mathbb{P}_n(G)\right|+o(n^2) \\
&= o(n^2).
\end{aligned}
$$

Here the last step follows from Proposition 10.

So we have

$$
\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}^2\right] = \sum_{u,v}\mathrm{E}_{\mathbb{P}_n}\left[\pi^2(f(i)) I_{\{\hat{\sigma}_u=f(i)\}} I_{\{\hat{\sigma}_v=f(i)\}}\right] + o(n^2) \qquad (2.7.14)
$$

Similar calculations will prove that

$$\mathrm{E}_{\mathbb{P}_n}\left[N_{ii}N_{i\cdot}N_{\cdot i}\right] = n\sum_{u,v}\mathrm{E}_{\mathbb{P}_n}\left[\pi^2(f(i))I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\right] + o(n^3) \tag{2.7.15}$$

and

$$\mathrm{E}_{\mathbb{P}_n}\left[N_{i\cdot}^2N_{\cdot i}^2\right] = n^2\sum_{u,v}\mathrm{E}_{\mathbb{P}_n}\left[\pi^2(f(i))I_{\{\hat{\sigma}_u=f(i)\}}I_{\{\hat{\sigma}_v=f(i)\}}\right] + o(n^4). \tag{2.7.16}$$

Plugging in these estimates we have

$$\frac{1}{n^2}\mathrm{E}_{\mathbb{P}_n}\left[N_{ii} - \frac{1}{n}N_{i\cdot}N_{\cdot i}\right]^2 = o(1).$$

This completes the proof. □

2.8. Appendix

2.8.1. More general words and their equivalence classes

Here we only give a very brief description about the combinatorial aspects of random matrix theory required to prove Lemma 3. For more general information one should look at Chapter 1 of Anderson et al. (2010) and Anderson and Zeitouni (2006). The definitions in this section have been taken from Anderson et al. (2010) and Anderson and Zeitouni (2006).

Definition 11. ($\mathcal{S}$ words) Given a set $\mathcal{S}$, an $\mathcal{S}$ letter s is simply an element of $\mathcal{S}$. An $\mathcal{S}$ word w is a finite sequence of letters $s_1 \ldots s_n$, at least one letter long. An $\mathcal{S}$ word w is closed if its first and last letters are the same. Two $\mathcal{S}$ words w_1, w_2 are called equivalent, denoted $w_1 \sim w_2$, if there is a bijection on $\mathcal{S}$ that maps one into the other.

When $\mathcal{S} = \{1, \ldots, N\}$ for some finite N, we use the term N word. Otherwise, if the set $\mathcal{S}$ is clear from the context, we refer to an $\mathcal{S}$ word simply as a word.

For any word $w = s_1 \ldots s_k$, we use $l(w) = k$ to denote the length of w, define the weight $wt(w)$ as the number of distinct elements of the set $s_1, \ldots, s_k$ and the support of w, denoted by $\mathrm{supp}(w)$, as the set of letters appearing in w. With any word w we may associate an undirected graph, with $wt(w)$ vertices and $l(w)-1$ edges, as follows.

Definition 12. (Graph associated with a word) Given a word $w = s_1 \dots s_k$, we let $G_w = (V_w, E_w)$ be the graph with set of vertices $V_w = \mathrm{supp}(w)$ and (undirected) edges $E_w = \{\{s_i, s_i+1\}, i = 1, \dots, k-1\}$.

The graph G_w is connected since the word w defines a path connecting all the vertices of G_w, which further starts and terminates at the same vertex if the word is closed. For $e \in E_w$, we use N_e^w to denote the number of times this path traverses the edge e (in any direction). We note that equivalent words generate the same graphs G_w (up to graph isomorphism) and the same passage-counts N_e^w.

Definition 13. (sentences and corresponding graphs) A sentence $a = [w_i]_{i=1}^n = [[\alpha_{i,j}]_{j=1}^{l(w_i)}]_{i=1}^n$ is an ordered collection of n words of length $(l(w_1), \dots, l(w_n))$ respectively. We define the graph $G_a = (V_a, E_a)$ to be the graph with

$$V_a = \mathrm{supp}(a), E_a = \{\{\alpha_{i,j}, \alpha_{i,j+1}\} | i = 1, \dots, n; j = 1, \dots, l(w_i) - 1\}\} .$$

Definition 14. (weak CLT sentences) A sentence $a = [w_i]_{i=1}^n$ is called a weak CLT sentence. If the following conditions are true:

1. All the words w_i's are closed.
2. Jointly the words w_i visit edge of G_a at least twice.
3. For each $i \in \{1, \dots, n\}$, there is another $j \neq i \in \{1, \dots, n\}$ such that G_{w_i} and G_{w_j} have at least one edge in common.

Note that these definitions are consistent with the ones given in Section 2.5. However, in Section 2.5, we defined these only for some specific cases required to solve the problem.

In order to prove Lemma 3, we require the following result from Anderson et al. (2010).

Lemma 6. *(Lemma 2.1.23 in Anderson et al. (2010)) Let $\mathcal{W}_{k,t}$ denote the equivalence classes corresponding to all closed words w of length $k+1$ with $\mathrm{wt}(w) = t$ such that each* *edge in G_w have been traversed at least twice. Then for $k > 2t-2$,*

$$\#\mathcal{W}_{k,t} \le 2^k k^{3(k-2t+2)}.$$

Assuming Lemma 6 we now prove Lemma 3.

Proof of Lemma 3: Let $a = [w_i]_{i=1}^m$ be a weak CLT sentence such that G_a have $\mathcal{C}(a)$ many connected components. At first we introduce a partition $\eta(a)$ in the following way. We put i and j in same block of $\eta(a)$ if G_{w_i} and G_{w_j} share an edge. At first we fix such a partition η and consider all the sentences such that $\eta(a) = \eta$. Let $\mathcal{C}(\eta)$ be the number of blocks in η. It is easy to observe that for any a with $\eta(a) = \eta$, we have $\mathcal{C}(\eta) = \mathcal{C}(a)$. From now on we denote $\mathcal{C}(\eta)$ by $\mathcal{C}$ for convenience.

Let a be any weak CLT sentence such that $\eta(a) = \eta$. We now propose an algorithm to embed a into $\mathcal{C}$ ordered closed words $(W_1, \ldots, W_{\mathcal{C}})$ such that the equivalence class of each W_i belongs to $\mathcal{W}_{L_i,t_i}$ for some numbers L_i and t_i.

A similar type of argument can be found in Claim 3 of the proof of Theorem 2.2 in Banerjee and Bose (2016).

An embedding algorithm: Let $B_1, \ldots, B_{\mathcal{C}}$ be the blocks of the partition η ordered in the following way. Let $m_i = \min\{j : j \in B_i\}$ and we order the blocks B_i such that $m_1 < m_2 \ldots < m_{\mathcal{C}}$. Given a partition η this ordering is unique. Let

$$B_i = \{i(1) < i(2) < \ldots < i(l(B_i))\}.$$

Here $l(B_i)$ denotes the number of elements in B_i.

For each B_i we embed the sentence $a_i = [w_{i(j)}]_{1 \le j \le l(B_i)}$ into W_i sequentially in the following manner.

1. Let $S_1 = \{i(1)\}$ and $\mathfrak{w}_1 = w_{i(1)}$.

2. For each $1 \le c \le l(B_i) - 1$ we perform the following.

- Consider $\mathfrak{w}_c = (\alpha_{1,c}, \ldots, \alpha_{l(\mathfrak{w}_c),c})$ and $S_c \subset B_i$. Let $ne \in B_i \backslash S_c$ be the minimum index such that the following two conditions hold.

  (a) $G_{\mathfrak{w}_c}$ and $G_{w_{ne}}$ shares at least one edge $e = \{\alpha_{\kappa_1,c}, \alpha_{\kappa_1+1,c}\}$.

  (b) κ_1 is minimum among all such choices.

- Let $w_{ne} = (\beta_{1,c}, \ldots, \beta_{l(w_{ne}),c})$ and $\{\beta_{\kappa_2,c}, \beta_{\kappa_2+1,c}\}$ be the first time e appears in w_{ne}. As $\{\beta_{\kappa_2,c}, \beta_{\kappa_2+1,c}\} = \{\alpha_{\kappa_1,c}, \alpha_{\kappa_1+1,c}\}$, $\alpha_{\kappa_1,c}$ is either equal to $\beta_{\kappa_2,c}$ or $\beta_{\kappa_2,c}$. Let $\kappa_3 \in \{\kappa_2, \kappa_2 + 1\}$ such that $\alpha_{\kappa_1,c} = \beta_{\kappa_3,c}$. If $\beta_{\kappa_2,c} = \beta_{\kappa_2+1,c}$, then we simply take $\kappa_3 = \kappa_2$.

- We now generate $\mathfrak{w}_{c+1}$ in the following way

$$\mathfrak{w}_{c+1} = (\alpha_{1,c}, \ldots, \alpha_{\kappa_1,c}, \beta_{\kappa_3+1,c}, \ldots, \beta_{l(w_{ne}),c}, \beta_{2,c}, \ldots, \beta_{\kappa_3,c}, \alpha_{\kappa_1+1,c}, \ldots, \alpha_{l(\mathfrak{w}_c),c}).$$

  Let $\tilde{a}_c := (\mathfrak{w}_c, w_{ne})$. It is easy to observe by induction that all $\mathfrak{w}_c$'s are closed words and so are all the w_{ne}'s. Also all the edges in the graph $G_{\tilde{a}_c}$ are preserved along with their passage counts in $G_{\mathfrak{w}_{c+1}}$.

- Generate $S_{c+1} = S_c \cup \{ne\}$.

3. Return $W_i = \mathfrak{w}_{l(B_i)}$.

In the preceding algorithm we have actually defined a function f which maps any weak CLT sentence a into $\mathcal{C}$ ordered closed words $(W_1, \ldots, W_{\mathcal{C}})$ such that the equivalence class of each W_i belongs to $\mathcal{W}_{L_i,t_i}$ for some numbers L_i and t_i. Observe that given two words $\mathfrak{w}_1$ and

$\mathfrak{w}_2$, application of step 2 gives rise to a closed word $\mathfrak{w}_3$ where $l(\mathfrak{w}_3) = l(\mathfrak{w}_1) + l(\mathfrak{w}_2) - 1$. So

$$\begin{aligned} L_i &= \sum_{j \in B_i} l(w_j) - (l(B_i) - 1) < \sum_{j \in B_i} l(w_j). \\ \Rightarrow L_i + 1 &\leq \sum_{j \in B_i} l(w_j) \\ \Rightarrow L_i + 1 - 2t_i &\leq \sum_{j \in B_i} l(w_j) - 2t_i. \end{aligned} \tag{2.8.1}$$

Unfortunately f is not an injective map. So given $(W_1, \ldots, W_{\mathcal{C}})$ we find an upper bound to the cardinality of the following set

$$f^{-1}(W_1, \ldots, W_{\mathcal{C}}) := \{a | f(a) = (W_1, \ldots, W_{\mathcal{C}})\}$$

We have argued earlier $\mathcal{C}$ is the number of blocks in η. However, in general $(W_1, \ldots, W_{\mathcal{C}})$ does neither specify the partition η nor the order in which the words are concatenated with in each block B_i of η. So we fix a partition η with $\mathcal{C}$ many blocks and an order of concatenation $\mathcal{O}$. Observe that

$$\mathcal{O} = (\sigma_1(\eta), \ldots, \sigma_{\mathcal{C}}(\eta))$$

where for each i, $\sigma_i(\eta)$ is a permutation of the elements in B_i. Now we give an uniform upper bound to the cardinality of the following set

$$f^{-1}_{\eta,\mathcal{O}}(W_1, \ldots, W_{\mathcal{C}}) := \{a | \eta(a) = \eta \ \ ; \mathcal{O}(a) = \mathcal{O} \ \ \& f(a) = (W_1, \ldots, W_{\mathcal{C}})\}.$$

According to the algorithm any word W_i is formed by recursively applying step 2 to $(\mathfrak{w}_c, w_{ne})$ for $1 \leq c \leq l(B_i)$. Given a word $\mathfrak{w}_3 = (\alpha_1, \ldots, \alpha_{l(\mathfrak{w}_3)})$, we want to find out the number of two word sentences $(\mathfrak{w}_1, \mathfrak{w}_2)$ such that applying step 2 of the algorithm on $(\mathfrak{w}_1, \mathfrak{w}_2)$ gives $\mathfrak{w}_3$ as an output. This is equivalent to choose three positions $i_1 < i_2 < i_3$ from the set

$\{1, \ldots, l(\mathfrak{w}_3)\}$ such that $\alpha_{i_1} = \alpha_{i_3}$. Once these three positions are chosen, $(\mathfrak{w}_1, \mathfrak{w}_2)$ can be constructed uniquely in the following manner

$$\mathfrak{w}_1 = (\alpha_1, \ldots, \alpha_{i_1}, \alpha_{i_3+1}, \ldots, \alpha_{l(\mathfrak{w}_3)})$$
$$\mathfrak{w}_2 = (\alpha_{i_2}, \ldots, \alpha_{i_3}, \alpha_{i_1+1}, \ldots, \alpha_{i_2}).$$

Total number of choices $i_1 < i_2 < i_3$ is bounded by $l(\mathfrak{w}_3)^3 \le \left(\sum_{i=1}^m l(w_i)\right)^3$. For each block B_i, step 2 of the algorithm has been used $l(B_i)$ many times. So

$$f_{\eta,\mathcal{O}}^{-1}(W_1, \ldots, W_{\mathcal{C}}) \le \left(\sum_{i=1}^m l(w_i)\right)^{3\sum_{i=1}^{\mathcal{C}} l(B_i)} = \left(\sum_{i=1}^m l(w_i)\right)^{3m}.$$

On the other hand, there are at most m^m many η's and for each η there are at most $\prod_{i=1}^{\mathcal{C}} l(B_i)! \le m^m$ choices of $\mathcal{O}$. So

$$f^{-1}(W_1, \ldots, W_{\mathcal{C}}) \le m^{2m} \left(\sum_{i=1}^m l(w_i)\right)^{3m} \le \left(D_1 \sum_{i=1}^m l(w_i)\right)^{D_2 m} \tag{2.8.2}$$

for some known constants D_1 and D_2. Now we fix the sequence (L_i, t_i) and find an upper bound to the number of $(W_1, \ldots, W_{\mathcal{C}})$. From Lemma 6 we know the number of choices of W_i is bounded by $2^{L_i - 1}(L_i - 1)^{L_i - 2t_i + 1} n^{t_i}$. So the total number of choices for $(W_1, \ldots, W_{\mathcal{C}})$ is bounded by

$$2^{\sum_{i=1}^{\mathcal{C}} L_i} \prod_{i=1}^{\mathcal{C}} (L_i - 1)^{3(L_i - 2t_i + 1)} n^{t_i} \le 2^{\sum_{i=1}^m l(w_i)} n^t \left(\sum_{i=1}^m l(w_i)\right)^{3\left(\sum_{i=1}^m l(w_i) - 2t\right)}. \tag{2.8.3}$$

Now the number of choices (L_i, t_i) such that $\sum_{i=1}^{\mathcal{C}} L_i = \sum_{i=1}^m l(w_i) - \sum_{i=1}^{\mathcal{C}} (l(B_i) - 1)$ and $\sum_{i=1}^{\mathcal{C}} t_i = t$ are bounded by

$$\binom{\sum_{i=1}^m l(w_i) - \sum_{i=1}^{\mathcal{C}} (l(B_i) - 1) - 1}{\mathcal{C} - 1} \binom{t-1}{\mathcal{C}-1} \le \binom{\sum_{i=1}^m l(w_i) - 1}{\mathcal{C} - 1} \binom{t-1}{\mathcal{C}-1} \le \left(\sum_{i=1}^m l(w_i)\right)^{2m}. \tag{2.8.4}$$

Here the inequality follows since $\mathcal{C} \le m$ and $t < \sum_{i=1}^m \frac{l(w_i) - 1}{2}$. Finally we using the fact that $1 \leq \mathcal{C} \leq m$ and combining (4.5.3), (4.5.4) and (4.5.5) we finally have

$$\#\mathcal{A} \leq \left(D_1 \sum_{i=1}^{m} l(w_i)\right)^{D_2 m} \times 2^{\sum_{i=1}^{m} l(w_i)} n^t \left(\sum_{i=1}^{m} l(w_i)\right)^{3(\sum_{i=1}^{m} l(w_i) - 2t)} \times \left(\sum_{i=1}^{m} l(w_i)\right)^{2m}$$
$$\Rightarrow \#\mathcal{A} \leq 2^{\sum_i l(w_i)} \left(C_1 \sum_i l(w_i)\right)^{C_2 m} \left(\sum_i l(w_i)\right)^{3(\sum_i l(w_i) - 2t)} n^t \tag{2.8.5}$$

as required. □

CHAPTER 3 : Optimal signal detection in some spiked random matrix models:likelihood ratio tests and linear spectral statistics

3.1. Introduction

An important class of signal detection problems share the following hypothesis testing framework. Under the null hypothesis, the observed data matrix consists of pure noise. Under the alternative, it has a "spiked signal + noise" structure, where the signal component is of low rank and certain knowledge can be encoded as a prior distribution on the signal.

We consider two different versions of the problem. In both cases, we may assume that the observed data matrix X is in $\mathbb{R}^{n\times p}$.
(i) In the *mean detection* case, let $Z = (Z_{ij}) \in \mathbb{R}^{n\times p}$ with $Z_{ij} \overset{iid}{\sim} N(0,1)$. We aim to test

$$H_0 : X = Z, \quad \text{vs.} \quad H_1 : X = \frac{1}{\sqrt{p}}\Theta U' + Z, \tag{3.1.1}$$

where $\Theta \in \mathbb{R}^{n\times\kappa}$ follows some prior distribution π_Θ and $U \in \mathbb{R}^{p\times\kappa}$ follows some prior distribution π_U. In addition, Θ, U and Z are mutually independent. Through out the paper the rows of the matrix X will be denoted by $X_1, \ldots, X_n$. Here and after, for any matrix $A \in \mathbb{R}^{n\times p}$, A' stands for its transposition, $A_{i*} \in \mathbb{R}^{1\times p}$ its i-th row, and $A_{*j} \in \mathbb{R}^{n\times 1}$ its j-th column. We use A_{ij} and $A(i,j)$ exchangeably to denote its (i,j)-th entry. For any positive integer l, $[l] = \{1, 2, \ldots, l\}$. Throughout, we assume that under π_Θ (and π_U, resp.), the rows of Θ (U, resp.) are i.i.d. random vectors with $\mathbb{E}[\Theta_{1*}] = 0$ and $\mathrm{Cov}(\Theta'_{1*}\Theta_{1*}) = \Sigma_\Theta$ (with $\mathbb{E}[U_{1*}] = 0$ and $\mathrm{Cov}(U'_{1*}U_{1*}) = \Sigma_U$, resp.). In other words, we assume that $X_{ij} \overset{iid}{\sim} N(0,1)$ under H_0, and under H_1, $X_{ij}|(\Theta, U) \overset{ind}{\sim} N(\frac{1}{\sqrt{p}}\sum_{l=1}^{\kappa}\Theta_{il}U_{jl}, 1)$ where $\mathbb{E}[\Theta_{il}] = 0$, $\mathbb{E}[U_{jl}] = 0$, $\mathbb{E}[\Theta_{il_1}\Theta_{il_2}] = \Sigma_\Theta(l_1, l_2)$ and $\mathbb{E}[U_{jl_1}U_{jl_2}] = \Sigma_U(l_1, l_2)$ for $l, l_1, l_2 \in [\kappa]$. Under H_1, if the distribution of Θ_{i*} is discrete and takes a finite number of values, then conditioning on U the rows of X in (3.1.1) can be viewed as i.i.d. observations from a Gaussian mixture distribution.

(ii) In the *covariance detection* case, we test

$$H_0 : X = Z, \quad \text{vs.} \quad H_1 : X = \Theta V' + Z, \tag{3.1.2}$$

where $V = U(U'U)^{-1/2}$ and U is defined as is in the previous case. In other words, V is a self-normalized version of U such that $V'V = I_\kappa$ and so $V \in O(p, \kappa)$, i.e., the Stiefel manifold consisting of all κ-frames in $\mathbb{R}^p$. Under H_1, if $\Theta_{i*} \overset{iid}{\sim} N(0, \Sigma_\Theta)$ with $\Sigma_\Theta = H = \text{diag}(h_1, \ldots, h_\kappa)$ with $h_1 \geq \cdots \geq h_\kappa > 0$, then conditioning on V the rows of X are i.i.d. observations from a p-dimensional normal distribution with mean 0 and multi-spiked covariance matrix $VHV' + I_p$. Here I_p is the p-dimensional identity matrix. In this case, the testing problem (3.1.2) reduces to the high dimensional sphericity test against multi-spiked alternative.

In either case, we aim to detect a spiked random matrix model against the null. Moreover, we deal with simple vs. simple hypothesis testing since we put some prior distribution on the signal component under the alternative. Therefore, the Neyman–Pearson lemma dictates that the likelihood ratio test (LRT) is optimal. In this paper, we are concerned with the asymptotic behavior of likelihood ratios in the aforementioned detection/testing problems. In particular, let $p = p_n$ scale with n. We are interested in the asymptotic regime where

$$p/n \to \gamma \in (0, \infty) \text{ as } n \to \infty. \tag{3.1.3}$$

Let $\mathbb{P}_{0,n}$ be the null distribution and $\mathbb{P}_{1,n}$ the alternative. Let $L_n = \frac{\mathrm{d}\mathbb{P}_{1,n}}{\mathrm{d}\mathbb{P}_{0,n}}$ denote the likelihood ratio, and we call $\log(L_n)$ the log-likelihood ratio.

A series of papers have discovered the following general phenomenon in these testing problems. Depending on the signal-to-noise ratio[1] (SNR), there are two different types of asymptotic behavior of the likelihood ratio. If the SNR is below certain bound, then L_n has a nontrivial weak limit, and the null and alternative distributions are mutually contiguous. When the SNR is sufficiently large, the likelihood ratio converges to zero under null and

[1]For now, this can be loosely understood as $\|\Sigma_\Theta \Sigma_U\|_2$ for (3.1.1) and $\|\Sigma_\Theta\|_2$ for (3.1.2) where $\|\cdot\|_2$ denotes the spectral norm of a matrix.

diverges to infinity under alternative in probability as n tends to infinity. In this case, the two distributions are asymptotically singular. **?** focused on providing good lower and upper bounds for the SNR thresholds between asymptotically contiguous and singular regimes in both detection problem (3.1.1) and its symmetric version known as detection for the Gaussian spiked Wigner model. They considered priors with i.i.d. Gaussian entries and those with bounded support sizes such as those with i.i.d. (sparse) Rademacher entries and those formed by membership matrices of uniformly random balanced partitions. **?** investigated the same issue in three single-spiked models: Gaussian spiked Wigner model, non-Gaussian spiked Wigner model and spiked Wishart model. In addition, they determined when spectral method, i.e., PCA, is optimal or sub-optimal in these models. When the spike is positive, the single-spiked Wishart model they considered is a special case of model (3.1.1) with $\kappa = 1$, a Gaussian prior for Θ and a sub-Gaussian prior for U. Neither **?** nor **?** studies asymptotic distributions of log-likelihood ratios. In a different line of research, Onatski et al. Onatski et al. (2013, 2014) (see also Johnstone and Onatski (2015)) derived asymptotic normality and uniform convergence of the log-likelihood ratio under both single and multiple spiked Wishart models for all contiguous alternative distributions, under the condition that the prior on the leading eigenvectors is the uniform probability measure on the corresponding Stiefel manifold. The scenario they considered can be identified as model (3.1.2) with the entries of U sampled i.i.d. from a standard normal distribution and the rows of Θ sampled i.i.d. from $N(0, H)$ with $H = \mathrm{diag}(h_1, \ldots, h_\kappa)$. Furthermore, **?** and **?**, among other results, derived asymptotic normality results for log-likelihood ratios for single spiked Wigner and Wishart models. Their approach is borrowed from spin glass literature and uses cavity method. However, the priors they considered were restricted by a uniformly bounded support size condition. For single spiked Wishart models, the result in **?** is complementary to that in Onatski et al. (2013) as the uniformly bounded support size condition excludes the spherically symmetric prior which ensures that the likelihood ratio of sample eigenvalues used in Onatski et al. (2013) is identical to the full likelihood ratio.

3.1.1. Main contributions

The main contributions of the present manuscript are three-folded:

(1) For both models (3.1.1) and (3.1.2) and for a large collection of sub-Gaussian priors on Θ and U, we show that when the SNR is below certain bound, the null and the alternative distributions are asymptotically mutually contiguous and that the log-likelihood ratio has normal limits under both null and alternative in asymptotic regime (3.1.3). The bound is defined jointly by the dimension-sample size ratio γ, second moments and sub-Gaussianity parameters of the priors. The limiting normal distributions have different means but the same variance, both of which depend only on γ and second moments of the priors. We allow any prior on Θ that assigns independent sub-Gaussian row vectors, and also any such prior on U. This allows the rows of Θ (and U) to be i.i.d. according to any multivariate normal distribution or any multivariate discrete/continuous distribution with bounded support, among other possibilities. To the best of our limited knowledge, the present manuscript is the first to give such results for these multi-spiked signal detection problems beyond the case of uniform priors.

(2) In either model, when the SNR is below the bound under which we have asymptotic normality for the log-likelihood ratio, we show that under either null or alternative the log-likelihood ratio can be decomposed as the weighted sum of a collection of statistics, defined later as *bipartite signed cycles*. The bipartite signed cycles are asymptotically independently and normally distributed. Thus, this provides an asymptotic analysis of variance (ANOVA) type decomposition for the log-likelihood ratio statistics. Such a result provides insights on the source of randomness in the asymptotic log-likelihood ratio.

(3) For both models, we show that below the SNR bound a special class of linear spectral statistics of the matrix $\frac{1}{n}X'X$ first proposed in Onatski et al. (2014) leads to a test that has the exact asymptotically optimal power of the LRT. Therefore, for such testing

problems, the asymptotically optimal detection power can be achieved within polynomial time complexity. The result is in a sense universal because it holds regardless of the specific sub-Gaussian priors one puts on the signal component!

The approach we take in the present manuscript is inspired by a parallel line of research on contiguity and signal detection for random graph models. **?** introduced a second moment argument to study asymptotically contiguous random graph models with respect to random d-regular graphs where the degree parameter d remains fixed as the graph size tends to infinity. In addition, he showed that the asymptotic likelihood ratios between these sparse random graph models are determined by counts of cycles on graphs. Mossel et al. (2015) established a comparable set of results when studying the detection of planted partition models (i.e., symmetric two block stochastic block models) against Erdős-Renyi graphs in the asymptotic regime where the average degree of nodes remain finite when the graph size tends to infinity. They determined the exact boundary between asymptotically contiguous and singular regimes and showed that within the contiguous regime, the asymptotic log-likelihood ratio is determined by counts of cycles and has a Poisson mixture limit. Banerjee (2018) studied the same Erdős-Renyi model vs. planted partition model detection problem in a different asymptotic regime where average degree and graph size tend to infinity together. Similar to Mossel et al. (2015), he determined the exact boundary between contiguity and singularity. In addition, he showed that in the contiguous regime, the asymptotic likelihood ratio is determined by a series of graph statistics called signed cycles as opposed to actual counts of cycles on graph. The major tool in Banerjee (2018) is a Gaussian version of Janson's second moment method in **?**. This second moment method also serves as the backbone of arguments in the present manuscript. In addition, Banerjee and Ma (2017a) considered approximation of LRTs in detecting planted partition model by linear spectral statistics of the adjacency matrix. Despite the apparent connections, the present paper is different from Banerjee (2018) and Banerjee and Ma (2017a) in two important aspects. First, the successful application of Janson's method requires explicit construction of a specific collection of statistics that determine the asymptotic likelihood ratios for the models of interest. In this paper, we shall construct a new set of statistics called bipartite signed cycles and analyze them under a large collection of prior distributions on Θ and U in a multiple spike setting. In contrast, Banerjee (2018) worked with a different set of statistics and the analysis was restricted to a single spike setting with a simple i.i.d. Rademacher prior. Second, we further provide an asymptotic ANOVA type decomposition for the log-likelihood ratios that deepens the connection of the log-likelihood ratios with the bipartite signed cycles. Neither Banerjee (2018) nor Banerjee and Ma (2017a) provided a result of this nature. This decomposition is critical for showing that a polynomial time test based on linear spectral statistics can *simultaneously* achieve the asymptotically optimal powers of LRTs for a large collection of priors. In comparison, the analysis of approximation by linear spectral statistics in Banerjee and Ma (2017a) was again restricted to single spike with i.i.d. Rademacher prior.

3.1.2. Organization

The rest of this manuscript is organized as the following. We introduce Janson's second moment method and other important definitions in Section 3.2. Section 3.3 states the main theorems. Section 3.4 establishes the asymptotic normality of bipartite signed cycles which is instrumental in the proof of main theorems on the asymptotic behavior of log-likelihood ratios in Section 3.5. We then prove the achievability of LRT powers by linear spectral statistics in Section 3.6. The appendices in **?** present the proofs of more technical results.

3.2. Preliminaries

Contiguity For two sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ defined on σ-fields $(\Omega_n, \mathcal{F}_n)$, we say that $\mathbb{Q}_n$ is contiguous with respect to $\mathbb{P}_n$, denoted by $\mathbb{Q}_n \triangleleft \mathbb{P}_n$, if for any event sequence A_n, $\mathbb{P}_n(A_n) \to 0$ implies $\mathbb{Q}_n(A_n) \to 0$. We say that they are (asymptotically) mutually contiguous, denoted by $\mathbb{P}_n \triangleleft \triangleright \mathbb{Q}_n$, if both $\mathbb{Q}_n \triangleleft \mathbb{P}_n$ and $\mathbb{P}_n \triangleleft \mathbb{Q}_n$ hold. We refer interested readers to Le Cam (2012) and Le Cam and Yang (2012) for general discussions on contiguity.

To establish our main results, we rely on the following proposition for establishing contiguity and asymptotic normality of log-likelihood ratios. For any two probability measure $\mathbb{P}$ and $\mathbb{Q}$ on the same probability space, we write $\mathbb{Q} \ll \mathbb{P}$ when $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$.

Proposition 2 (Janson's second moment method)**.** *Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures such that for each n, both are defined on the common σ-algebra $(\Omega_n, \mathcal{F}_n)$. Suppose that for each $i \geq 1$, $W_{n,i}$ are random variables defined on $(\Omega_n, \mathcal{F}_n)$. The sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if the following conditions hold simultaneously:*

(i) $\mathbb{Q}_n$ *is absolutely continuous with respect to* $\mathbb{P}_n$ *for each* n*;*

(ii) *For any fixed* $k \geq 1$*, one has* $(W_{n,1}, \ldots, W_{n,k})\,|\mathbb{P}_n \xrightarrow{d} (Z_1, \ldots, Z_k)$ *and* $(W_{n,1}, \ldots, W_{n,k})\,|\mathbb{Q}_n \xrightarrow{d} (Z'_1, \ldots, Z'_k)$.

(iii) $Z_i \sim N(0, \sigma_i^2)$ *and* $Z'_i \sim N(\mu_i, \sigma_i^2)$ *are sequences of independent random variables.*

(iv) *The likelihood ratio statistic* $Y_n = \frac{\mathrm{d}\mathbb{Q}_n}{\mathrm{d}\mathbb{P}_n}$ *satisfies*

$$\limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}\left[Y_n^2\right] \leq \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\} < \infty. \tag{3.2.1}$$

In addition, under these four conditions, we have that under $\mathbb{P}_n$*,*

$$Y_n \xrightarrow{d} \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i Z_i - \frac{1}{2}\mu_i^2}{\sigma_i^2}\right\}. \tag{3.2.2}$$

Furthermore, given any $\epsilon, \delta > 0$ *there exists a natural number* $K = K(\delta, \epsilon)$ *such that for any sequence* n_l *there is a further subsequence* n_{l_m} *such that*

$$\limsup_{m\to\infty} \mathbb{P}_{n_{l_m}}\left(\left|\log(Y_{n_{l_m}}) - \sum_{k=1}^{K} \frac{2\mu_k W_{n_{l_m},k} - \mu_k^2}{2\sigma_k^2}\right| \geq \epsilon\right) \leq \delta. \tag{3.2.3}$$

Remark 4. The proposition can be viewed as a Gaussian version of Theorem 1 in Janson **?**

which dealt with convergence to a Poisson mixture. In addition, it generalizes Proposition 3.4 in Banerjee (2018) where a more specific version of it appeared with an additional redundant condition ($\mathbb{P}_n$ being absolutely continuous with respect to $\mathbb{Q}_n$). Moreover, Proposition 3.4 in Banerjee (2018) did not have conclusion (3.2.3). Intuitively speaking, one may interpret the result as the collection of statistics $\{W_{n,i} : i \geq 1\}$ are "asymptotically sufficient" for determining the behavior of the likelihood ratio statistic.

Bipartite signed cycles In view of Proposition 2, our proofs rely on explicitly constructing a class of random variables that play the roles of the $W_{n,i}$'s. To this end, we define the following set of statistics for testing problems (3.1.1) and (3.1.2).

Definition 15 (Bipartite signed cycle of length $2k$)**.** For each $k \in [n \wedge p]$, we define the bipartite signed cycle of length $2k$ as

$$B_{n,k} = \frac{1}{n^k} \sum_{i_0,j_0,i_1,j_1,\ldots,i_{k-1},j_{k-1}} X_{i_0,j_0} X_{i_1,j_0} X_{i_1,j_1} \ldots X_{i_{k-1},j_{k-1}} X_{i_0,j_{k-1}} \tag{3.2.4}$$

where $i_0, i_1, \ldots, i_{k-1} \in [n]$ are all distinct, and so are $j_0, j_1, \ldots, j_{k-1} \in [p]$.

It is worth noting that statistics based on summing over certain types of "paths" like (3.2.4) have appeared previously in the random matrix theory literature. See for instance **???**. However, an important distinction is that the cycle considered here is simple in the sense that no vertex is repeated.

We are to show that for both testing problems (3.1.1) and (3.1.2), a linear combination of bipartite signed cycles of increasing lengths approximates the asymptotic log-likelihood ratio in probability, at least for a large collection of prior distributions on Θ and U which we now define.

Sub-Gaussian prior distributions We first recall the definition of sub-Gaussian random vectors and their variance proxies.

Definition 16. Suppose X is a random vector of dimension d. We say the random vector

X is sub-Gaussian with variance proxy $\widetilde{\Sigma}_X$ if $\mathbb{E}[X] = 0$ and $\mathbb{E}[\exp(t'X)] \leq \exp(\frac{1}{2}t'\widetilde{\Sigma}_X t)$ for any $t \in \mathbb{R}^d$. Here $\widetilde{\Sigma}_X$ is a non-negative definite matrix.

By definition, if $\widetilde{\Sigma}_X$ is a variance proxy for X, then so is any matrix $\widetilde{\Sigma}$ such that $\widetilde{\Sigma} - \widetilde{\Sigma}_X$ is non-negative definite. For any multivariate normal distribution the variance proxy can be chosen to match the true covariance matrix. If X is a random vector with i.i.d. Rademacher entries then X is sub-Gaussian with variance proxy I_d. Furthermore, if X is sub-Gaussian with variance proxy $\widetilde{\Sigma}_X$ then for any A, AX is sub-Gaussian with variance proxy $A'\widetilde{\Sigma}_X A$. Finally, if X is a random variable taking values within $[a, b]$, then $X - \mathbb{E}[X]$ is sub-Gaussian with variance proxy $\frac{1}{4}(a-b)^2$.

Definition 17 (Sub-Gaussian prior)**.** For any given number $\kappa < \min(n, p)$, let $\mathcal{P}(n, \kappa, \Sigma_\Theta, \widetilde{\Sigma}_\Theta)$ be the collection of all priors π_Θ on Θ such that under π_Θ, the row vectors $\{\Theta_{i*} : i \in [n]\}$ are i.i.d. sub-Gaussian random vectors in $\mathbb{R}^\kappa$ with mean zero, covariance matrix Σ_Θ and variance proxy $\widetilde{\Sigma}_\Theta$. Let $\mathcal{P}(p, \kappa, \Sigma_U, \widetilde{\Sigma}_U)$ be defined analogously for U.

3.3. Main results

3.3.1. Asymptotic behavior of log-likelihood ratios

We first state the theorem for testing problem (3.1.1). Recall that for any matrix A, $\|A\|_2$ denotes its spectral norm which equals the largest singular value of A. For any square matrix A, let $\mathsf{Tr}(A)$ denote its trace. For any event E, let $\mathbf{1}_E$ be its indicator function.

Theorem 4. *Consider the testing problem defined in (3.1.1) with the Θ prior $\pi_\Theta \in \mathcal{P}(n, \kappa, \Sigma_\Theta, \widetilde{\Sigma}_\Theta)$ and the U prior $\pi_U \in \mathcal{P}(p, \kappa, \Sigma_U, \widetilde{\Sigma}_U)$. Denote the null distribution by $\mathbb{P}_{0,n}$, the alternative distribution $\mathbb{P}_{1,n}$ and the likelihood ratio $L_n = \frac{\mathrm{d}\mathbb{P}_{1,n}}{\mathrm{d}\mathbb{P}_{0,n}}$. Suppose as $n \to \infty$, $p/n \to \gamma \in (0, \infty)$ while $\kappa, \Sigma_\Theta, \Sigma_U, \widetilde{\Sigma}_\Theta, \widetilde{\Sigma}_U$ remain fixed. The following hold whenever*

$$\|\widetilde{\Sigma}_\Theta \widetilde{\Sigma}_U\|_2 \|\Sigma_\Theta \Sigma_U\|_2 < \gamma. \tag{3.3.1}$$

1. *$\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ are asymptotically mutually contiguous.*

2. *Under* H_0,

$$L_n \xrightarrow{d} \exp\left\{\sum_{k=1}^{\infty} \frac{2\mu_k Z_k - \mu_k^2}{4k\gamma^k}\right\} \tag{3.3.2}$$

where Z_k *are independent* $N\left(0, 2k\gamma^k\right)$ *random variables and for any* k, $\mu_k = \mathsf{Tr}\left((\Sigma_\Theta \Sigma_U)^k\right)$. *By continuous mapping, under* H_0, $\log(L_n) \xrightarrow{d} N(-\frac{1}{2}\sigma_b^2, \sigma_b^2)$ *with*

$$\sigma_b^2 = \sum_{k=1}^{\infty} \frac{\mu_k^2}{2k\gamma^k} = \frac{1}{2}\sum_{i=1}^{\kappa}\sum_{j=1}^{\kappa} \log\left(1 - \frac{h_i h_j}{\gamma}\right) \tag{3.3.3}$$

where $h_1 \geq \cdots \geq h_\kappa$ *are the eigenvalues of* $\Sigma_\Theta \Sigma_U$. *Under* H_1, *we have* (3.3.2) *with* $Z_k \overset{ind}{\sim} N(\mu_k, 2k\gamma^k)$ *and* $\log(L_n) \xrightarrow{d} N(\frac{1}{2}\sigma_b^2, \sigma_b^2)$.

3. *Further under both null and alternative the log-likelihood ratio satisfies the following ANOVA type decomposition:*

$$\log(L_n) - \sum_{k=1}^{m_n} \frac{2\mu_k\left(B_{n,k} - p\mathbf{1}_{k=1}\right) - \mu_k^2}{4k\gamma^k} \xrightarrow{p} 0 \tag{3.3.4}$$

where m_n *is any sequence growing to* ∞ *at a rate* $o(\sqrt{\log n})$.

The following counterpart of Theorem 4 holds for testing problem (3.1.2).

Theorem 5. *Consider the testing problem defined in (3.1.2). Denote the null distribution by* $\mathbb{P}_{0,n}$, *the alternative distribution* $\mathbb{P}_{1,n}$ *and the likelihood ratio* $L_n = \frac{\mathrm{d}\mathbb{P}_{1,n}}{\mathrm{d}\mathbb{P}_{0,n}}$. *Under the condition of Theorem 4, whenever*

$$\|\Sigma_U^{-1/2}\widetilde{\Sigma}_U\Sigma_U^{-1/2}\widetilde{\Sigma}_\Theta\|_2 \|\Sigma_\Theta\|_2 < \gamma, \tag{3.3.5}$$

the three conclusions of Theorem 4 hold with $\mu_k = \mathsf{Tr}(\Sigma_\Theta^k)$ *and* $h_1 \geq \cdots \geq h_\kappa$ *the eigenvalues of* Σ_Θ.

It is not surprising that the results for the testing problems (3.1.1) and (3.1.2) are essentially the same. By law of large number, one has $\|\sqrt{p}(UU')^{-1/2} - \Sigma_U^{-1/2}\|_2 \xrightarrow{p} 0$. Hence $(UU')^{-1/2}$ is essentially same as $\frac{1}{\sqrt{p}}\Sigma_U^{-1/2}$ for large values of p. In addition, the distribution of X in

(3.1.2) remains unchanged if we replace U with $U\Sigma_U^{-1/2}$. Hence the testing problem (3.1.2) is essentially the same as (3.1.1) with $\Sigma_U = I_\kappa$.

It is worth noting that the bounds in (3.3.1) or (3.3.5) are sufficient but are not necessary in general. This is partly related to the second moment method used in the proof of these results. This being said, there are cases such as that in Corollary 1 for which these conditions are tight.

Sphericity test against multi-spiked Wishart covariance matrix Suppose π_Θ^0 assigns i.i.d. $N_\kappa(0, H)$ rows vectors in Θ where $H = \text{diag}(h_1, \dots, h_\kappa)$ and π_U^0 assigns i.i.d. $N_\kappa(0, I_\kappa)$ rows vectors in U. Then $V = U(U'U)^{-1/2}$ follows the uniform distribution on the Stiefel manifold $O(p, \kappa)$, and (3.1.2) reduces to the high-dimensional sphericity testing problem considered in Onatski et al. (2014) because in this case the full data likelihood ratio reduces to the likelihood ratio of the eigenvalues of the sample covariance matrix. Since $\pi_\Theta^0 \in \mathcal{P}(n, \kappa, H, H)$ and $\pi_U^0 \in \mathcal{P}(p, \kappa, I_\kappa, I_\kappa)$, we obtain the following corollary of Theorem 5 which reconstructs the key result in Onatski et al. (2014) for normal data for all fixed h values. However, the combinatorics approach we take to obtain the result is completely different from the analysis approach used in Onatski et al. (2014). On the other hand, Onatski et al. (2014) established uniform convergence over all h values in $[0, \sqrt{\gamma} - \epsilon]^r$ which our result does not cover.

Corollary 1. *Let $X_1, \dots, X_n \overset{iid}{\sim} N_p(0, \Sigma)$. Consider testing $H_0 : \Sigma = I_p$ vs. $H_1 : \Sigma = I_p + V'HV$ where $H = \text{diag}(h_1, \dots, h_\kappa)$ with $h_1 \geq \cdots \geq h_\kappa > 0$ and V follows the uniform distribution on $O(p, \kappa)$. Denote the null distribution by $\mathbb{P}_{0,n}$, the alternative distribution $\mathbb{P}_{1,n}$ and the likelihood ratio $L_n = \frac{d\mathbb{P}_{1,n}}{d\mathbb{P}_{0,n}}$. Suppose as $n \to \infty$, $p/n \to \gamma \in (0, \infty)$ while κ remains fixed. Then the conclusions of Theorem 5 hold whenever $h_1 < \sqrt{\gamma}$.*

Remark 5. Since our main theorems depend on the priors only through covariance matrices and sub-Gaussian variance proxies, Corollary 1 holds for any prior $\pi_U \in \mathcal{P}(p, \kappa, I_\kappa, I_\kappa)$, such as the prior that assigns entries of U with i.i.d. Rademacher random variables.

3.3.2. Polynomial time achievability of asymptotically optimal power

Theorems 4 and 5 give precise characterizations of the asymptotic behavior of log-likelihood ratios in (3.1.1) and (3.1.2) when SNRs are below the bounds specified in (3.3.1) and (3.3.5), respectively. However, even in one of the simplest cases where $\kappa = 1$ and both π_Θ and π_U assign i.i.d. Rademacher entries, direct evaluation of the likelihood ratio in (3.1.1) requires summing over 2^{n+p} different possible configurations of (Θ, U). Thus, directly performing the LRT can be computationally intractable. As a pleasant surprise, the following theorem ensures that under condition (3.3.1) or (3.3.5), one can achieve the asymptotically optimal power of the LRTs by using a special class of linear spectral statistics of the matrix $\frac{1}{n}X'X$. The result holds universally for all sub-Gaussian priors on Θ and U and simultaneously for testing problems (3.1.1) and (3.1.2). Therefore, for the weak SNR regimes defined by (3.3.1) and (3.3.5) respectively, one could always achieve the asymptotically optimal powers of LRTs for (3.1.1) and (3.1.2) by a test that is of time complexity at most $O((n+p)^3)$.

Theorem 6. *For testing problem* (3.1.1)*, suppose the conditions of Theorem 4 hold. When the SNR satisfies* (3.3.1)*, under both H_0 and H_1:*

$$\log(L_n) + \frac{1}{2}\sum_{l=1}^{\kappa} \Delta_p\left(Z(h_l)\right) - \frac{1}{2}\sum_{l_1=}^{\kappa}\sum_{l_2=1}^{\kappa} \log\left(1 - \frac{h_{l_1}h_{l_2}}{\gamma}\right) \xrightarrow{p} 0. \qquad (3.3.6)$$

Here $h_1 \geq \ldots \geq h_\kappa$ are eigenvalues of $\Sigma_\Theta \Sigma_U$, and for any $l \in \{1, 2, \ldots, \kappa\}$,

$$\Delta_p\left(Z(h_l)\right) := \sum_{i=1}^{p} \log\left(z(h_l) - \lambda_i\right) - p\int \log\left(z(h_l) - \lambda\right) dF^{\mathrm{MP}}(\lambda)$$

where $z(h_l) := (\gamma + h_l)(1 + h_l)/h_l$, $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\frac{1}{n}X'X$ and F^{MP} is the Marchenko-Pastur distribution **??***Anderson et al. (2010) with parameter γ.*

For testing problem (3.1.2)*, suppose the conditions of Theorem 5 hold. If the SNR satisfies* (3.3.5)*, then the conclusion* (3.3.6) *continues to hold under both H_0 and H_1 while $h_1 \geq \ldots \geq h_\kappa$ are eigenvalues of Σ_Θ.*

Remark 6. One might observe that the statistics $\Delta_p\left(Z(h_l)\right)$ is well defined with high probability under both H_0 and H_1. This is true because of the following reason. We know under H_0 the largest eigenvalue of $\frac{1}{n}X'X$ converges in probability to $\left(1+\sqrt{\gamma}\right)^2$. On the other hand the function $z(t) = (\gamma + t)(1 + t)/t$ achieves its minimum when $t = \sqrt{\gamma}$ and the corresponding minimum value is $(1 + \sqrt{\gamma})^2$. As a consequence, given $\Sigma_\Theta \Sigma_U$, $\min_{h \in \{h_1,\ldots,h_\kappa\}} z(h) > (1 + \sqrt{\gamma})^2$. Hence $\Delta_p\left(Z(h_l)\right)$ is well defined under H_0 with high probability. On the other hand by mutual contiguity of $\mathbb{P}_n$ and $\mathbb{Q}_n$, under the measure $\mathbb{Q}_n$ the largest eigenvalue of $\frac{1}{n}X'X$ also converges in probability to $\left(1+\sqrt{\gamma}\right)^2$. As a consequence,$\Delta_p\left(Z(h_l)\right)$ is well defined under H_1. Furthermore, there is no need to further scale the linear spectral statistics here which is a noticeable feature of linear spectral statistics. See for instance **?**.

The linear spectral statistic

$$-\frac{1}{2}\sum_{l=1}^{\kappa} \Delta_p\left(Z(h_l)\right) + \frac{1}{2}\sum_{l_1=}^{\kappa}\sum_{l_2=1}^{\kappa} \log\left(1 - \frac{h_{l_1}h_{l_2}}{\gamma}\right)$$

was first introduced in Onatski et al. (2014) for the special case of the testing problem considered in Corollary 1. Here we adopt the notation used in Onatski et al. (2014). The foregoing theorem shows that it approximates the log-likelihood ratio in probability in a much broader setting. Below the SNR bound, tests that reject the null for large values of this test statistic achieve the asymptotic power of the LRTs simultaneously for testing problems (3.1.1) and (3.1.2) for all sub-Gaussian priors on the signal component such that the spectrum of $\Sigma_\theta \Sigma_U$ (Σ_Θ when testing (3.1.2)) is given by $\{h_1, \ldots, h_\kappa\}$.

Remark 7. An exchange of the order of summation readily shows that the statistic in the last display is a linear spectral statistic. It is a function of $\{h_1, \ldots, h_\kappa\}$ and γ (which are given by the problem specification) as well as the eigenvalues $\lambda_1, \ldots, \lambda_p$ of $\frac{1}{n}X'X$ (which are given by data). Its computation is of polynomial time complexity.

3.4. Asymptotic normality of bipartite signed cycles

As we have announced earlier, we are to establish Theorems 4 and 5 by using bipartite signed cycles defined in Definition 15 as the set of "asymptotically sufficient" statistics in Proposition 2. To this end, we need the following asymptotic normality result for these statistics under both null and alternative.

Proposition 3. *Consider both testing problems* (3.1.1) *and* (3.1.2)*. Suppose that* $\pi_\Theta \in \mathcal{P}(n, \kappa, \Sigma_\Theta, \widetilde{\Sigma}_\Theta)$*,* $\pi_U \in \mathcal{P}(p, \kappa, \Sigma_U, \widetilde{\Sigma}_U)$*, and as* $n \to \infty$*,* $p/n \to \gamma \in (0, \infty)$ *while* $\kappa, \Sigma_\Theta, \Sigma_U, \widetilde{\Sigma}_\Theta, \widetilde{\Sigma}_U$ *remain fixed. Then for any fixed integer* $l > 0$*, the following results hold:*

(i) Under H_0*, when* $1 \le k_1 < \ldots < k_l = o(\sqrt{\log n})$*,*

$$\left(\frac{B_{n,k_1} - p\mathbf{1}_{k_1=1}}{\sqrt{2k_1\gamma^{k_1}}}, \ldots, \frac{B_{n,k_l}}{\sqrt{2k_l\gamma^{k_l}}} \right) \xrightarrow{d} N_l(0, I_l). \tag{3.4.1}$$

(ii) Under H_1*, when* $1 \le k_1 < \ldots < k_l = o(\sqrt{\log n})$*,*

$$\left(\frac{B_{n,k_1} - p\mathbf{1}_{k_1=1} - \mu_{k_1}}{\sqrt{2k_1\gamma^{k_1}}}, \ldots, \frac{B_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l\gamma^{k_l}}} \right) \xrightarrow{d} N_l(0, I_l), \tag{3.4.2}$$

where for testing problem (3.1.1)*,*

$$\mu_k = \mathsf{Tr}\left((\Sigma_\Theta\Sigma_U)^k\right), \tag{3.4.3}$$

and for testing problem (3.1.2)

$$\mu_k = \mathsf{Tr}(\Sigma_\Theta^k). \tag{3.4.4}$$

In the foregoing proposition, the indices $\{k_1, \ldots, k_l\}$ are allowed to diverge to infinity together with n and p. Therefore, the results are stronger than what is required in condition (ii) of Proposition 2. This is also the reason that we are able to improve (3.2.3) to an asymptotic ANOVA type decomposition (3.3.4). Finally, the asymptotic normality of bipartite signed cycles holds under H_1 even when the SNR conditions (3.3.1) and (3.3.5) are not met.

3.5. Proof of Theorems 4 and 5

For conciseness, we focus on the proof of Theorem 4. For Theorem 5, in view of the discussion after its statement, the modification of the formula for μ_k for (3.1.2) is natural as we could in some sense treat this case similarly to (3.1.1) with $\Sigma_U = I_\kappa$. In particular, we could modify the proof below by considering a sequence of high probability events Ω_n such that $\|\mathbf{1}_{\Omega_n}\sqrt{p}(UU')^{-1/2} - \Sigma_U^{-1/2}\|_{\max} \le \delta_n \to 0$ and then establish all the weak limits on Ω_n. Here and after, for any matrix A, $\|A\|_{\max} = \max_{i,j}|A_{ij}|$ is its vector ℓ_∞ norm.

Throughout the proof, without further specification, all probability and expectation calculations are conducted with respect to $\mathbb{P}_{0,n}$, i.e., under the null hypothesis. For any two matrices $A = (a_{i,j}) \in \mathbb{R}^{m_1\times m_2}$ and $B = (b_{i,j}) \in \mathbb{R}^{n_1\times n_2}$, we define their Kronecker product $A \otimes B$ as

$$A \otimes B = \begin{pmatrix} a_{1,1}B & a_{1,2}B & \dots & a_{1,m_2}B \\ a_{2,1}B & a_{2,2}B & \dots & a_{2,m_2}B \\ \vdots & \vdots & \dots & \vdots \\ a_{m_1,1}B & a_{m_1,2}B & \dots & a_{m_1,m_2}B \end{pmatrix}.$$

In addition, $\mathrm{vec}(A) = (A'_{*1}, \dots, A'_{*m_2})' \in \mathbb{R}^{m_1 m_2 \times 1}$ is the vector obtained from stacking all column vectors of A in order.

3.5.1. Proof of parts 1 and 2

The proof technique we use here is a second moment method with conditioning which is similar in spirit to those used in **?** and **?**. Recall that $p = p_n$ is a sequence depending on n. In this proof we shall use the following two sequences of σ-fields:

$$\mathcal{G}_n = \sigma\left(\{X_i\}_{i=1}^n\right), \qquad \mathcal{F}_n = \sigma\big(\{\Theta_{i*}\}_{i=1}^n, \{U_{j*}\}_{j=1}^p\big). \tag{3.5.1}$$

Here the X_i's are the observed data which consist of the rows of X defined in (3.1.1). It is straightforward to verify that

$$L_n = \mathbb{E}[L_n^f | \mathcal{G}_n] \tag{3.5.2}$$

where the expectation is taken over Θ and U and

$$L_n^f := \exp\left\{ \sum_{i=1}^{n} \sum_{j=1}^{p} \left(X_{i,j} M_{i,j} - \frac{1}{2} M_{i,j}^2 \right) \right\}$$

for

$$M_{i,j} = \frac{1}{\sqrt{p}} \sum_{l=1}^{\kappa} \Theta_{i,l} U_{j,l}. \tag{3.5.3}$$

Step 1. We now consider any sequence of events $E_n \in \mathcal{F}_n$ such that $\mathbb{P}\left[E_n^c\right] \to 0$ as $n \to \infty$. An explicit description of the E_n's of our interest will be given in step 2. Now define

$$\widetilde{L}_n := \mathbb{E}\left[L_n^f \mathbf{1}_{E_n} \mid \mathcal{G}_n\right].$$

In the rest of this step, we argue that it suffices to prove the desired results for $\widetilde{L}_n$. Note that the measure $\widetilde{\mathbb{Q}}_n$ on $\mathcal{G}_n$ defined as

$$\widetilde{\mathbb{Q}}_n(A_n) = \frac{1}{\mathbb{P}[E_n]} \mathbb{E}_{\mathbb{P}_{0,n}}\left[\widetilde{L}_n \mathbf{1}_{A_n}\right], \qquad \forall\ A_n \in \mathcal{G}_n,$$

is a probability measure. By definition,

$$\begin{aligned} 0 &\le \left| \mathbb{P}_{1,n}(A_n) - \widetilde{\mathbb{Q}}_n(A_n) \right| \\ &\le \frac{1}{\mathbb{P}\left[E_n\right]} \mathbb{E}_{\mathbb{P}_{0,n}}[(L_n - \widetilde{L}_n)\mathbf{1}_{A_n}] + \mathbb{P}_{1,n}(A_n) \frac{\mathbb{P}[E_n^c]}{\mathbb{P}[E_n]} \\ &\le \frac{1}{\mathbb{P}\left[E_n\right]} \mathbb{E}_{\mathbb{P}_{0,n}}\left[L_n - \widetilde{L}_n\right] + \frac{\mathbb{P}\left[E_n^c\right]}{\mathbb{P}\left[E_n\right]} = \frac{1}{\mathbb{P}\left[E_n\right]} \mathbb{E}\left[L_n^f \mathbf{1}_{E_n^c}\right] + \frac{\mathbb{P}\left[E_n^c\right]}{\mathbb{P}\left[E_n\right]} \\ &= \frac{1}{\mathbb{P}\left[E_n\right]} \mathbb{E}\Big[\mathbf{1}_{E_n^c} \mathbb{E}\left[L_n^f | \mathcal{F}_n\right]\Big] + \frac{\mathbb{P}\left[E_n^c\right]}{\mathbb{P}\left[E_n\right]} = 2\, \frac{\mathbb{P}\left[E_n^c\right]}{\mathbb{P}\left[E_n\right]}. \end{aligned} \tag{3.5.4}$$

In other words, the total variation distance between $\mathbb{P}_{1,n}$ and $\widetilde{\mathbb{Q}}_n$ converges to zero. Here the third inequality holds since $\widetilde{L}_n \le L_n$ almost surely under $\mathbb{P}_{0,n}$. As a consequence, Proposition 3 implies that for any fixed $l \in \mathbb{N}$ and any $1 \le k_1 < \cdots < k_l = o(\sqrt{\log(n)})$, under $\widetilde{\mathbb{Q}}_n$,

$$\left(\frac{B_{n,k_1} - p\mathbf{1}_{k_1=1} - \mu_{k_1}}{\sqrt{2k_1\gamma^{k_1}}}, \ldots, \frac{B_{n,k_l} - \mu_{k_l}}{\sqrt{2k_l\gamma^{k_l}}} \right) \xrightarrow{d} N_l(0, I_l).$$

Now if one can choose E_n in such a way that

$$\limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_{0,n}} \left[\widetilde{L}_n^2 \right] = \limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_{0,n}} \left[\left(\frac{1}{\mathbb{P}\left[E_n\right]} \widetilde{L}_n \right)^2 \right] \le \exp\left\{ \sum_{k=1}^{\infty} \frac{\mu_k^2}{2k\gamma^k} \right\}, \tag{3.5.5}$$

then one can use Proposition 2 to conclude that

$$\frac{1}{\mathbb{P}\left[E_n\right]} \widetilde{L}_n \bigg| \mathbb{P}_{0,n} \xrightarrow{d} \exp\left\{ \sum_{k=1}^{\infty} \frac{2\mu_k Z_k - \mu_k^2}{4k\gamma^k} \right\}.$$

Hence, Slutsky's theorem implies that $\widetilde{L}_n \,|\, \mathbb{P}_{0,n}$ converges in distribution to the same limit. Then it remains to prove that

$$L_n - \widetilde{L}_n \,|\, \mathbb{P}_{0,n} \xrightarrow{p} 0.$$

Observe that $L_n \ge \widetilde{L}_n$ almost surely under $\mathbb{P}_{0,n}$. For any $\epsilon > 0$ let $A = \{L_n - \widetilde{L}_n > \epsilon\}$. By definition $|\mathbb{P}_{1,n}(A) - \mathbb{Q}_n(A)| \le d_{\text{TV}}\left(\mathbb{P}_{1,n}, \mathbb{Q}_n\right)$, where $d_{\text{TV}}(P, Q)$ stands for the total variation distance between probability measures P and Q. Then we have

$$\frac{1}{\mathbb{P}[E_n]} \epsilon \mathbb{P}_{0,n}(A) - \frac{\mathbb{P}[E_n^c]}{\mathbb{P}[E_n]} \mathbb{P}_{1,n}(A) \le d_{\text{TV}}\left(\mathbb{P}_{1,n}, \mathbb{Q}_n\right).$$

Rearranging terms leads to

$$\mathbb{P}_{0,n}(A) \le \frac{1}{\epsilon} \left[d_{\text{TV}}\left(\mathbb{P}_{1,n}, \mathbb{Q}_n\right) + \frac{\mathbb{P}[E_n^c]}{\mathbb{P}[E_n]} \right] \to 0. \tag{3.5.6}$$

Step 2. Now we prove (3.5.5) by making appropriate choices of the sequence of E_n's which guarantee Θ and U matrices are well-behaved. First observe that

$$\begin{aligned}
\mathbb{E}\big[\widetilde{L}_n^2\big] &= \mathbb{E}\Big[\mathbb{E}\big[L_n^f \mathbf{1}_{E_n} \,|\, \mathcal{G}_n\big]^2\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[L_n^{f(1)} L_n^{f(2)} \mathbf{1}_{E_n^{(1)}} \mathbf{1}_{E_n^{(2)}} \,|\, \mathcal{G}_n\big]\Big] \\
&= \mathbb{E}\Big[L_n^{f(1)} L_n^{f(2)} \mathbf{1}_{E_n^{(1)}} \mathbf{1}_{E_n^{(2)}}\Big] \\
&= \mathbb{E}\Big[\mathbf{1}_{E_n^{(1)}} \mathbf{1}_{E_n^{(2)}} \mathbb{E}\big[L_n^{f(1)} L_n^{f(2)} \,|\, \mathcal{F}_n\big]\Big].
\end{aligned} \tag{3.5.7}$$

Here $L_n^{f(1)}$ and $L_n^{f(2)}$ are two independent copies of L_n^f where the X_i's are kept fixed but one takes two i.i.d. copies of the Θ's and U's. *This is feasible (only) under the null hypothesis when the X_i's are equal to the Z_i's and hence are independent of Θ and U.* With slight abuse of notation, we use $\mathcal{F}_n$ to denote the σ-field generated by both copies. We call the corresponding random variables $\{\Theta^{(1)}, U^{(1)}\}$ and $\{\Theta^{(2)}, U^{(2)}\}$. For any matrices A and B of the same dimensions, let $\langle A, B\rangle = \mathsf{Tr}(A'B)$ be the trace inner product. Then we have

$$\begin{aligned}
&\mathbb{E}\left[L_n^{f(1)} L_n^{f(2)} |\mathcal{F}_n\right] \\
&= \exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\sum_{l=1}^{\kappa}\frac{1}{\sqrt{p}}\Theta_{i,l}^{(1)}U_{j,l}^{(1)}\right)\left(\sum_{l=1}^{\kappa}\frac{1}{\sqrt{p}}\Theta_{i,l}^{(2)}U_{j,l}^{(2)}\right)\right\} \\
&= \exp\left\{\sum_{l_1=1}^{\kappa}\sum_{l_2=1}^{\kappa}\frac{1}{p}\langle\Theta_{*l_1}^{(1)}, \Theta_{*l_2}^{(2)}\rangle\langle U_{*l_1}^{(1)}, U_{*l_2}^{(2)}\rangle\right\} \\
&:= \psi_n = \psi_n(\Theta^{(1)}, \Theta^{(2)}, U^{(1)}, U^{(2)}).
\end{aligned} \tag{3.5.8}$$

Define

$$\begin{aligned}
E_n^{(1)} := \Bigg\{\max_{1\le l_1,l_2\le\kappa}\Bigg(&\left|\frac{1}{n}\langle\Theta_{*l_1}^{(1)}, \Theta_{*l_2}^{(1)}\rangle - \Sigma_\Theta(l_1,l_2)\right|, \\
&\left|\frac{1}{p}\langle U_{*l_1}^{(1)}, U_{*l_2}^{(1)}\rangle - \Sigma_U(l_1,l_2)\right|\Bigg) \le \delta_n\Bigg\}
\end{aligned} \tag{3.5.9}$$

where $\delta_n \to 0$ and $\mathbb{P}((E_n^{(1)})^c) \to 0$ as $n \to \infty$. Such a sequence of δ_n exists due to law of large numbers. Define $E_n^{(2)}$ as an identical and independent copy of $E_n^{(1)}$ that depends on $\Theta^{(2)}, U^{(2)}$. Conditioning on $\Theta^{(1)}$, $U^{(1)}$ and $U^{(2)}$, the exponent of the rightmost expression in (3.5.8) can be written as

$$\sqrt{\frac{n}{p}}\,\langle Z, V\rangle \tag{3.5.10}$$

where

$$V = A\,\mathrm{vec}(U^{(2)}) \in \mathbb{R}^{\kappa^2} \qquad \text{for} \quad A = \frac{1}{\sqrt{p}}\, I_\kappa \otimes (U^{(1)})' \in \mathbb{R}^{\kappa^2\times\kappa p}, \tag{3.5.11}$$

$$Z = B\,\mathrm{vec}(\Theta^{(2)}) \in \mathbb{R}^{\kappa^2} \qquad \text{for} \quad B = \frac{1}{\sqrt{n}}\, I_\kappa \otimes (\Theta^{(1)})' \in \mathbb{R}^{\kappa^2\times\kappa n}. \tag{3.5.12}$$

Our goal is to prove the random variables $\{\psi_n\,\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{E_n^{(2)}} : n \geq 1\}$ are uniformly integrable. To this end, it suffices to show that $\mathbb{E}[\psi_n^{1+\eta}\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{E_n^{(2)}}]$ is uniformly bounded for some $\eta > 0$. Now from the sub-Gaussianity assumption on the priors, we have for sufficiently large values of n,

$$\begin{aligned}
&\mathbb{E}\left[\psi_n^{1+\eta}\,\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{E_n^{(2)}}\middle|\Theta^{(1)}, U^{(1)}, U^{(2)}\right]\\
&= \mathbb{E}\left[\psi_n^{1+\eta}\,\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{\widetilde{E}_n^{(2)}}\mathbf{1}_{\widehat{E}_n^{(2)}}\middle|\Theta^{(1)}, U^{(1)}, U^{(2)}\right]\\
&\leq \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{\widetilde{E}_n^{(2)}}\exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'BD_\Theta B'V\right\}\middle|\Theta^{(1)}, U^{(1)}, U^{(2)}\right].
\end{aligned} \tag{3.5.13}$$

Here

$$\begin{aligned}
\widetilde{E}_n^{(2)} &= \left\{\max_{1\leq l_1,l_2\leq\kappa}\left|\frac{1}{p}\langle U_{*l_1}^{(2)}, U_{*l_2}^{(2)}\rangle - \Sigma_U(l_1,l_2)\right| \leq \delta_n\right\},\\
\widehat{E}_n^{(2)} &= \left\{\max_{1\leq l_1,l_2\leq\kappa}\left|\frac{1}{n}\langle \Theta_{*l_1}^{(2)}, \Theta_{*l_2}^{(2)}\rangle - \Sigma_\Theta(l_1,l_2)\right| \leq \delta_n\right\}
\end{aligned}$$

and

$$D_\Theta = \widetilde{\Sigma}_\Theta \otimes I_n \in \mathbb{R}^{\kappa n\times\kappa n}. \tag{3.5.14}$$

As a consequence, for B defined in (3.5.12), we have

$$BD_\Theta B' = \widetilde{\Sigma}_\Theta \otimes \left[\frac{1}{n}(\Theta^{(1)})'\Theta^{(1)}\right]. \tag{3.5.15}$$

Recall that for any matrix A, $\|A\|_{\max} = \max_{i,j} |A_{ij}|$ is the vector ℓ_∞-norm of A. On the event $\widetilde{E}_n^{(2)} \cap E_n^{(1)}$, we have $\|BD_\Theta B' - \widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta\|_{\max} < \|\widetilde{\Sigma}_\Theta\|_{\max}\delta_n$. Now we know that for any symmetric matrix Σ of dimension $\kappa^2 \times \kappa^2$, $\|\Sigma\|_2 \leq \|\Sigma\|_{\mathrm{F}} \leq \kappa^2\|\Sigma\|_{\max}$ where $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ denote the spectral norm and Frobenius norm respectively. So

$$\begin{aligned}\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{\widetilde{E}_n^{(2)}} V'BD_\Theta B'V &\leq \mathbf{1}_{E_n^{(1)}}\mathbf{1}_{\widetilde{E}_n^{(2)}} V'\Big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \kappa^2\|\widetilde{\Sigma}_\Theta\|_{\max}\delta_n I_{\kappa^2}\Big)V \\ &\leq \mathbf{1}_{E_n^{(1)}}\mathbf{1}_{\widetilde{E}_n^{(2)}} V'\Big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\Big)V\end{aligned}$$

where $\delta_n' \to 0$ is a sequence depending only on κ, $\widetilde{\Sigma}_\Theta$ and δ_n. Therefore, we have

$$\begin{aligned}&\mathbb{E}\left[\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{\widetilde{E}_n^{(2)}} \exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'BD_\Theta B'V\right\}\right] \\ &\leq \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{\widetilde{E}_n^{(2)}} \exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'\big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\big)V\right\}\right] \\ &= \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}}\mathbb{E}\left[\mathbf{1}_{\widetilde{E}_n^{(2)}} \exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'\big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\big)V\right\}\Big|\Theta^{(1)}, U^{(1)}\right]\right] \\ &\leq \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}}\mathbb{E}\left[\exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'\big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\big)V\right\}\Big|\Theta^{(1)}, U^{(1)}\right]\right] \\ &= \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}} \exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'\big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\big)V\right\}\right].\end{aligned}$$

In Step 3 we are to prove that for some $\eta > 0$ the sequence

$$\limsup_{n\to\infty} \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}} \exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'\big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\big)V\right\}\right] < \infty \tag{3.5.16}$$

under the assumption that $\|\widetilde{\Sigma}_\Theta\widetilde{\Sigma}_U\|_2\|\Sigma_\Theta\Sigma_U\|_2 < \gamma$.

If we assume (3.5.16), then uniform integrability of $\{\psi_n \mathbf{1}_{E_n^{(1)}}\mathbf{1}_{E_n^{(2)}} : n \geq 1\}$ is established. The rest of the proof can be completed as follows. Observe that by central limit theorem, we have $\frac{1}{\sqrt{n}}\langle\Theta_{*l_1}^{(1)}, \Theta_{*l_2}^{(2)}\rangle \xrightarrow{d} T_{l_1,l_2}$ and $\frac{1}{\sqrt{p}}\langle U_{*l_1}^{(1)}, U_{*l_2}^{(2)}\rangle \xrightarrow{d} Y_{l_1,l_2}$, and so by continuous mapping,

$$\psi_n \mathbf{1}_{E_n^{(1)}}\mathbf{1}_{E_n^{(2)}} \xrightarrow{d} \exp\left\{\frac{1}{\sqrt{\gamma}}\sum_{l_1=1}^{\kappa}\sum_{l_2=1}^{\kappa} T_{l_1,l_2}Y_{l_1,l_2}\right\}.$$

In addition, the collections $\{T_{l_1,l_2} : 1 \leq l_1, l_2 \leq \kappa\}$ and $\{Y_{l_1,l_2} : 1 \leq l_1, l_2 \leq \kappa\}$ are mutually independent. Furthermore, the random variables $\{T_{l_1,l_2} : 1 \leq l_1, l_2 \leq \kappa\}$ are jointly Gaussian with mean 0 and $\mathrm{Cov}(T_{l_1,l_2}, T_{l_3,l_4}) = \Sigma_\Theta(l_1,l_3)\Sigma_\Theta(l_2,l_4)$ and analogous results hold for $\{Y_{l_1,l_2} : 1 \leq l_1, l_2 \leq \kappa\}$. Let $T = (T_{l_1,l_2})$ and $Y = (Y_{l_1,l_2})$ be $\kappa \times \kappa$ matrices. Then the foregoing discussion implies that $\mathrm{vec}(T) \sim N_{\kappa^2}(0, \Sigma_\Theta \otimes \Sigma_\Theta)$ and is independent of $\mathrm{vec}(Y) \sim N_{\kappa^2}(0, \Sigma_U \otimes \Sigma_U)$. This, together with the uniform integrability of $\psi_n \mathbf{1}_{E_n^{(1)}} \mathbf{1}_{E_n^{(2)}}$, implies that

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}}\mathbf{1}_{E_n^{(2)}}\psi_n\right] &= \mathbb{E}\left[\exp\left\{\frac{1}{\sqrt{\gamma}}\langle \mathrm{vec}(T), \mathrm{vec}(Y)\rangle\right\}\right] \\ &= \mathbb{E}\left[\exp\left\{\frac{1}{2\gamma}\mathrm{vec}(Y)'(\Sigma_\Theta \otimes \Sigma_\Theta)\mathrm{vec}(Y)\right\}\right] \\ &= \exp\left\{-\frac{1}{2}\sum_{i=1}^{\kappa^2}\log\left(1 - \frac{s_i}{\gamma}\right)\right\}.\end{aligned}$$

The Taylor series of $\log(1-x)$ further gives that the last display equals

$$\exp\left\{\sum_{k=1}^{\infty}\frac{\sum_{i=1}^{\kappa^2} s_i^k}{2k\gamma^k}\right\} = \exp\left\{\sum_{k=1}^{\infty}\frac{\mathsf{Tr}\left((\Sigma_\Theta\Sigma_U)^k \otimes (\Sigma_\Theta\Sigma_U)^k\right)}{2k\gamma^k}\right\}.$$

Here $\{s_i\}_{1\leq i\leq\kappa^2}$ are the eigenvalues of the matrix $(\Sigma_U \otimes \Sigma_U)^{1/2}(\Sigma_\Theta \otimes \Sigma_\Theta)(\Sigma_U \otimes \Sigma_U)^{1/2}$. We complete the proof by noting that $\mathsf{Tr}((\Sigma_\Theta\Sigma_U)^k \otimes (\Sigma_\Theta\Sigma_U)^k) = [\mathsf{Tr}((\Sigma_\Theta\Sigma_U)^k)]^2 = \mu_k^2$.

Step 3. In the final step of the proof, we verify (3.5.16). Recall (3.5.11) and observe that

$$\begin{aligned}&\exp\left\{\frac{1}{2\gamma}(1+2\eta)^2 V'\big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\big)V\right\} \\ &= \exp\left\{\frac{1}{2\gamma}(1+2\eta)^2\mathrm{vec}(U^{(2)})'A'\big(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\big)A\mathrm{vec}(U^{(2)})\right\}.\end{aligned} \tag{3.5.17}$$

Write $\widetilde{U}^{(2)} = D_U^{-1/2}U^{(2)}$ where $D_U = \widetilde{\Sigma}_U \otimes I_p \in \mathbb{R}^{\kappa p\times\kappa p}$. We have

$$\begin{aligned}&\exp\left\{\frac{1}{2\gamma}(1+2\eta)^2(\mathrm{vec}(U)^{(2)})'A'\left(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\right)A\mathrm{vec}(U)^{(2)}\right\} \\ &= \exp\left\{\frac{1}{2\gamma}(1+2\eta)^2\mathrm{vec}(\widetilde{U}^{(2)})'D_U^{1/2}A'\left(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\right)AD_U^{1/2}\mathrm{vec}(\widetilde{U}^{(2)})\right\}.\end{aligned}$$

Theorem 1 from **?** implies for any non-random non-negative definite $\widetilde{\Sigma}$ and all $t > 0$,

$$\mathbb{P}\left(\mathrm{vec}(\widetilde{U}^{(2)})'\widetilde{\Sigma}\mathrm{vec}(\widetilde{U}^{(2)}) > \mathsf{Tr}(\widetilde{\Sigma}) + \sqrt{\mathsf{Tr}(\widetilde{\Sigma}^2)t} + 2\|\widetilde{\Sigma}\|_2 t\right) \leq e^{-t}. \tag{3.5.18}$$

In particular, the tail bound in (3.5.18) only depends on the nonzero eigenvalues of $\widetilde{\Sigma}$. Now the nonzero eigenvalues of

$$D_U^{1/2}A'\left(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\right)AD_U^{1/2}$$

are the same as those of

$$AD_U A'\left(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2}\right). \tag{3.5.19}$$

On $E_n^{(1)}$, we have $AD_U A' = \widetilde{\Sigma}_U \otimes \Sigma_U + P$ where P is a perturbation matrix with $\|P\|_{\max} = O(\delta_n)$. As a consequence, Theorem 5.5.4 of **?** implies that the nonzero eigenvalues of (3.5.19) are the eigenvalues of

$$(\widetilde{\Sigma}_U \otimes \Sigma_U)(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta) + O(\delta_n).$$

Here the constant in the $O(\delta_n)$ term depends on the eigenvalues of $(\widetilde{\Sigma}_U \otimes \Sigma_U)(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta)$ and γ, but not on n and p.

For convenience, we define $\widetilde{\Sigma} := D_U^{1/2}A'(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta + \delta_n' I_{\kappa^2})AD_U^{1/2}$. On $E_n^{(1)}$, $\mathsf{Tr}(\widetilde{\Sigma})$ and $\mathsf{Tr}(\widetilde{\Sigma}^2)$ are uniformly bounded. So given any $\epsilon > 0$, there exists a sufficiently large $t_0 > 0$ that is independent of n such that for all $t \geq t_0$,

$$\mathsf{Tr}(\widetilde{\Sigma}) + \sqrt{\mathsf{Tr}(\widetilde{\Sigma}^2)t} \leq t\epsilon.$$

So for all $t > t_0$ we have

$$\mathbf{1}_{E_n^{(1)}}\mathbb{P}\left(\mathrm{vec}(\widetilde{U}^{(2)})'\widetilde{\Sigma}\mathrm{vec}(\widetilde{U}^{(2)}) > \left(2\|\widetilde{\Sigma}\|_2 + \epsilon\right)t \,\middle|\, \Theta^{(1)}, U^{(1)}\right) \leq e^{-t},$$

and hence for

$$\log(t) > \frac{(1+2\eta)^2 \left(2\|\widetilde{\Sigma}\|_2 + \epsilon\right) t_0}{2\gamma},$$

$$\begin{aligned}
&\mathbf{1}_{E_n^{(1)}} \mathbb{P}\left(\exp\left\{ \frac{1}{2\gamma}(1+2\eta)^2 \mathrm{vec}(\widetilde{U}^{(2)})' \widetilde{\Sigma} \mathrm{vec}(\widetilde{U}^{(2)}) \right\} > t \,\middle|\, \Theta^{(1)}, U^{(1)} \right) \\
&= \mathbf{1}_{E_n^{(1)}} \mathbb{P}\left(\mathrm{vec}(\widetilde{U}^{(2)})' \widetilde{\Sigma} \mathrm{vec}(\widetilde{U}^{(2)}) > \frac{2\gamma \log t}{(1+2\eta)^2} \,\middle|\, \Theta^{(1)}, U^{(1)} \right) \\
&= \mathbf{1}_{E_n^{(1)}} \mathbb{P}\left(\mathrm{vec}(\widetilde{U}^{(2)})' \widetilde{\Sigma} \mathrm{vec}(\widetilde{U}^{(2)}) > \left(2\|\widetilde{\Sigma}\|_2 + \epsilon\right) \frac{2\gamma \log t}{\left(2\|\widetilde{\Sigma}\|_2 + \epsilon\right)(1+2\eta)^2} \,\middle|\, \Theta^{(1)}, U^{(1)} \right) \\
&\le \left(\frac{1}{t}\right)^{\frac{2}{(1+2\eta)^2 (2\|\widetilde{\Sigma}\|_2/\gamma + \epsilon/\gamma)}}.
\end{aligned} \tag{3.5.20}$$

Since $\|(\widetilde{\Sigma}_U \otimes \Sigma_U)(\widetilde{\Sigma}_\Theta \otimes \Sigma_\Theta)\|_2 < \gamma$, we can choose ϵ and η small enough such that on $E_n^{(1)}$,

$$\frac{2}{(1+2\eta)^2 (2\|\widetilde{\Sigma}\|_2/\gamma + \epsilon/\gamma)} \ge \alpha_0 > 1.$$

Hence we have the last expression in (3.5.20) is bounded from above by $t^{-\alpha_0}$. We know that

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{E_n^{(1)}} \exp\left\{ \frac{1}{2\gamma}(1+2\eta)^2 \mathrm{vec}(\widetilde{U}^{(2)})' \widetilde{\Sigma} \mathrm{vec}(\widetilde{U}^{(2)}) \right\} \right] \\
&= \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}} \mathbb{E}\left[\exp\left\{ \frac{1}{2\gamma}(1+2\eta)^2 \mathrm{vec}(\widetilde{U}^{(2)})' \widetilde{\Sigma} \mathrm{vec}(\widetilde{U}^{(2)}) \right\} \,\middle|\, \Theta^{(1)}, U^{(1)} \right] \right] \\
&= \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}} \int_0^\infty \mathbb{P}\left[\exp\left\{ \frac{1}{2\gamma}(1+2\eta)^2 \mathrm{vec}(\widetilde{U}^{(2)})' \widetilde{\Sigma} \mathrm{vec}(\widetilde{U}^{(2)}) \right\} > t \,\middle|\, \Theta^{(1)}, U^{(1)} \right] \right] \\
&\le \mathbb{E}\left[\mathbf{1}_{E_n^{(1)}} \left(t_0 + \int_{t_0}^\infty \frac{1}{t^{\alpha_0}} dt \right) \right] \\
&\le t_0 + \frac{1}{\alpha_0 - 1} \frac{1}{t_0^{\alpha_0 - 1}}.
\end{aligned} \tag{3.5.21}$$

As a consequence,

$$\mathbb{E}\left[\mathbf{1}_{E_n^{(1)}} \exp\left\{ \frac{1}{2\gamma}(1+2\eta)^2 \mathrm{vec}(\widetilde{U}^{(2)})' \widetilde{\Sigma} \mathrm{vec}(\widetilde{U}^{(2)}) \right\} \right]$$

is uniformly bounded. This completes the proof.

3.5.2. Proof of part 3

By (3.2.3), for any given $\epsilon, \delta > 0$ there exists $K = K(\epsilon, \delta)$ and for any subsequence n_l there exists a further subsequence n_{l_q} such that

$$\mathbb{P}_{n_{l_q}}\left(\left|\log(L_{n_{l_q}}) - \sum_{k=1}^{K} \frac{2\mu_k(B_{n_{l_q},k} - p\mathbf{1}_{k=1}) - \mu_k^2}{2\sigma_k^2}\right| \geq \frac{\epsilon}{2}\right) \leq \frac{\delta}{2}. \tag{3.5.22}$$

Now choose $K' \geq K$ such that

$$\sum_{K'+1}^{\infty} \frac{\mu_k^2}{\sigma_k^2} \leq \max\left\{\frac{\delta\epsilon^2}{100}, \frac{\epsilon}{100}\right\}.$$

For any $K' < k_1 < k_2 < m_n = o(\sqrt{\log n})$, the proof of Proposition 3 in the supplement **?** implies that $\mathbb{E}_{\mathbb{P}_n}[B_{n,k_1}] = 0$, $\mathrm{Cov}(B_{n,k_1}, B_{n,k_2}) = 0$ and $\mathrm{Var}(B_{n,k_i}) = 2k_1\gamma^{k_i}(1 + O(k_i^2/n))$ for $i \in \{1, 2\}$. So

$$\mathrm{Var}\left(\sum_{k=K'+1}^{m_{n_{l_q}}} \frac{2\mu_k B_{n_{l_q},k} - \mu_k^2}{2\sigma_k^2}\right) = (1 + o(1)) \sum_{k=K'+1}^{m_{n_{l_q}}} \frac{\mu_k^2}{\sigma_k^2} \leq \frac{\delta\epsilon^2}{100}.$$

Now for large values of n_{l_q},

$$\begin{aligned} &\mathbb{P}_{n_{l_q}}\left(\left|\sum_{k=K+1}^{m_{n_{l_q}}} \frac{2\mu_k B_{n_{l_q},k}}{\sigma_k^2}\right| \geq \frac{\epsilon}{4}\right) \leq \frac{16\delta\epsilon^2}{100\epsilon^2}, \qquad \text{and so} \\ &\mathbb{P}_{n_{l_q}}\left(\left|\sum_{k=K+1}^{m_{n_{l_q}}} \frac{2\mu_k B_{n_{l_q},k} - \mu_k^2}{2\sigma_k^2}\right| \geq \frac{\epsilon}{4} + \frac{\epsilon}{100}\right) \leq \frac{16\delta\epsilon^2}{100\epsilon^2}. \end{aligned} \tag{3.5.23}$$

Plugging in the estimates of (3.5.22) and (3.5.23) we have for all large values of n_{l_q},

$$\mathbb{P}_{n_{l_q}}\left(\left|\log(L_{n_{l_q}}) - \sum_{k=1}^{m_{n_{l_q}}} \frac{2\mu_k(B_{n_{l_q},k} - p\mathbf{1}_{k=1}) - \mu_k^2}{2\sigma_k^2}\right| \geq \epsilon\right) \leq \delta. \tag{3.5.24}$$

Since (3.5.24) occurs to any subsequence and any (ϵ, δ) pair, this completes the proof. $\square$

3.6. Proof of Theorem 6

The proof relies on two key ingredients: (i) the asymptotic ANOVA type decomposition for $\log(L_n)$ that holds below the SNR bound for both testing problems (3.1.1) and (3.1.2), and (ii) the approximation of log-likelihood ratio of *sample eigenvalues* (as opposed to full data matrix) by linear spectral statistics in the multiple spiked Wishart model setting. The former was established in Theorems 4 and 5. The latter was first proved in Onatski et al. (2014).

Step 1. For any sample size n and dimension $p = p_n$, let $\mathbb{P}_{0,n}$ denote the null distribution of the data matrix X in both (3.1.1) and (3.1.2) as they are identical. Given a pair of sub-Gaussian priors π_Θ and π_U such that condition (3.3.1) is met, we denote the alternative distribution of X in (3.1.1) by $\mathbb{P}^M_{1,n}(\pi_\Theta, \pi_U)$. If condition (3.3.5) is met, we denote the alternative distribution of X in (3.1.2) by $\mathbb{P}^C_{1,n}(\pi_\Theta, \pi_U)$. The subscript indicates whether it is the mean version (M) or the covariance version (C) of the detection problem. By Theorems 4 and 5, for any sub-Gaussian priors (π^1_Θ, π^1_U) and (π^2_Θ, π^2_U) satisfying (3.3.1) and any $(\widetilde{\pi}^1_\Theta, \widetilde{\pi}^1_U)$ and $(\widetilde{\pi}^2_\Theta, \widetilde{\pi}^2_U)$ satisfying (3.3.5), we have

$$\mathbb{P}^M_{1,n}(\pi^1_\Theta, \pi^1_U) \ \triangleleft\triangleright\ \mathbb{P}^M_{1,n}(\pi^2_\Theta, \pi^2_U) \ \triangleleft\triangleright\ \mathbb{P}_{0,n} \ \triangleleft\triangleright\ \mathbb{P}^C_{1,n}(\widetilde{\pi}^1_\Theta, \widetilde{\pi}^1_U) \ \triangleleft\triangleright\ \mathbb{P}^C_{1,n}(\widetilde{\pi}^2_\Theta, \widetilde{\pi}^2_U). \tag{3.6.1}$$

In other words, they are all mutually contiguous with respect to one another. In what follows, unless otherwise specified, we always work with priors satisfying (3.3.1) for problem (3.1.1), and (3.3.5) for (3.1.2). Hence, we may write $\mathbb{P}^M_{1,n}$ and $\mathbb{P}^C_{1,n}$ directly without causing confusion.

By Theorems 4 and 5 we have

$$\log(L_n) - \sum_{k=1}^{m_n} \frac{2\mu_k \left(B_{n,k} - p\mathbf{1}_{k=1}\right) - \mu_k^2}{4k\gamma^k} \xrightarrow{p} 0 \tag{3.6.2}$$

under $\mathbb{P}_{0,n}$ and all $\mathbb{P}^M_{1,n}$ and $\mathbb{P}^C_{1,n}$ that satisfies the respective SNR condition. Here, $\mu_k =$

$\sum_{l=1}^{\kappa} h_l^k$. Note that the alternative model in (3.1.1) is identifiable up to a linear transform $U \to UA$ and $\Theta \to \Theta A^{-1}$ where A is any full rank $\kappa \times \kappa$ matrix. Hence, we may always choose $A = \Sigma_U^{-1/2}$ and thus assume that $\Sigma_U = I_\kappa$. Then $h_1 \geq \cdots \geq h_\kappa$ are ordered eigenvalues of Σ_Θ for both (3.1.1) and (3.1.2), and both (3.3.1) and (3.3.5) reduces to

$$h_1 \|\widetilde{\Sigma}_\Theta\|_2 < \gamma. \tag{3.6.3}$$

By the discussion following Definition 16, we always have $h_1 \leq \|\widetilde{\Sigma}_\Theta\|_2$ and hence (3.6.3) implies that $h_1 < \sqrt{\gamma}$.

In view of the foregoing discussions, it suffices to show that for any given $H = \text{diag}(h_1, \ldots, h_\kappa)$ such that $h_1 < \sqrt{\gamma}$, we can find a particular set of sub-Gaussian priors (π_Θ, π_U) satisfying $\Sigma_\Theta = H$, $\Sigma_U = I_\kappa$ and (3.6.3) such that

$$\sum_{k=1}^{m_n} \frac{2\mu_k \left(B_{n,k} - p\mathbf{1}_{k=1}\right) - \mu_k^2}{4k\gamma^k} + \frac{1}{2}\sum_{l=1}^{\kappa} \Delta_p \left(Z(h_l)\right) - \frac{1}{2}\sum_{l_1=}^{\kappa}\sum_{l_2=1}^{\kappa} \log\left(1 - \frac{h_{l_1} h_{l_2}}{\gamma}\right) \xrightarrow{p} 0 \tag{3.6.4}$$

under H_0 for testing problem (3.1.2). Then the definition of mutual contiguity and (3.6.1) implies that (3.6.4) also holds for any $\mathbb{P}_1^M(\pi_\Theta, \pi_U)$ such that the eigenvalues of $\Sigma_\Theta \Sigma_U$ are the h_l's and that (3.3.1) holds. Similarly, we have (3.6.4) hold for any $\mathbb{P}_1^C(\pi_\Theta, \pi_U)$ such that the eigenvalues of Σ_Θ are the h_l's and that (3.3.5) holds. This is true because given X and the h_l's, the left side of (3.6.4) is completely determined! Finally, the desired conclusion is a direct consequence of (3.6.2) and the fact that (3.3.1) or (3.3.5) implies that $h_1 < \sqrt{\gamma}$.

Step 2. Given any $H = \text{diag}(h_1, \ldots, h_\kappa)$ such that $h_1 < \sqrt{\gamma}$, we now work with a particular set of sub-Gaussian priors (π_Θ, π_U) satisfying $\Sigma_\Theta = H$, $\Sigma_U = I_\kappa$ and (3.6.3) such that (3.6.4) holds under H_0 for (3.1.2). In particular, let π_Θ assign i.i.d. $N_\kappa(0, H)$ row vectors in Θ and π_U assign i.i.d. $N_\kappa(0, I_\kappa)$ row vectors in U. Then (3.6.3) is satisfied as long as $h_1 < \sqrt{\gamma}$ since one can choose $\widetilde{\Sigma}_\Theta = H$. From Corollary 1 we have in this case

$$\log(L_n) - \sum_{k=1}^{m_n} \frac{2\mu_k \left(B_{n,k} - p\mathbf{1}_{k=1}\right) - \mu_k^2}{4k\gamma^k} \xrightarrow{p} 0.$$

On the other hand, the full data likelihood ratio reduces to the likelihood ratio of sample eigenvalues of $\frac{1}{n}X'X$ for these priors. So Theorem 3 of Onatski et al. (2014) implies that under H_0,

$$\log(L_n) + \frac{1}{2}\sum_{l=1}^{\kappa}\Delta_p\left(Z(h_l)\right) - \frac{1}{2}\sum_{l_1=}^{\kappa}\sum_{l_2=1}^{\kappa}\log\left(1 - \frac{h_{l_1}h_{l_2}}{\gamma}\right) \xrightarrow{p} 0.$$

As a consequence, we obtain the desired claim (3.6.4) by Slutsky's theorem. This concludes the proof. □

3.7. Proof of Proposition 1

At first we introduce the concept of Wasserstein's metric which will be used in the proof of Proposition 2. Let F and G be two distribution functions with finite p-th moment. Then the Wasserstein distance W_p between F and G is defined to be

$$W_p(F, G) = \left[\inf_{X\sim F, Y\sim G}\mathbb{E}|X - Y|^p\right]^{1/p}.$$

Here X and Y are random variables having distribution functions F and G respectively. The following result will be useful in our proof. See, for instance, Mallows (1972) for its proof.

Proposition 11. *Let F_n be a sequence of distribution functions and F be a distribution function. Then $F_n \xrightarrow{d} F$ in distribution and $\int x^2 dF_n(x) \to \int x^2 dF(x)$ if $W_2(F_n, F) \to 0$.*

Proof of Proposition 2. We now prove the proposition.

Proof of mutual contiguity and (3.2.2) This proof is broken into two steps. We focus on proving (3.2.2). Given (3.2.2), mutual contiguity is a direct consequence of Le Cam's first lemma Le Cam (2012).

Step 1. We first prove the random variable on the right hand side of (3.2.2) is almost

surely positive and has mean 1. Let us define

$$L := \exp\left\{\sum_{i=1}^{\infty} \frac{2\mu_i Z_i - \mu_i^2}{2\sigma_i^2}\right\}, \qquad L^{(m)} := \exp\left\{\sum_{i=1}^{m} \frac{2\mu_i Z_i - \mu_i^2}{2\sigma_i^2}\right\}, \quad \forall m \in \mathbb{N}.$$

As $Z_i \sim N(0, \sigma_i^2)$, for any $i \in \mathbb{N}$, and so

$$\mathbb{E}\left[\exp\left\{\frac{2\mu_i Z_i - \mu_i^2}{2\sigma_i^2}\right\}\right] = 1.$$

So $\{L^{(m)}\}_{m=1}^{\infty}$ is a martingale sequence and

$$\mathbb{E}\left[\left(L^{(m)}\right)^2\right] = \prod_{i=1}^{m} \exp\left\{\frac{\mu_i^2}{\sigma_i^2}\right\} = \exp\left\{\sum_{i=1}^{m} \frac{\mu_i^2}{\sigma_i^2}\right\}.$$

Now by the righthand side of (3.2.1), $L^{(m)}$ is a L^2 bounded martingale. Hence, L is a well defined random variable with

$$\mathbb{E}[L] = 1, \qquad \mathbb{E}[L^2] = \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\}.$$

On the other hand $\log(L)$ is a limit of Gaussian random variables, hence $\log(L)$ is Gaussian with

$$\mathbb{E}[\log(L)] = -\frac{1}{2}\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}, \qquad \mathrm{Var}(\log(L)) = \sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}.$$

Hence $\mathbb{P}(L = 0) = \mathbb{P}(\log(L) = -\infty) = 0$.

Step 2. Now we prove $Y_n \xrightarrow{d} L$. Since

$$\limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}\left[Y_n^2\right] < \infty,$$

condition (iv) implies that the sequence Y_n is tight. Prokhorov's theorem further implies that there is a subsequence $\{n_k\}_{k=1}^{\infty}$ such that Y_{n_k} converge in distribution to some random variable $L(\{n_k\})$. In what follows, we prove that the distribution of $L(\{n_k\})$ does not depend on the subsequence $\{n_k\}$. In particular, $L(\{n_k\}) \overset{d}{=} L$. To start with, note that since Y_{n_k} converges in distribution to $L(\{n_k\})$, for any further subsequence $\{n_{k_l}\}$ of $\{n_k\}$, $Y_{n_{k_l}}$ also converges in distribution to $L(\{n_k\})$.

Given any fixed $\epsilon > 0$ take m large enough such that

$$\exp\left\{\sum_{i=1}^{\infty}\frac{\mu_i^2}{\sigma_i^2}\right\} - \exp\left\{\sum_{i=1}^{m}\frac{\mu_i^2}{\sigma_i^2}\right\} < \epsilon.$$

For this fixed number m, consider the joint distribution of $(Y_{n_k}, W_{n_k,1}, \ldots, W_{n_k,m})$. This sequence of $m+1$ dimensional random vectors with respect to $\mathbb{P}_{n_k}$ is tight by condition (ii). So it has a further subsequence such that

$$(Y_{n_{k_l}}, W_{n_{k_l},1}, \ldots, W_{n_{k_l},m})|\mathbb{P}_{n_{k_l}} \xrightarrow{d} (L(\{n_k\}), Z_3 \ldots, Z_m)\,.$$

We are to show that we can define the random variables $L^{(m)}$ and $L(\{n_k\})$ in such a way that there exist suitable σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ such that $L^{(m)} \in \mathcal{F}_1$, $L(\{n_k\}) \in \mathcal{F}_2$, and $\mathbb{E}\left[L(\{n_k\}) \mid \mathcal{F}_1\right] = L^{(m)}$.

Since $\limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}\left[Y_n^2\right] < \infty$, the sequence $Y_{n_{k_l}}$ is uniformly integrable. This, together with condition (i), leads to

$$\mathbb{E}[L(\{n_k\})] = \lim_{l\to\infty} \mathbb{E}_{\mathbb{P}_{n_{k_l}}}[Y_{n_{k_l}}] = 1. \tag{3.7.1}$$

Now take any positive bounded continuous function $f : \mathbb{R}^m \to \mathbb{R}$. By Fatou's lemma

$$\liminf_{l\to\infty} \mathbb{E}_{\mathbb{P}_{n_{k_l}}}\left[f(W_{n_{k_l},1}, \ldots, W_{n_{k_l},m})Y_{n_{k_l}}\right] \geq \mathbb{E}\left[f\left(Z_1, \ldots, Z_m\right) L(\{n_k\})\right]. \tag{3.7.2}$$

However for any constant ξ, (3.7.1) implies $\xi = \xi\mathbb{E}_{\mathbb{P}_{n_{k_l}}}[Y_{n_{k_l}}] \to \xi\mathbb{E}[L(\{n_k\})] = \xi$. Observe that given any bounded continuous function f we can find ξ large enough so that $f + \xi$ is a bounded continuous function. So (3.7.2) is indeed implied by Fatou's lemma. Now

$$\begin{aligned}
&\liminf \mathbb{E}_{\mathbb{P}_{n_{k_l}}} \left[\left(f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m})+\xi\right) Y_{n_{k_l}}\right] \\
&= \liminf \mathbb{E}_{\mathbb{P}_{n_{k_l}}} \left[f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m}) Y_{n_{k_l}}\right] + \xi \\
&\geq \mathbb{E}\left[\left(f(Z_1,\ldots,Z_m)+\xi\right) L(\{n_k\})\right]
\end{aligned} \tag{3.7.3}$$

So (3.7.2) holds for any bounded continuous function f. On the other hand, replacing f by $-f$ we have

$$\lim_{l\to\infty} \mathbb{E}_{\mathbb{P}_{n_{k_l}}} \left[f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m}) Y_{n_{k_l}}\right] = \mathbb{E}\left[f(Z_1,\ldots,Z_m)L(\{n_k\})\right]. \tag{3.7.4}$$

Now condition (ii) leads to

$$\int f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m}) Y_{n_{k_l}} \mathrm{d}\mathbb{P}_{n_{k_l}} = \int f(W_{n_{k_l},1},\ldots,W_{n_{k_l},m}) \mathrm{d}\mathbb{Q}_{n_{k_l}} \to \int f(Z'_1,\ldots,Z'_m)\mathrm{d}Q.$$

Here Q is the measure induced by $(Z'_1,\ldots,Z'_m)$. In particular, one can take the measure Q such that $(Z_1,\ldots,Z_m)$ themselves are distributed as $(Z'_1,\ldots,Z'_m)$ under the measure Q. This is true since

$$\int f(Z'_1,\ldots,Z'_m)\mathrm{d}Q = \mathbb{E}\left[f(Z_1,\ldots,Z_m)L^{(m)}\right].$$

for any bounded continuous function f, and so $\int_A \mathrm{d}Q = \mathbb{E}[\mathbf{1}_A L^{(m)}]$ for any $A \in \sigma(Z_1,\ldots,Z_m)$. Now looking back into (3.7.4), we have for any $A \in \sigma(Z_1,\ldots,Z_m)$, $\mathbb{E}[\mathbf{1}_A L^{(m)}] = \mathbb{E}\left[\mathbf{1}_A L(\{n_k\})\right]$. Since by definition $L^{(m)}$ is $\sigma(Z_1,\ldots,Z_m)$ measurable, we have

$$L^{(m)} = \mathbb{E}\left[L(\{n_k\}) \mid \sigma(Z_1,\ldots,Z_m)\right].$$

From Fatou's lemma

$$\mathbb{E}[L(\{n_k\})^2] \leq \liminf_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[Y_n^2] = \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\}.$$

As a consequence, we have

$$0 \leq \mathbb{E}|L(\{n_k\}) - L^{(m)}|^2 = \mathbb{E}[L(\{n_k\})^2] - \mathbb{E}[L^{(m)2}] < \epsilon.$$

So $L_2(F^{L^{(m)}}, F^{L(\{n_k\})}) < \sqrt{\epsilon}$. Here $F^{L^{(m)}}$ and $F^{L(\{n_k\})}$ denote the distribution functions corresponding to $L^{(m)}$ and $L(\{n_k\})$ respectively. As a consequence, $W_2(F^{L^{(m)}}, F^{L(\{n_k\})}) \to 0$ as $m \to \infty$. Hence by Proposition 11, $L^{(m)} \xrightarrow{d} L(\{n_k\})$. On the other hand, we have already proved $L^{(m)}$ converges to L in L^2. So $L(\{n_k\}) \stackrel{d}{=} L$.

Proof of (3.2.3) Consider any fixed pair of $(\epsilon, \delta) \in (0,1) \times (0,1)$. First observe that the sequence $\log(Y_n)$ is tight from the proof of the previous part. For the given δ, there exists a fixed number $M < \infty$ such that $\mathbb{P}_n\left(-M \leq \log(Y_n) \leq M\right) \geq 1 - \frac{1}{100}\delta$ for all n, implying $\mathbb{P}_n\left(e^{-M} \leq Y_n \leq e^M\right) \geq 1 - \frac{1}{100}\delta$. Now consider $\tau \in (0, e^{-M})$. The function $\log(\cdot)$ is uniformly continuous on $[\tau, e^{M+1}]$. On this interval consider $\tilde{\epsilon}$ such that $|\log(x) - \log(y)| < \frac{\epsilon}{4}$ for all x, y on this interval with $|x - y| < \tilde{\epsilon}$. Let $\epsilon_1 = \min\{\tilde{\epsilon}, e^{-M} - \tau, e^{M+1} - e^M\}$ and pick a sufficiently large $K \in \mathbb{N}$ such that

$$\exp\left\{\sum_{k=1}^{\infty} \frac{\mu_k^2}{\sigma_k^2}\right\} - \exp\left\{\sum_{k=1}^{K} \frac{\mu_k^2}{\sigma_k^2}\right\} \leq \frac{\delta\epsilon_1^2}{100}. \tag{3.7.5}$$

From the proof of the previous part, we also know given any subsequence n_l there exists a further subsequence $\{n_{l_m}\}$ so that under $\mathbb{P}_n$,

$$\left(Y_{n_{l_m}}, \exp\left\{\sum_{k=1}^{K} \frac{2\mu_k W_{n_{l_m},k} - \mu_k^2}{2\sigma_k^2}\right\}\right) \xrightarrow{d} \left(L, \exp\left\{\sum_{k=1}^{K} \frac{2\mu_k Z_k - \mu_k^2}{2\sigma_k^2}\right\}\right)$$

and

$$\mathbb{E}\left[\left(L - \exp\left\{\sum_{k=1}^{K} \frac{2\mu_k Z_k - \mu_k^2}{2\sigma_k^2}\right\}\right)^2\right] \leq \frac{\delta\epsilon_1^2}{100}.$$

As a consequence,

$$\limsup_{n_{l_m}\to\infty} \mathbb{P}_{n_{l_m}}\left(\left|Y_{n_{l_m}} - \exp\left\{\sum_{k=1}^{K}\frac{2\mu_k W_{n_{l_m},k}-\mu_k^2}{2\sigma_k^2}\right\}\right| \geq \frac{\epsilon_1}{2}\right) \leq \mathbb{P}\left(\left|(L - \exp\left\{\sum_{k=1}^{K}\frac{2\mu_k Z_k-\mu_k^2}{2\sigma_k^2}\right\}\right| \geq \frac{\epsilon_1}{2}\right) \leq \frac{\delta}{25}. \tag{3.7.6}$$

As a consequence, for large values of n_{l_m},

$$\mathbb{P}_{n_{l_m}}\left(\left|Y_{n_{l_m}} - \exp\left\{\sum_{k=1}^{K}\frac{2\mu_k W_{n_{l_m},k}-\mu_k^2}{2\sigma_k^2}\right\}\right| \geq \frac{\epsilon_1}{2} \text{ and } Y_{n_{l_m}} \notin [e^{-M}, e^{M}]\right) \leq \frac{\delta}{25} + \frac{\delta}{100} < \frac{\delta}{2}. \tag{3.7.7}$$

Therefore, for large values of n_{l_m},

$$\mathbb{P}_{n_{l_m}}\left(\left|\log(Y_{n_{l_m}}) - \left\{\sum_{k=1}^{K}\frac{2\mu_k W_{n_{l_m},k}-\mu_k^2}{2\sigma_k^2}\right\}\right| \geq \frac{\epsilon}{2}\right) \leq \frac{\delta}{2}.$$

This completes the proof. □

3.8. Proof of Proposition 2

3.8.1. Preliminaries

The proof of the proposition is inspired by Anderson and Zeitouni (2006). The fundamental idea is to prove the asymptotic normality by using the method of moments and showing that moments of the limiting distributions satisfy Wick's formula. We first state the method of moments.

Lemma 7. *Let $Y_{n,1},\ldots,Y_{n,l}$ be a random vector of l dimension. Then $(Y_{n,1},\ldots,Y_{n,l}) \xrightarrow{d} (Z_1,\ldots,Z_l)$ if the following conditions are satisfied:*

(i) $\lim_{n\to\infty}\mathbb{E}[X_{n,1}\ldots X_{n,m}]$ *exists for any fixed m and $X_{n,i}\in\{Y_{n,1},\ldots,Y_{n,l}\}$ for $1\leq i\leq m$.*

(ii) (Carleman's Condition)Carleman (1926)

$$\sum_{h=1}^{\infty} \left(\lim_{n\to\infty} \mathbb{E}[X_{n,i}^{2h}] \right)^{-\frac{1}{2h}} = \infty \quad \forall\ 1 \le i \le l.$$

Further,

$$\lim_{n\to\infty} \mathbb{E}[X_{n,1} \dots X_{n,m}] = \mathbb{E}[X_1 \dots X_m].$$

Here $X_{n,i} \in \{Y_{n,1}, \dots, Y_{n,l}\}$ *for* $1 \le i \le m$ *and* X_i *is the in distribution limit of* $X_{n,i}$*. In particular, if* $X_{n,i} = X_{n,j}$ *for some* $i \ne j \in \{1, \dots, l\}$ *then* $X_i = X_j$*.*

Now we state Wick's formula for Gaussian random variables which was first proved by Isserlis (1918) and later on introduced by Wick (1950) in the physics literature.

Lemma 8. *Let* $(Y_1, \dots, Y_l)$ *be a multivariate mean* 0 *random vector of dimension* l *with covariance matrix* Σ *(possibly singular). Then* $(Y_1, \dots, Y_l)$ *is jointly Gaussian if and only if for any integer* m *and* $X_i \in \{Y_1, \dots, Y_l\}$ *for* $1 \le i \le m$

$$\mathbb{E}[X_1 \dots X_m] = \begin{cases} \sum_\eta \prod_{i=1}^{\frac{m}{2}} \mathbb{E}[X_{\eta(i,1)} X_{\eta(i,2)}] & \text{for } m \text{ even} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{3.8.1}$$

Here η *is a partition of* $\{1, \dots, m\}$ *into* $\frac{m}{2}$ *blocks such that each block contains exactly* 2 *elements and* $\eta(i,j)$ *denotes the* j*th element of the* i*th block of* η *for* $j = 1, 2$*.*

The proofs of the aforesaid lemmas are omitted. However, we note that the random variables $Y_1, \dots, Y_l$ may be the same. In particular, taking $Y_1 = \dots = Y_l$, Lemma 8 provides a description of the moments of multivariate Gaussian random variables.

3.8.2. Proof

In this part, we focus on the proof of Proposition 3 for testing problem (3.1.1). The result for (3.1.2) can be established analogously using the same strategy as that spelled out at the beginning of Section 3.5.

Additional notation and definition Given a set $\mathcal{S}$, an $\mathcal{S}$ letter s is simply an element $s \in \mathcal{S}$. With two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, a *bi-word* for $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined as an alternating ordered sequence of letters where the letters at odd positions come from $\mathcal{S}_1$ and the letters at even positions come from $\mathcal{S}_2$; The final letter is required to come from $\mathcal{S}_1$. We call the letters from $\mathcal{S}_1$ *type I* and those from $\mathcal{S}_2$ *type II*. Given any bi-word w, the ith type I letter is denoted by α_i and the ith type II letter by β_i. As a convention, we start the subscripts for letters in a bi-word with 0. Observe that any bi-word w starts from and ends with a type I letter and so the total number of letters in w is always odd. In particular, any bi-word w looks like $(\alpha_0, \beta_0, \alpha_1, \beta_1, \dots, \alpha_k)$. We use $l(w) = 2k + 1$ to denote the length of w. A bi-word is called *closed* if $\alpha_0 = \alpha_k$.

Throughout the proof, we take $\mathcal{S}_1 = \{1, \dots, p\}$ and $\mathcal{S}_2 = \{1, \dots, n\}$. The bipartite graph induced by a bi-word $w = (\alpha_0, \beta_0, \alpha_1, \beta_1, \dots, \alpha_k)$ is denoted by G_w. It is defined as follows. One treats the letters $(\alpha_0, \beta_0, \alpha_1, \beta_1, \dots, \alpha_k)$ as nodes and one puts an edge between α_i and β_j whenever $|i - j| = 1$. In this paper we shall focus on a special class of bi-words where the letters $(\alpha_0, \dots, \alpha_{k-1})$ and $(\beta_0, \dots, \beta_{k-1})$ are all distinct. In the subsequent part of the paper whenever we consider a word w we shall always assume that w belongs to the aforesaid restricted class of words. We call this class $\mathcal{W}_{2k}$. Observe that for any closed bi-word $w \in \mathcal{W}_{2k}$, G_w is a cycle of even length[2]. Two bi-words $w_1, w_2 \in \mathcal{W}_{2k}$ are called *paired* if the graphs G_{w_1} and G_{w_2} are the same. For a closed bi-word $w = (\alpha_0, \beta_0, \alpha_1, \beta_1, \dots, \alpha_k)$, its *mirror image* is $\tilde{w} = (\alpha_0, \beta_{k-1}, \alpha_{k-1}, \beta_{k-2}, \dots, \alpha_k)$. Furthermore, for a cyclic permutation τ of the set $\{0, 1, \dots, k-1\}$ and a closed bi-word w, we define $w^\tau := (\alpha_{\tau(0)}, \beta_{\tau(0)}, \alpha_{\tau(1)}, \beta_{\tau(1)}, \dots, \beta_{\tau(k-1)}, \alpha_{\tau(0)})$. If two closed bi-words $w_1, w_2 \in \mathcal{W}_2$ are paired, then there exists a cyclic permutation τ such that either $w_1^\tau = w_2$ or $\tilde{w}_1^\tau = w_2$.

Remark 8. These bi-words are not fundamentally different from the words defined in Anderson and Zeitouni (2006) and Anderson et al. (2010). In particular, they form a restricted class of words where the alphabet set is taken to be $\mathcal{S}_1 \cup \mathcal{S}_2$. Hence all the properties of the words can be derived with minimal modifications of the proofs in Anderson and Zeitouni

[2] Cycles of odd length in a bipartite graph do not exist.

(2006) and Anderson et al. (2010).

We call an ordered tuple of m words $(w_1, \ldots, w_m)$ a *sentence*. For any sentence $a = (w_1, \ldots w_m)$, $G_a = (V_a, E_a)$ is the graph with $V_a = \cup_{i=1}^m V_{w_i}$ and $E_a = \cup_{i=1}^m E_{w_i}$. A sentence a is called a *weak CLT sentence* if the following conditions hold:

1. Each word w_i is closed.

2. Each edge in G_a is traversed at least twice by the words w_i jointly.

3. For any word w_i, there exists another word w_j such that G_{w_i} shares an edge with G_{w_j}.

Although the definition of weak CLT sentences do not need $w_i \in \mathcal{W}_{2k_i}$ for some k_i, for the purpose of this paper we shall assume $w_i \in \mathcal{W}_{2k_i}$. The following lemma gives a bound on the number of weak CLT sentences. For any numbers b and c, $b \vee c = \max(b, c)$ and $b \wedge c = \min(b, c)$.

Lemma 9. *Let $\mathcal{A}_t = \mathcal{A}_t(l_1, \ldots, l_m)$ be the set of weak CLT sentences such that for each $a \in \mathcal{A}_t$, it consists of m words of lengths $l_1, \ldots, l_m$ respectively and $\#V_a = t$. Then*

$$\#\mathcal{A}_t \leq 2^{\sum_i l_i} \left(C_1 \sum_i l_i \right)^{C_2 m} \left(\sum_i l_i \right)^{3(\sum_i l_i - 2t)} n^t \left(\gamma \vee 1\right)^t . \tag{3.8.2}$$

Proof. The proof of this lemma is almost identical to the proof of Lemma 4.3 in Banerjee (2018). The only difference is in the possible choices of vertices of V_a. Here this choice will be $n^{t_1} p^{t_2}$ where $t_1 + t_2 = t$ and t_1 is the number of vertices which are from $\mathcal{S}_1$ and t_2 is the number of vertices which are from $\mathcal{S}_2$. It is easy to see in this case $n^{t_1} p^{t_2} = n^t \gamma^{t_2} \leq n^t \left(\gamma \vee 1\right)^t$. □

Remark 9. One might note that the proof of Lemma 4.3 in Banerjee (2018) does not require w_i to belong to a restrictive class of words. In fact the bound in Lemma 4.3 in Banerjee (2018) holds when w_i's are arbitrary closed words of proper length.

Proof of part (i) We complete the proof of this part in two steps. In the first step we calculate the asymptotic variances of $(B_{n,k_1}, \ldots, B_{n,k_l})$. The second step is dedicated towards proving the asymptotic normality and independence of $(B_{n,k_1}, \ldots, B_{n,k_l})$.

Step 1 (Calculation of variance). Under H_0, the case $k_1 = 1$ is simple as it is a sum of i.i.d. random variables and hence its variance calculation is omitted. One important thing to note is that the case $k = 1$ depends on $\mathrm{E}[X_{i,j}^4]$ (which is equal to 3 in the current case). This makes the asymptotic variance of $B_{n,1}$ equal to 2γ, which is not the case in general.

In what follows, we focus on the case when $k \geq 2$. Now we prove that $\mathrm{Var}(B_{n,k}) = (1 + o(1))2k\gamma^k$ for any finite k. Define for any bi-word $w = (\alpha_0, \beta_0, \alpha_1, \beta_1, \ldots, \alpha_k) \in \mathcal{W}_{2k}$,

$$X_w := X_{\alpha_0,\beta_0} X_{\alpha_1,\beta_0} X_{\alpha_1,\beta_1} X_{\alpha_2,\beta_1} \ldots X_{\alpha_{k-1},\beta_{k-1}} X_{\alpha_0,\beta_{k-1}} - \mathbf{1}_{k=1}. \tag{3.8.3}$$

Now observe that

$$\mathrm{Var}(B_{n,k}) = \left(\frac{1}{n}\right)^{2k} \mathbb{E}\left[\left(\sum_w X_w\right)^2\right] = \left(\frac{1}{n}\right)^{2k} \mathbb{E}\left[\sum_{w_1,w_2} X_{w_1} X_{w_2}\right]. \tag{3.8.4}$$

Since both X_{w_1} and X_{w_2} are products of independent mean 0 random variables that appears exactly once with X_{w_1} or X_{w_2}, $\mathbb{E}[X_{w_1} X_{w_2}] \neq 0$ if and only if all the edges in G_{w_1} are repeated in G_{w_2}. This happens only if w_1 and w_2 are paired. Now there are $(1 + o(1))n^k p^k$ choices for w_1 and for each w_1 there are exactly $2k$ w_2's such that w_1 and w_2 are paired (images of cyclic permutations of w_1 and of $\tilde{w}_1$). As a consequence,

$$\mathrm{Var}(B_{n,k}) = (1 + o(1))2k \frac{n^k p^k}{n^{2k}} = (1 + o(1))2k\gamma^k.$$

Step 2 (Proof of asymptotic normality). In order to complete this step, it suffices to prove the following two limits:

$$\lim_{n\to\infty} \mathbb{E}\left[\left(B_{n,k_1} - p\mathbf{1}_{k_1=1}\right) B_{n,k_2}\right] \to 0 \tag{3.8.5}$$

whenever $k_1 < k_2$ and there exist random variables $Z_1, \ldots, Z_m$ such that for any fixed m

$$\lim_{n\to\infty} \mathbb{E}[X_{n,1} \ldots X_{n,m}] \to \begin{cases} \sum_\eta \prod_{i=1}^{\frac{m}{2}} \mathbb{E}[Z_{\eta(i,1)} Z_{\eta(i,2)}] & \text{for } m \text{ even,} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{3.8.6}$$

where $X_{n,i} \in \left\{ \frac{B_{n,k_1} - p\mathbf{1}_{k_1=1}}{\sqrt{2k_1\gamma^{k_1}}}, \ldots, \frac{B_{n,k_l}}{\sqrt{2k_l\gamma^{k_l}}} \right\}$. To see this, observe that (3.8.6) simultaneously imply parts (i) and (ii) of Lemma 7. The implication of part (i) is obvious. For part (ii) one can take $X_{n,i}$'s to be all equal and from Wick's formula (Lemma 8) the limiting distribution of $X_{n,i}$'s are normal and it is well known that normal random variables satisfy Carleman's condition. In addition, (3.8.6) also implies that the limiting distribution of $\left(\frac{B_{n,k_1} - p\mathbf{1}_{k_1=1}}{\sqrt{2k_1\gamma^{k_1}}}, \ldots, \frac{B_{n,k_l}}{\sqrt{2k_l\gamma^{k_l}}} \right)$ is multivariate normal. Hence one gets the asymptotic independence by applying (3.8.5).

We first prove (3.8.5). Observe that

$$\mathbb{E}\left[\left(B_{n,k_1} - p\mathbf{1}_{k_1=1}\right) B_{n,k_2}\right] = \left(\frac{1}{n}\right)^{k_1+k_2} \mathbb{E}\left[\sum_{w_1,w_2} X_{w_1} X_{w_2}\right].$$

However, here $l(w_1) \neq l(w_2)$. So $\mathbb{E}\left[X_{w_1} X_{w_2}\right] = 0$. As a consequence, (3.8.5) holds.

Now we prove (3.8.6). Let $l_i - 1$ be the length of the bipartite cycle corresponding to $X_{n,i}$ (so that l_i is the length of the word corresponding to the bipartite cycle). Observe that $\frac{l_i-1}{2} \in \{k_1, \ldots, k_l\}$ for any i. At first we expand the left hand side of (3.8.6) as

$$\mathbb{E}[X_{n,1} \ldots X_{n,m}] = \left(\frac{1}{n}\right)^{\frac{1}{2}\sum_i (l_i - 1)} \sum_{w_1,\ldots,w_m} \mathbb{E}\left[X_{w_1} \ldots X_{w_m}\right]. \tag{3.8.7}$$

Here each of the graphs $G_{w_1}, \ldots, G_{w_m}$ are cycles of length $l_1 - 1, \ldots, l_m - 1$ respectively. We at first prove that one needs the sentence $a = [w_1, \ldots, w_m]$ to be a weak CLT sentence in order to have $\mathbb{E}\left[X_{w_1} \ldots X_{w_m}\right] \neq 0$. First observe that each edge in G_a needs to be traversed more than once. Otherwise let w_i be a word where a particular edge $e^* := \{\alpha_{e^*}, \beta_{e^*}\} \in G_{w_i}$ is traversed exactly once in G_a. In this case $l(w_i) > 3$. Now observe that X_{e^*} is independent of the rest of the random variables in $X_{w_1} \ldots X_{w_m}$. Thus $\mathrm{E}[X_{w_1} \ldots X_{w_m}] = 0$. To verify the other condition, suppose there is an i such that G_{w_i} does not share any edge with G_{w_j} for any $j \neq i$. Hence the random variables X_{w_i} and $\prod_{j \neq i} X_{w_j}$ are independent, and so $\mathbb{E}\left[X_{w_1} \cdots X_{w_m}\right] = \mathbb{E}[X_{w_i}]\,\mathbb{E}[\prod_{j \neq i} X_{w_j}] = 0$ from definition. Thus, $a = (w_1, \ldots, w_m)$ is a weak CLT sentence. Given any weak CLT sentence a, we introduce a partition $\eta(a)$ of $\{1, \ldots, m\}$ in the following way: If i and j are in same block of the partition $\eta(a)$, then G_{w_i} and G_{w_j} have at least one edge in common. Observe that each block in $\eta(a)$ has cardinality more than or equal to 2. As a consequence, we can further expand the right hand side of (3.8.7) as

$$\left(\frac{1}{n}\right)^{\frac{1}{2}\sum_i (l_i - 1)} \sum_{\eta} \sum_{\substack{w_1, \ldots, w_m: \\ \eta(w_1, \ldots, w_m) = \eta}} \mathbb{E}\left[X_{w_1} \ldots X_{w_m}\right], \tag{3.8.8}$$

where the number of blocks in $\eta \leq \lfloor \frac{m}{2} \rfloor$. We now show that only those η's such that the number of blocks in them are exactly $\frac{m}{2}$ contribute to a non-vanishing asymptotic mean. Note that this necessarily requires m to be even.

When $\eta(w_1, \ldots, w_m)$ have strictly less than $\lfloor \frac{m}{2} \rfloor$ blocks (including all cases of odd m and the case of even m when the number of blocks is strictly less than $\frac{m}{2}$), G_a has strictly less than $\lfloor \frac{m}{2} \rfloor$ connected components. From Lemma 4.10 of Anderson and Zeitouni (2006) it follows that in this case $\# V_a < \sum_{i=1}^{m} \frac{l_i - 1}{2}$. Applying Lemma 9 and noting that the a's are weak CLT sentences, we have

$$
\begin{aligned}
&\left(\frac{1}{n}\right)^{\frac{1}{2}\sum_i(l_i-1)} \sum_{a:\#V_a<\sum_{i=1}^{m}\frac{l_i-1}{2}} \mathbb{E}\left[X_{w_1}\dots X_{w_m}\right] \\
&\leq \left(\frac{1}{n}\right)^{\frac{1}{2}\sum_i(l_i-1)} 2^{\sum_i l_i} \sum_{t<\frac{1}{2}\sum_i(l_i-1)} \left(C_1\sum_i l_i\right)^{C_2 m}\left(\sum_i l_i\right)^{3(\sum_i l_i-2t)} n^t\,(\gamma\vee 1)^t\,\mathbb{E}\left[|X_{11}|^{\sum_i l_i}\right] \\
&\leq \mathbb{E}\left[|X_{11}|^{\sum_i l_i}\right] 2^{\sum_i l_i}\left(C_1\sum_i l_i\right)^{C_2 m}\left(\sum_i l_i\right)^{3m}(\gamma\vee 1)^{\frac{1}{2}\sum_i(l_i-1)} \\
&\qquad \sum_{t<\frac{1}{2}\sum_i(l_i-1)}\left(\frac{(\sum_i l_i)^3}{\sqrt{n}}\right)^{\sum_i(l_i-1)-2t} \\
&\leq 2^{\sum_i l_i}\left(C_3\sum_i l_i\right)^{C_4\sum_i l_i}\left(C_1\sum_i l_i\right)^{C_2 m}\left(\sum_i l_i\right)^{3m}(\gamma\vee 1)^{\frac{1}{2}\sum_i(l_i-1)}\,O\left(\frac{(\sum_i l_i)^3}{\sqrt{n}}\right)
\end{aligned}
\tag{3.8.9}
$$

the last expression seems redundant... the previous version was correct anyway and more concise. Here we have also used the fact for any standard Gaussian random variable $\mathbb{E}[|X|^l] \leq (C_3 l)^{C_4 l}$. Observe that in the final expression of (3.8.9) the dominant term is $(C_3\sum_{i=1}^m l_i)^{C_4\sum_{i=1}^m l_i}$ and for $l_1,\dots,l_m = o(\sqrt{\log n})$, $(C_3\sum_{i=1}^m l_i)^{C_4\sum_{i=1}^m l_i}/n^\alpha \to 0$ whenever $\alpha > 0$ and m is finite[3].

Now the only remaining partitions are pair partitions which have exactly $\frac{m}{2}$ many blocks (and so naturally m is even). We now fix a partition η of this kind. Let for any $i \in \{1,\dots,\frac{m}{2}\}$, $\eta(i,1) < \eta(i,2)$ be the elements in the ith block. Observe now that fixing a pair partition η and $(w_1,\dots,w_m)$ such that $\eta(w_1,\dots,w_m) = \eta$, the random variables $X_{w_{\eta(i_1,j)}}$ and $X_{w_{\eta(i_2,j)}}$ are independent when ever $i_1 \neq i_2$ for any $j \in \{1,2\}$. As a consequence, we

[3] In fact the term $\mathbb{E}[|X_{11}|^{\sum_i l_i}]$ is not optimal. One can prove the CLT under the null upto $o(\log n)$ order by the arguments similar to (2.1.32) in Anderson et al. Anderson et al. (2010). However for our purpose this suffices.

now can rewrite (3.8.8) as

$$\begin{aligned}&\left(\frac{1}{n}\right)^{\frac{1}{2}\sum_i(l_i-1)}\sum_{\eta}\sum_{\substack{w_1,\ldots,w_m:\\ \eta(w_1,\ldots,w_m)=\eta}}\mathbb{E}\left[X_{w_1}\ldots X_{w_m}\right]\\ &=o(1)+\left(\frac{1}{n}\right)^{\frac{1}{2}\sum_i(l_i-1)}\sum_{\eta\text{ pair partition}}\sum_{\substack{w_1,\ldots,w_m:\\ \eta(w_1,\ldots,w_m)=\eta}}\prod_{i=1}^{\frac{m}{2}}\mathbb{E}\left[X_{w_{\eta(i,1)}}X_{w_{\eta(i,2)}}\right]\end{aligned} \tag{3.8.10}$$

Now observe that whenever $\prod_{i=1}^{\frac{m}{2}}\mathbb{E}[X_{w_{\eta(i,1)}}X_{w_{\eta(i,2)}}]\neq 0$, we have $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$ are paired. When $l(w_{\eta(i,1)})=l(w_{\eta(i,2)})\neq 3$, there are $(1+o(1))(l_{\eta(i,1)}-1)(n\sqrt{\gamma})^{l_{\eta(i,1)}-1}$ many such choices of $(w_{\eta(i,1)},w_{\eta(i,2)})$ for every i. Here $l_{\eta(i,1)}-1$ equals the common length of the cycles induced by $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$. In this case $\mathbb{E}[X_{w_{\eta(i,1)}}X_{w_{\eta(i,2)}}]=1$. On the other hand, when $l(w_{\eta(i,1)})=3$, there are $(1+o(1))n^{l_{\eta(i,1)}-1}\gamma$ many such choices of $(w_{\eta(i,1)},w_{\eta(i,2)})$ for every i and in this case $\mathbb{E}[X_{w_{\eta(i,1)}}X_{w_{\eta(i,2)}}]=2$. Hence, we get the following further reduction of the right side of (3.8.10):

$$\begin{aligned}&o(1)+(1+o(1))\left(\frac{1}{n}\right)^{\frac{1}{2}\sum_i(l_i-1)}\sum_{\eta\text{ pair partition}}\prod_{i=1}^{\frac{m}{2}}(l_{\eta(i,1)}-1)\mathbf{1}_{l_{\eta(i,1)}=l_{\eta(i,2)}}(n\sqrt{\gamma})^{l_{\eta(i,1)}-1}\\ &=o(1)+(1+o(1))\sum_{\eta\text{ pair partition}}\prod_{i=1}^{\frac{m}{2}}(l_{\eta(i,1)}-1)\gamma^{\frac{1}{2}(l_{\eta(i,1)}-1)}\mathbf{1}_{l_{\eta(i,1)}=l_{\eta(i,2)}}.\end{aligned} \tag{3.8.11}$$

Recalling that $l_i=2k_i+1$ we complete the proof. □

Proof of part (ii) We at first look at the case when $k=1$. This is an exceptional case and needs to be handled differently. Then we deal with the general case of $k\geq 2$.

Analysis of $B_{n,1}$. Recall that $B_{n,1}=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}X_{i,j}^2$. We have

$$B_{n,1}\,|\Theta,U=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}(Z_{i,j}+M_{i,j})^2 \tag{3.8.12}$$

where for any (i, j), $M_{i,j}$ is defined as in (3.5.3) and $Z_{i,j} \stackrel{iid}{\sim} N(0, 1)$. Observe that in this case one can apply the Lindeberg–Feller central limit theorem. So it suffices to calculate the limiting mean and variance of $B_{n,1} \,|\Theta, U$. Now

$$\mathbb{E}\left[X_{i,j}^2 \,|\Theta, U\right] = 1 + M_{i,j}^2, \tag{3.8.13}$$

and

$$\begin{aligned} \text{Var}\left(X_{i,j}^2|\Theta, U\right) &= \text{Var}\left(Z_{i,j}^2 + 2Z_{i,j}M_{i,j}|\Theta, U\right) \\ &= \text{Var}\left(Z_{i,j}^2\right) + 4\text{Var}[Z_{i,j}]M_{i,j}^2 \\ &= 2 + 4M_{i,j}^2. \end{aligned} \tag{3.8.14}$$

So it is enough to prove

$$\frac{1}{n}\sum_{i,j} M_{i,j}^2 \xrightarrow{p} \sum_{l_1,l_2} \Sigma_\Theta(l_1, l_2)\Sigma_U(l_1, l_2). \tag{3.8.15}$$

As a consequence,

$$\text{Var}\left(B_{n,1}\right) = \frac{1}{n^2}\left(2np + \sum_{i,j} 4M_{i,j}^2\right) \to 2\gamma.$$

To this end, note that

$$\begin{aligned} \frac{1}{n}\sum_{i,j} M_{i,j}^2 &= \frac{1}{n}\left[\sum_{i,j}\sum_{l,l'} \frac{1}{p}\Theta_{i,l}\Theta_{i,l'}U_{j,l}U_{j,l'}\right] \\ &= \sum_{l=1}^{\kappa}\sum_{l'=1}^{\kappa}\left(\frac{1}{n}\sum_{i=1}^{n}\Theta_{i,l}\Theta_{i,l'}\right)\left(\frac{1}{p}\sum_{j=1}^{p} U_{j,l}U_{j,l'}\right). \end{aligned} \tag{3.8.16}$$

The weak law of large numbers then gives

$$\frac{1}{n}\sum_{i=1}^{n}\Theta_{i,l}\Theta_{i,l'} \xrightarrow{p} \Sigma_\Theta(l, l') \qquad \text{and} \qquad \frac{1}{p}\sum_{j=1}^{p} U_{j,l}U_{j,l'} \xrightarrow{p} \Sigma_U(l, l').$$

Since κ is fixed, we obtain (3.8.15).

Analysis of $B_{n,k}$ with $k \geq 2$. We first write

$$\begin{aligned} B_{n,k} &= \frac{1}{n^k} \sum_{i_0,j_0,\ldots,i_{k-1},j_{k-1}} X_{i_0,j_0} \ldots X_{i_0,j_{k-1}} \\ &= \frac{1}{n^k} \sum_{i_0,j_0,\ldots,i_{k-1},j_{k-1}} (Z_{i_0,j_0} + M_{i_0,j_0}) \ldots \left(Z_{i_0,j_{k-1}} + M_{i_0,j_{k-1}}\right) \\ &= \frac{1}{n^k} \sum_{i_0,j_0,\ldots,i_{k-1},j_{k-1}} Z_{i_0,j_0} \ldots Z_{i_0,j_{k-1}} + \mu_{n,k} + V_{n,k}, \end{aligned} \tag{3.8.17}$$

where

$$\mu_{n,k} := \frac{1}{n^k} \sum_{i_0,j_0,\ldots,i_{k-1},j_{k-1}} M_{i_0,j_0} \cdots M_{i_0,j_{k-1}}, \tag{3.8.18}$$

and $V_{n,k}$ collects all the terms involving cross-products.

The proof of the asymptotic normality of $\frac{1}{n^k} \sum_{i_0,j_0,\ldots,i_{k-1},j_{k-1}} Z_{i_0,j_0} \ldots Z_{i_0,j_{k-1}}$ is the same as the proof we have just finished for the null distribution. We shall prove later that μ_k satisfies (3.4.3). Now we focus on $V_{n,k}$. Observe that $\mathbb{E}\left[V_{n,k} \,|\, \Theta, U\right] = 0$ and hence $\mathbb{E}[V_{n,k}] = 0$. So our goal is to prove $\mathbb{E}[V_{n,k}^2] \to 0$ which implies $V_{n,k} \xrightarrow{p} 0$.

Note that $V_{n,k} = \sum_w V_{n,k,w}$ where the summation is over all closed bi-words of length $2k+1$. Fix such a bi-word w and let $\emptyset \subsetneq E_f \subsetneq E_w$ be a subset. Then

$$V_{n,k,w} = \frac{1}{n^k} \sum_{\emptyset \subsetneq E_f \subsetneq E_w} \mu(E_f, w) \prod_{e \in E_w \setminus E_f} Z_e.$$

Here

$$\mu(E_f, w) = \prod_{e \in E_f} M_{\alpha_e, \beta_e}.$$

where for any edge e, α_e and β_e denote its two end points which belong to $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. Now

$$\mathbb{E}\left[V_{n,k}^2 \,|\Theta, U\right] = \sum_{w_1,w_2} \mathbb{E}\left[V_{n,k,w_1} V_{n,k,w_2} \,|\Theta, U\right]. \tag{3.8.19}$$

We now give an upper bound to $\mathbb{E}\left[V_{n,k,w_1} V_{n,k,w_2}\right]$. At first fix any word w_1 and the set

$\emptyset \subsetneq E_f \subsetneq E_{w_1}$ and consider all the words w_2 such that $E_{w_1} \cap E_{w_2} = E_{w_1} \backslash E_f$. As every edge in G_{w_1} and G_{w_2} appear exactly once within G_{w_1} and G_{w_2},

$$\begin{aligned} &\mathbb{E}[V_{n,k,w_1} V_{n,k,w_2} | \Theta, U] \\ &= \sum_{E_{w_1} \backslash E' \subset E_{w_1} \backslash E_f} \left(\frac{1}{n}\right)^{2k} \left[\mu(E', w_1)\mu(E_{w_2} \backslash (E_{w_1} \backslash E'), w_2)\right] \mathbb{E} \prod_{e \in E_w \backslash E'} (Z_e)^2 \\ &= \sum_{E_{w_1} \backslash E' \subset E_{w_1} \backslash E_f} \left(\frac{1}{n}\right)^{2k} \left[\mu(E', w_1)\mu(E_{w_2} \backslash (E_{w_1} \backslash E'), w_2)\right]. \end{aligned} \tag{3.8.20}$$

Now it is enough to prove

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{1}{n}\right)^{2k} \sum_{w_1} \sum_{\emptyset \subsetneq E_f \subsetneq E_w} \sum_{E_f \subset E'} \sum_{\{w_2 | E_{w_1} \cap E_{w_2} = E_{w_1} \backslash E_f\}} \left[\mu(E', w_1)\mu(E_{w_2} \backslash (E_{w_1} \backslash E'), w_2)\right]\right] \\ &\leq \left(\frac{1}{n}\right)^{2k} \sum_{w_1} \sum_{\emptyset \subsetneq E_f \subsetneq E_w} \sum_{E_f \subset E'} \sum_{\{w_2 | E_{w_1} \cap E_{w_2} = E_{w_1} \backslash E_f\}} \mathbb{E}\left|\mu(E', w_1)\mu(E_{w_2} \backslash (E_{w_1} \backslash E'), w_2)\right| \to 0. \end{aligned} \tag{3.8.21}$$

Now observe that for any w in consideration and any subset E of E_w,

$$|\mu(E, w)| = \left(\frac{1}{p}\right)^{\frac{\#E}{2}} \prod_{e \in E} \left|\sum_{l=1}^{\kappa} \Theta_{\alpha_e, l} U_{\beta_e, l}\right|.$$

Hence we have for any $E \subset E_{w_1}$ and $\bar{E} \subset E_{w_2}$ such that $\#E = \#\bar{E}$,

$$\begin{aligned} &\mathbb{E}\left|\mu(E, w_1)\mu(\bar{E}, w_2)\right| \\ &\leq \left(\frac{1}{p}\right)^{\#E} \prod_{e \in E} \mathbb{E}\left[\left|\sum_{l=1}^{\kappa} \Theta_{\alpha_e, l} U_{\beta_e, l}\right|^{2\#E}\right]^{\frac{1}{2\#E}} \prod_{\bar{e} \in \bar{E}} \mathbb{E}\left[\left|\sum_{l=1}^{\kappa} \Theta_{\alpha_{\bar{e}}, l} U_{\beta_{\bar{e}}, l}\right|^{2\#E}\right]^{\frac{1}{2\#E}} \\ &\leq \left(\frac{1}{p}\right)^{\#E} (C_5 \#E)^{C_6 \#E} \leq \left(\frac{1}{p}\right)^{\#E} (C_7 k)^{C_8 k}. \end{aligned} \tag{3.8.22}$$

The last step follows from the fact no matter what the value of e is, $\sum_{l=1}^{\kappa} \Theta_{\alpha_e, l} U_{\beta_e, l}$ is sub-exponential with parameter C for some constant C that depends on κ, $\widetilde{\Sigma}_\Theta$ and $\widetilde{\Sigma}_U$.

Plugging the estimate obtained in (3.8.22) in (3.8.21), we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{n}\right)^{2k}\sum_{w_1}\sum_{\emptyset\subsetneq E_f\subsetneq E_w}\sum_{E_f\subset E'}\sum_{\{w_2|E_{w_1}\cap E_{w_2}=E_{w_1}\backslash E_f\}}\left[\mu(E',w_1)\mu(E_{w_2}\backslash(E_{w_1}\backslash E'),w_2)\right]\right]\\
&\leq\left(\frac{1}{n}\right)^{2k}\sum_{w_1}\sum_{\emptyset\subsetneq E_f\subsetneq E_w}\sum_{E_f\subset E'}\sum_{\{w_2|E_{w_1}\cap E_{w_2}=E_{w_1}\backslash E_f\}}\left(\frac{1}{p}\right)^{\#E'}(C_7k)^{C_8k}\\
&\leq\left(\frac{1}{n}\right)^{2k}(C_7k)^{C_8k}\sum_{w_1}\sum_{\emptyset\subsetneq E_f\subsetneq E_w}\left(\frac{1}{p}\right)^{\#E_f}\sum_{E_f\subset E'}\sum_{\{w_2|E_{w_1}\cap E_{w_2}=E_{w_1}\backslash E_f\}}1.
\end{aligned}\tag{3.8.23}$$

Observe that the graph corresponding to the edges $E_{w_1}\backslash E_f$ is a disjoint collection of line segments. Let the number of such line segments be ζ. Obviously $\zeta\leq\#(E_{w_1}\backslash E_f)$. The number of ways these ζ components can be placed in w_2 is bounded by $(2k)^{\zeta}\leq(2k)^{\#(E_{w_1}\backslash E_f)}\leq(2k)^{2k}$ and all other nodes in w_2 can be chosen freely. So there are at most $(1+o(1))\,[(\gamma\vee 1)\,n]^{2k-\#V_{E_{w_1}\backslash E_f}}(2k)^{2k}$ choices of such w_2. Here $V_{E_w\backslash E_f}$ is the set of vertices of the graph corresponding to G_w with all edges in E_f removed, i.e., $E_w\backslash E_f$. Observe that, whenever $2k=\mathrm{E}_w>\#E_f>0$, $E_w\backslash E_f$ is a forest and so $\#V_{E_w\backslash E_f}\geq\#(E_w\backslash E_f)+1$ which is equivalent to

$$2k-\#V_{E_w\backslash E_f}\leq\#E_f-1.$$

Also observe that there are no more than 2^{2k} many choices of E_f's and for each E_f there are no more than 2^{2k} many choices for E''s. Combining all these, we have the rightmost side of (3.8.23) is bounded by

$$\begin{aligned}
&\left(\frac{1}{n}\right)^{2k}(C_7k)^{C_8k}\sum_{w_1}\sum_{\emptyset\subsetneq E_f\subsetneq E_w}\left(\frac{1}{p}\right)^{\#E_f}(2)^{2k}\times(2k)^{2k}\left[(\gamma\vee 1)\,n\right]^{\#E_f-1}\\
&\leq\frac{1}{p}(C_7k)^{C_8k}(2k)^{2k}2^{4k}\left[\frac{\gamma\vee 1}{\gamma}\right]^{2k}\to 0.
\end{aligned}\tag{3.8.24}$$

Now our final task is to prove $\mu_{n,k}\xrightarrow{p}\mu_k$ defined in (3.4.3). First we expand $\mu_{n,k}$ in (3.8.18) as

$$\mu_{n,k} = \frac{1}{n^k}\frac{1}{p^k} \sum_{i_0,j_0,\ldots,i_{k-1},j_{k-1}} \sum_{l_1,\ldots,l_{2k}} \Theta_{i_0,l_1} U_{j_0,l_1} \ldots \Theta_{i_0,l_{2k}} U_{j_{k-1},l_{2k}}$$

$$= \sum_{l_1,\ldots,l_{2k}} \left(\frac{1}{n^k} \left[\sum_{i_0,\ldots,i_{k-1}} \Theta_{i_0,l_1}\Theta_{i_0,l_{2k}}\Theta_{i_1,l_2}\Theta_{i_1,l_3}\ldots\Theta_{i_{k-1},l_{2k-2}}\Theta_{i_{k-1},l_{2k-1}} \right] \times \frac{1}{p^k} \left[\sum_{j_0,\ldots,j_{k-1}} U_{j_0,l_1}U_{j_0,l_2}U_{j_1,l_3}U_{j_1,l_4}\ldots U_{j_{k-1},l_{2k-1}}U_{j_{k-1},l_{2k}} \right] \right). \tag{3.8.25}$$

Now fix the values of $l_1,\ldots,l_{2k}$ and for this value of the group assignment we have

$$\mathbb{E}\left[\frac{1}{n^k} \sum_{i_0,\ldots,i_{k-1}} \Theta_{i_0,l_1}\Theta_{i_0,l_{2k}}\Theta_{i_1,l_2}\Theta_{i_1,l_3}\ldots\Theta_{i_{k-1},l_{2k-2}}\Theta_{i_{k-1},l_{2k-1}} \right] = m^{\Theta}_{l_1,\ldots,l_{2k}} = (1+o(1))\Sigma_\Theta(l_1,l_{2k})\ldots\Sigma_\Theta(l_{2k-2},l_{2k-1}).$$

Now

$$\mathrm{Var}\left(\frac{1}{n^k} \sum_{i_0,\ldots,i_{k-1}} \Theta_{i_0,l_1}\Theta_{i_0,l_{2k}}\Theta_{i_1,l_2}\Theta_{i_1,l_3}\ldots\Theta_{i_{k-1},l_{2k-2}}\Theta_{i_{k-1},l_{2k-1}} \right)$$
$$= \frac{1}{n^{2k}} \sum_{i_0^{(1)},\ldots,i_{k-1}^{(1)}} \sum_{i_0^{(2)},\ldots,i_{k-1}^{(2)}} \mathbb{E}\left[\left(\Theta_{i_0^{(1)},l_1}\Theta_{i_0^{(1)},l_{2k}}\ldots\Theta_{i_{k-1}^{(1)},l_{2k-2}}\Theta_{i_{k-1}^{(1)},l_{2k-1}} - m^{\Theta}_{l_1,\ldots,l_{2k}} \right) \times \left(\Theta_{i_0^{(1)},l_2}\Theta_{i_0^{(1)},l_{2k}}\ldots\Theta_{i_{k-1}^{(2)},l_{2k-2}}\Theta_{i_{k-1}^{(2)},l_{2k-1}} - m^{\Theta}_{l_1,\ldots,l_{2k}} \right) \right]. \tag{3.8.26}$$

However, if the indices $(i_0^{(1)},\ldots,i_{k-1}^{(1)})$ and $(i_0^{(2)},\ldots,i_{k-1}^{(2)})$ are disjoint,

$$\mathbb{E}\left[\left(\Theta_{i_0^{(1)},l_1}\Theta_{i_0^{(1)},l_{2k}}\ldots\Theta_{i_{k-1}^{(1)},l_{2k-2}}\Theta_{i_{k-1}^{(1)},l_{2k-1}} - m^{\Theta}_{l_1,\ldots,l_{2k}} \right) \times \left(\Theta_{i_0^{(1)},l_2}\Theta_{i_0^{(1)},l_{2k}}\ldots\Theta_{i_{k-1}^{(2)},l_{2k-2}}\Theta_{i_{k-1}^{(2)},l_{2k-1}} - m^{\Theta}_{l_1,\ldots,l_{2k}} \right) \right] = 0.$$

Now consider the indices

$$\mathcal{A} := \left\{ (i_0^{(1)}, \ldots, i_{k-1}^{(1)}), (i_0^{(2)}, \ldots, i_{k-1}^{(2)}) \;\middle|\; \#\big(\{i_0^{(1)}, \ldots, i_{k-1}^{(1)}\} \cap \{i_0^{(2)}, \ldots, i_{k-1}^{(2)}\}\big) \geq 1 \right\}.$$

It is easy to see $\#\mathcal{A} \leq (c_1 k)^{c_2 k}\, n^{2k-1}$. Further from sub-Gaussianity and Hölder's inequality we also have

$$\mathbb{E}\left| \left(\Theta_{i_0^{(1)}, l_1} \Theta_{i_0^{(1)}, l_{2k}} \ldots \Theta_{i_{k-1}^{(1)}, l_{2k-2}} \Theta_{i_{k-1}^{(1)}, l_{2k-1}} - m^{\Theta}_{l_1, \ldots, l_{2k}} \right) \times \right.$$
$$\left. \left(\Theta_{i_0^{(1)}, l_2} \Theta_{i_0^{(1)}, l_{2k}} \ldots \Theta_{i_{k-1}^{(2)}, l_{2k-2}} \Theta_{i_{k-1}^{(2)}, l_{2k-1}} - m^{\Theta}_{l_1, \ldots, l_{2k}} \right) \right| = (c_3 k)^{c_4 k}$$

uniformly over the indices. This gives us the final expression of (3.8.26) to be bounded by $\frac{(c_1 c_3 k)^{(c_2 + c_4)k}}{n} \to 0$. The proof for

$$\frac{1}{p^k} \sum_{j_0, \ldots, j_{k-1}} U_{j_0, l_1} U_{j_0, l_2} U_{j_1, l_3} U_{j_1, l_4} \ldots U_{j_{k-1}, l_{2k-1}} U_{j_{k-1}, l_{2k}} \xrightarrow{p} \Sigma_U(l_1, l_2) \Sigma_U(l_3, l_4) \ldots \Sigma_U(l_{2k-1}, l_{2k})$$

is analogous and so we omit the details. □

CHAPTER 4 : Optimal hypothesis testing for planted partition model with growing degrees

4.1. Introduction

Stochastic block model (SBM) Holland et al. (1983) is an active domain of modern research in statistics, computer science and many other related fields. A stochastic block model for random graphs encodes a community structure where a pair of nodes from the same community are expected to be connected in a different manner from those from different communities. This model, together with the related community detection problem, has drawn substantial attentions in statistics and machine learning. Throughout the paper, let $\mathcal{G}_1(n, p_n)$ denote the Erdős-Renyi with n nodes in which the edges are i.i.d. Bernoulli random variables with success probability p_n. For any integer $\kappa \geq 2$, let $\mathcal{G}_\kappa(n, p_n, q_n)$ denote the symmetric stochastic block model with κ different blocks where the label σ_u of any node u is assigned independently and uniformly at random from the set $\{1, 2, \ldots, \kappa\}$. The edges are independent Bernoulli random variables, and two nodes are connected with probability p_n if they share the same label and q_n otherwise.

A fundamental question related to stochastic block models is community detection where one aims to recover the partition of nodes into communities based on one instance of the random graph. Depending on the signal-to-noise ratio, there are three different regimes for recovery, namely partial recovery, almost exact recovery and exact recovery. In the asymptotic regime of bounded degrees (i.e. np_n and nq_n remain constants as $n \to \infty$), the seminal papers by Mossel et al. Mossel et al. (2015); Mossel et al. (2013) and Massoulié (2013) established sharp threshold for $\mathcal{G}_2(n, p_n, q_n)$ on when it is possible and impossible to achieve a partial recovery of community labels that is strictly better than random guessing, which confirmed the conjecture in Decelle et al. (2011). See Abbe and Sandon (2015) for an extension to multiple blocks and Banerjee (2018) for an extension to the regime of growing degrees (i.e. $np_n, nq_n \to \infty$ as $n \to \infty$). In the regime of growing degrees, Mossel et al. (2016) established the necessary and sufficient condition for achieving almost exact recovery in $\mathcal{G}_2(n, p_n, q_n)$, i.e. when only a vanishing proportion of node labels are not recovered correctly. See also Abbe and Sandon (2015); Gao et al. (2015); Zhang et al. (2016); Yun and Proutiere (2016); Gao et al. (2016) for results on more general SBMs. Furthermore, Abbe et al. (2014) and Mossel et al. (2016) established the necessary and sufficient condition for achieving exact recovery of labels for $\mathcal{G}_2(n, p_n, q_n)$ which was later extended by Hajek et al. (2016a,b); Abbe and Sandon (2015); Jog and Loh (2015); Yun and Proutiere (2016); Gao et al. (2016) to more general cases. See Abbe (2017) for a survey of some recent results.

In addition to the literature on information-theoretic limits, many community detection algorithms have been proposed, including but not limited to spectral clustering and likelihood based clustering. An almost universal assumption of these algorithms is the knowledge of the number of blocks κ, which usually is unknown in practice. For data-driven choice of κ, researchers have proposed different methods. One popular way is information criterion based model selection. See, e.g., Daudin et al. (2008); Latouche et al. (2012); Peixoto (2013); Wang et al. (2017); Saldana et al. (2017). In addition, several block-wise cross-validation methods have been proposed and studied. See, e.g., Chen and Lei (2014); Dabbs and Junker (2016). Furthermore, Bickel and Sarkar (2016) proposed to recursively apply the largest eigenvalue test for partitioning the nodes and for determining κ. The proposal was based on the GOE Tracy–Widom limit Tracy and Widom (1996) of the largest eigenvalue distribution for adjacency matrices of Erdős-Renyi when the average degree grows linearly with n. Lei (2016) extended it to a procedure based on sequential largest eigenvalue tests in the regime where exact recovery can be achieved. See also Le and Levina (2015) for another spectral method for choosing κ.

Let the observed adjacency matrix be $A \in \{0, 1\}^{n\times n}$. The major focus of the present paper is to test the following hypotheses:

$$H_0 : A \sim \mathcal{G}_1\left(n, \frac{p_n + q_n}{2}\right) \quad \text{vs.} \quad H_1 : A \sim \mathcal{G}_2(n, p_n, q_n) \tag{4.1.1}$$

when the average degree of the random graph grows to infinity with the graph size. The parameters in the hypotheses are so chosen that the expected numbers of edges match under null and alternative. Let $a_n = np_n$ and $b_n = nq_n$. Our primary interest lies in the cases where the signal-to-noise ratio

$$c := \frac{(a_n - b_n)^2}{a_n + b_n} \tag{4.1.2}$$

is a constant, and we call any such alternative a *local* one. For such cases, one has growing average degree if and only if $np_n \to \infty$. In what follows, we denote the null and alternative hypotheses in (4.1.1) by $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ respectively. This testing problem is not only fundamental to inference for SBMs but is also the foundation of any test based method for choosing κ.

For (4.1.1), Mossel et al. (2015) (resp. Banerjee (2018)) proved that when $a_n \equiv a$ and $b_n \equiv b$ are fixed constants (resp. when $a_n \to \infty$ and $a_n/n = p_n \to p \in [0, 1)$), if $c < 2$ (resp. $c < 2(1 - p)$), then the measures $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ are *mutually contiguous* Le Cam (2012), i.e. for a sequence of events E_n, $\mathbb{P}_{0,n}(E_n) \to 0$ if and only if $\mathbb{P}_{1,n}(E_n) \to 0$. On the other hand, if $c > 2$ (resp. $c > 2(1 - p)$), then $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ are *asymptotically singular*. These results imply that whenever $c < 2$ (resp. $c < 2(1 - p)$), it is impossible to find a consistent test for (4.1.1).

In the respective asymptotic regimes, Mossel et al. (2015) and Banerjee (2018) further obtained explicit descriptions of the asymptotic log-likelihood ratio within the contiguous regime. Let $L_n := \frac{\mathrm{d}\mathbb{P}_{1,n}}{\mathrm{d}\mathbb{P}_{0,n}}$ be the likelihood ratio. In the growing degree asymptotic regime, Banerjee (2018, Proposition 3.4) showed that if $c < 2(1 - p)$ where $p = \lim_{n\to\infty} p_n$, then

$$\log(L_n) \,|\, \mathbb{P}_{0,n} \xrightarrow{d} \sum_{i=3}^{\infty} \frac{2t^i Z_i - t^{2i}}{4i}, \quad \log(L_n) \,|\, \mathbb{P}_{1,n} \xrightarrow{d} \sum_{i=3}^{\infty} \frac{2t^i Z_i + t^{2i}}{4i}, \tag{4.1.3}$$

where

$$t = \sqrt{\frac{c}{2(1-p)}} \quad \text{and} \quad Z_i \overset{\text{ind}}{\sim} N(0, 2i). \tag{4.1.4}$$

Each random variable Z_i comes from the weak limit of the signed cycle of length i. See (Banerjee, 2018, Definition 4.1) and Eq. (4.2.2) below for the exact definition. Asymptotically the log-likelihood is a measurable function of the signed cycles. As a consequence, in the contiguous regime, knowing the signed cycles is enough for obtaining the asymptotically optimal test for (4.1.1). On the other hand, in the singular regime, one has a consistent test for (4.1.1) by using the signed cycle statistic the length k_n of which tends to infinity with n at a rate of $o(\min\{\log(np_n), \sqrt{\log n}\})$ (Banerjee, 2018, Proposition 4.1). Here and after, for any two sequences of positive numbers x_n and y_n, we write $x_n = O(y_n)$ if x_n/y_n is uniformly bounded by a numeric constant and $x_n = o(y_n)$ or $x_n \ll y_n$ if x_n/y_n converges to zero as $n \to \infty$.

These results are satisfying from a statistical optimality viewpoint because the Neyman–Pearson lemma dictates that the likelihood ratio test is optimal for the simple vs. simple testing problem (4.1.1). However, there are two major drawbacks. First, neither the likelihood ratio test nor any test involving signed cycles of diverging lengths is computationally tractable. In particular, evaluation of the likelihood function of the alternative is of exponential time complexity, and calculating the signed cycle of length k directly requires enumeration of all node subsets of size k which is of $O(\binom{n}{k})$ time complexity. It grows faster than any polynomial of n as long as k diverges with n. In addition, to decide on which test statistic to use, one needs to know the null and alterative (or at least the value of t in (4.1.4)), and so one does not yet have a test procedure that is adaptive to different null and alternative hypotheses.

In view of the foregoing shortcomings, we pursue answers to the following two questions in the present paper:

- Can one achieve the sharp asymptotic optimal power of the likelihood ratio test in the contiguous regime with a test of polynomial time complexity?

- Can one design an adaptive test which achieves nontrivial power in the contiguous regime and consistency in the singular regime?

4.1.1. Main contributions

In this paper, we provide affirmative answers to both of the foregoing questions under appropriate conditions, which are summarized as the following main contributions:

1. For appropriately rescaled graph adjacency matrices, we derive joint central limit theorems for their linear spectral statistics (LSSs) of power functions under both the null and local alternative hypotheses in the growing degree asymptotic regime. An important feature of the central limit theorems is that we allow the powers in LSSs to grow to infinity with the graph size. The proof of these CLTs based on the ideas of Füredi–Komlós enumeration and unicyclic graphs further reveals a deep connection between LSSs of Chebyshev polynomials and signed cycles.

2. Based on the connection between the spectrum of an adjacency matrix and signed cycles, given the knowledge of both hypotheses in (4.1.1) (or the quantity t in (4.1.4)), we propose a test based on a special linear spectral statistic. The test statistic can be evaluated within $\widetilde{O}(n^3)$ time complexity and achieves the sharp optimal asymptotic power as the likelihood ratio test in the contiguous regime under the additional condition that $np_n^2 \to \infty$. If only $np_n \to \infty$ holds, we have a slightly different test with $\widetilde{O}(n^3)$ time complexity that achieves a nontrivial fraction of the optimal power in the contiguous regime. It is worth noting that regardless of the rate at which p_n scales with n, no community detection method can perform better than random guessing within the contiguous regime. In other words, we can only tell with nontrivial probability that the random graph comes from $\mathcal{G}_2(n, p_n, q_n)$ while having little idea about how the nodes are partitioned into communities. Based on our limited knowledge, the present paper is one of the first to achieve the exact asymptotic power of the

likelihood ratio test on a non-Gaussian model in high dimensions.

3. Further exploiting the connection between LSSs and signed cycles, we propose several adaptive tests for (4.1.1). These tests are data-driven, of $O(n^3 \log n)$ time complexity and do not require knowledge of p_n or q_n. Moreover, they achieve nontrivial power in the contiguous regime and consistency in the singular regime. We also show that they remain consistent when the alternative is some symmetric SBM with $\kappa > 2$ blocks and the separation is above the Kesten–Stigum threshold Decelle et al. (2011).

4.1.2. Related works

The present work is closely related to a large body of work on the edge scaling limits of Erdős-Renyi, and more generally of Wigner matrices. Consider first the case where $p_n \to p > 0$ and so the average degree of the graph grows linearly with the graph size n. In this case, the rescaled graph adjacency matrix (s.t. the entries are i.i.d. random variables with mean zero and variance $1/n$) is a Wigner matrix under the null hypothesis in (4.1.1), and can essentially be viewed as a rank-one perturbation of a Wigner matrix under the alternative with operator norm of the perturbation given by t in (4.1.4). It is well known that the largest eigenvalue of a Wigner matrix converges to 2 almost surely and have the GOE Tracy–Widom scaling limit under a fourth moment condition. Assuming the perturbation is positive semi-definite (corresponding to $p_n > q_n$ in (4.1.1)), the largest eigenvalue of a rank-one deformed Wigner matrix undergoes a phase transition. In particular, it converges to 2 or $t+1/t$ depending on whether $t < 1$ or $t > 1$. In addition, it has a GOE Tracy–Widom limit when $t < 1$ and a non-universal scaling limit when $t > 1$. See for instance Féral and Péché (2007); Capitaine et al. (2009); Pizzo et al. (2013) for more details. Note that the threshold $t = 1$ is exactly the threshold between the contiguous and the singular regimes for the null and alternative hypotheses in (4.1.1). The phase transition thus suggests that any test based solely on the largest eigenvalue has trivial power within the contiguous regime for having the same scaling limit under null and alternative. A result of this flavor was first discovered by Baik et al. (2005) for complex sample covariance matrices.

When p_n and $q_n \to 0$, the average degree of the graph grows sub-linearly with the graph size. In this regime, results about the edge scaling limits under either the null or the alternative are less complete compared with the linear degree growth regime. Under the null, the convergence to the GOE Tracy–Widom limit was established in Erdős et al. (2012) under the assumption that $np_n^3 \to \infty$. Lee and Schnelli (2016) weakened the condition to $n^2p_n^3 \to \infty$. Turning to the alternative. Suppose $p_n > q_n$. Erdős et al. (2013) showed that the largest eigenvalue of the rescaled graph adjacency matrix converges in probability to $t + 1/t$ whenever $np_n \gg (\log n)^{6\xi}$ for some $1 < \xi = O(\log \log n)$ and $t > 1$. Further, it was proved in Erdős et al. (2013) when $t > C_0(\log n)^{2\xi}$ for some large constant C_0, the largest eigenvalue has a $\sqrt{2/n}$ fluctuation and a normal scaling limit. To the best of our knowledge, little is known about the asymptotic null distribution of the largest eigenvalue when $n^2p_n^3$ is bounded or about its distribution under any local alternative when t is a constant.

As discussed earlier, one of the main contributions of the present paper is to link signed cycles and linear spectral statistics. In that sense analyzing linear spectral statistics of the rescaled adjacency matrices of Erdős-Renyi lies at the heart of our technical analysis. There are a series of papers on linear spectral statistics of Wigner and Wishart matrices relying on the methods introduced by Bai and Silverstein (2004). See, in particular, Bai et al. (2009) for CLTs of linear spectral statistics of Wigner matrices. These techniques are however specific to the asymptotic regime where the average degree grows linearly with graph size. In this paper we adopt the combinatorial methods developed by Anderson and Zeitouni (2006) which we modify and use for all growing degree cases, regardless of the growth rate.

In addition, our results are connected with the literature on optimal hypothesis testing in high dimensions. Onatski et al. Onatski et al. (2013, 2014) studied the optimal tests for an identical covariance (or correlation) matrix against a spiked local alternative for Gaussian data when the sample size and the ambient dimension grow proportionally to infinity. Remarkably, they further studied the asymptotic powers of the Gaussian likelihood ratio tests for non-Gaussian data. Dharmawansa et al. (2014) and Johnstone and

Onatski (2015) studied analogous questions for an exhaustive collection of testing problems in various "double Wishart" scenarios where the sufficient statistics of the observations are essentially two independent Wishart matrices. See also Dobriban (2017) for an important extension of Onatski et al. (2013, 2014) where one is allowed to have general covariance matrices as the null model. From a slightly different viewpoint, Cai and Ma (2013) and Cai et al. (2015) studied minimax optimal hypothesis testing for an identity covariance matrix. The concurrent work by Gao and Lafferty (2017) studied minimax rates for testing Erdős-Renyi against SBMs and more general alternatives. Interestingly, one of their proposed test statistics is asymptotically equivalent to the signed cycle of length three, also known as the signed triangle Bubeck et al. (2014).

4.1.3. Organization

The rest of the paper is organized as the following. After a brief introduction of definitions and notation in Section 4.2, we establish in Section 4.3 joint central limit theorems under both the null and local alternatives for linear spectral statistics of rescaled graph adjacency matrices and their connection to signed cycles. In addition, we propose a computationally tractable testing procedure based on these findings that achieves the same optimal asymptotic power as the likelihood ratio test. Section **??** investigates adaptive testing procedures for (4.1.1) and we also study the powers of the proposed tests under symmetric multi-block alternatives. We give an outline of the proofs in Section 4.4 and the detailed proofs are presented in a supplement Banerjee and Ma (2017b). Finally, we conclude in Section **??**.

4.2. Definitions and notation

We first introduce some preliminary definitions and notation to be used throughout the paper. We let $\mathrm{E}_{i,n}$ and $\mathrm{Var}_{i,n}$ denote expectation and variance under $\mathbb{P}_{i,n}$ for $i = 0$ and 1. For any random graph G, its adjacency matrix will be denoted by A and $x_{i,j}$ (instead of $a_{i,j}$) will be used to denote the indicator random variable corresponding to an edge between the nodes i and j. We denote the expected average connection probability and its sample counterpart by

$$p_{n,\text{av}} = \frac{1}{n(n-1)} \sum_{i \neq j} \mathrm{E}_{0,n}[x_{i,j}], \quad \text{and} \quad \widehat{p}_{n,\text{av}} = \frac{1}{n(n-1)} \sum_{i \neq j} x_{i,j}. \tag{4.2.1}$$

Under out settings, $p_{n,\text{av}}$ remains unchanged if we replace $\mathrm{E}_{0,n}$ with $\mathrm{E}_{1,n}$ in its definition. The signed cycle of length k of the graph G is defined to be

$$C_{n,k}(G) = \left(\frac{1}{\sqrt{np_{n,\text{av}}(1-p_{n,\text{av}})}} \right)^k \sum_{i_0, i_1, \ldots, i_{k-1}} (x_{i_0,i_1} - p_{n,\text{av}}) \ldots (x_{i_{k-1}i_0} - p_{n,\text{av}}) \tag{4.2.2}$$

where $i_0, i_1, \ldots, i_{k-1}$ are all distinct. We define the following centered and scaled versions of the adjacency matrix A. For any $n \in \mathbb{N}$, let $1_n = (1, \ldots, 1)' \in \mathbb{R}^n$ and I_n be the $n \times n$ identity matrix. Then

$$A_{\text{cen1}} := \frac{A - p_{n,\text{av}}(1_n 1_n' - I_n)}{\sqrt{np_{n,\text{av}}(1-p_{n,\text{av}})}}\,, \tag{4.2.3}$$

and

$$A_{\text{cen2}} := \frac{A - \widehat{p}_{n,\text{av}}(1_n 1_n' - I_n)}{\sqrt{n\widehat{p}_{n,\text{av}}(1-\widehat{p}_{n,\text{av}})}}\,. \tag{4.2.4}$$

Note that A_{cen2} is completely data-driven. If A is a random instance of the Erdős-Renyi $\mathcal{G}_1(n, p_{n,\text{av}})$, then A_{cen1} has zeros on the diagonal and the sub-diagonal entries (subject to symmetry) are i.i.d. with mean zero and variance $1/n$.

We now introduce an important generating function. Given any $r \in \mathbb{N}$, let

$$\left(\frac{1 - \sqrt{1-4z^2}}{2z} \right)^r = \sum_{m=r}^{\infty} f(m,r) z^m. \tag{4.2.5}$$

The coefficients $f(m,r)$'s are key quantities for defining the variances and covariances of linear spectral statistics constructed from different power functions. For any $k \in \mathbb{N}$ denote

$$\psi_k = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{1}{\frac{k}{2}+1}\binom{k}{\frac{k}{2}} & \text{if } k \text{ is even.} \end{cases} \tag{4.2.6}$$

So ψ_k is the $\frac{k}{2}$-th Catalan number for every even k. Finally, we define a set of rescaled Chebyshev polynomials. These polynomials are important for drawing the connection between signed cycles and the spectrum of adjacency matrix. The standard Chebyshev polynomial of degree m is denoted by $S_m(x)$ and can be defined by the identity

$$S_m(\cos(\theta)) = \cos(m\theta). \tag{4.2.7}$$

In this paper we use a slight variant of S_m, denoted by P_m and defined as

$$P_m(x) = 2S_m\left(\frac{x}{2}\right). \tag{4.2.8}$$

In particular, $P_m(2\cos(\theta)) = 2\cos(m\theta)$. It is easy to note that $P_m\left(z + z^{-1}\right) = z^m + z^{-m}$ for all $z \in \mathbb{C}$. One also notes that $P_m(\cdot)$ is even and odd whenever m is even or odd respectively.

Throughout the paper, we use $C, C_1, C_2, \ldots$ to denote positive numeric constants and their values may value from occurrence to occurrence. For any matrix (and vector) U, U' stands for its transpose.

4.3. Linear spectral statistics and likelihood ratio tests

4.3.1. Joint CLTs for LSSs of power functions

We first characterize the asymptotic joint normality of linear spectral statistics of the form $\sum_{i=1}^n g(\lambda_i)$ where $\lambda_1 \geq \cdots \geq \lambda_n$ are ordered eigenvalues of either A_{cen1} in (4.2.3) or A_{cen2} in (4.2.4) and $g(\lambda) = \lambda^k$ for some integer $k \geq 2$. For convenience, we often write the statistic as $\mathsf{Tr}(A_{\text{cen}i}^k)$ for $i = 1, 2$. In what follows, we separate the discussion under the null and the alternative hypotheses in (4.1.1).

Results under the null

Recall the definition of $f(m,r)$ in (4.2.5). For any $1 \le k_1 < \cdots < k_l$, define an $l \times l$ symmetric matrix $\Sigma_{2k_1+1,\ldots,2k_l+1}$ by setting its (i,j)-th entry as

$$\Sigma_{2k_1+1,\ldots,2k_l+1}(i,j) = \sum_{r=3:r \text{ odd}}^{\min(2k_i+1,2k_j+1)} 2f(2k_i+1,r)f(2k_j+1,r)\frac{(2k_i+1)(2k_j+1)}{r}. \quad (4.3.1)$$

In addition, define a second $l \times l$ symmetric matrix $\widetilde{\Sigma}_{2k_1,\ldots,2k_l}$ by setting its (i,j)-th entry as

$$\begin{aligned}\widetilde{\Sigma}(2k_1,\ldots,2k_l)(i,j) = &\sum_{r=4:r \text{ even}}^{\min(2k_i,2k_j)} 2f(2k_i,r)f(2k_j,r)\frac{(2k_i)(2k_j)}{r} \\ &+ 2(k_ik_j\psi_{2k_i}\psi_{2k_j})\lim_{n\to\infty}\frac{\mathrm{Var}_{0,n}\left[(x_{1,2}-\mathrm{E}_{0,n}[x_{1,2}])^2\right]}{p_{n,\mathrm{av}}^2(1-p_{n,\mathrm{av}})^2}.\end{aligned} \quad (4.3.2)$$

Here and after, we may omit the subscripts in variance and expectation when there is no ambiguity. With the foregoing definitions, we have the following results under $\mathbb{P}_{0,n}$.

Theorem 7. *Suppose $A \sim \mathbb{P}_{0,n}$ and $np_{n,\mathrm{av}} \to \infty$. For any fixed $l \ge 1$, we have:*

(i) If $1 \le k_1 < \ldots < k_l = o(\log(np_{n,\mathrm{av}}))$, then for $\Sigma = \Sigma_{2k_1+1,\ldots,2k_l+1}$ defined in (4.3.1)

$$\Sigma^{-\frac{1}{2}}\left(\mathsf{Tr}(A_{\mathrm{cen1}}^{2k_1+1}),\ldots,\mathsf{Tr}(A_{\mathrm{cen1}}^{2k_l+1})\right)' \xrightarrow{d} N_l(0,I_l). \quad (4.3.3)$$

(ii) Suppose $1 \le k_1 < \ldots < k_l = o(\log(np_{n,\mathrm{av}}))$. If $p_{n,\mathrm{av}} \to 0$,

$$\frac{\sqrt{p_{n,\mathrm{av}}}\,(\beta_{2k_i})}{\sqrt{2}k_i\psi_{2k_i}} \xrightarrow{d} N(0,1) \quad (4.3.4)$$

where

$$\beta_{2k_i} = \mathsf{Tr}(A_{\mathrm{cen1}}^{2k_i}) - \mathrm{E}_{0,n}(\mathsf{Tr}(A_{\mathrm{cen1}}^{2k_i})).$$

Further, $\mathrm{Cov}\left(\sqrt{p_{n,\mathrm{av}}}\beta_{2k_i},\sqrt{p_{n,\mathrm{av}}}\beta_{2k_j}\right) - 2(k_ik_j\psi_{2k_i}\psi_{2k_j}) = O\left(\frac{1}{np_{n,\mathrm{av}}}\right)$. *In other words,* *when* $p_{n,\mathrm{av}} \to 0$*, asymptotically the even moments are constant multiples of each other after rescaling.*

If $p_{n,\mathrm{av}} \to p \in (0,1)$*, then for* $\widetilde{\Sigma} = \widetilde{\Sigma}_{2k_1,\ldots,2k_l}$ *defined in* (4.3.2)

$$(\widetilde{\Sigma})^{-\frac{1}{2}} \left(\beta_{2k_1}, \ldots, \beta_{2k_l}\right)' \xrightarrow{d} N_l(0, I_l). \tag{4.3.5}$$

(iii) For any $k_i = o(\log(np_{n,\mathrm{av}}))$*,*

$$\mathsf{Tr}(A_{\mathrm{cen1}}^{2k_i+1}) - \sum_{r=3:r\ odd}^{2k_i+1} f(2k_i+1, r)\frac{2k_i+1}{r} C_{n,r}(G) \xrightarrow{p} 0. \tag{4.3.6}$$

(iv) If $1 \le k_1 < \ldots < k_l = o\left(\min\left(\log(np_{n,\mathrm{av}}), \sqrt{\log n}\right)\right)$*, results in (i) and (iii) continue to hold when* A_{cen1} *is replaced with* A_{cen2}*.*

For finite k, parts (i) and (ii) in Theorem 7 are known in the literature. See, for instance, Anderson and Zeitouni (2006) or Anderson et al. (2010, pp.30-35) for reference. In particular, the variance expression given in (Anderson et al., 2010, Equation (2.1.44)) matches with (4.3.2) and differs from (4.3.1) at only the term involving $\mathrm{E}[Y_1^2]$. This term comes from the diagonal entries which are zeros in our case. The significance of (i) and (ii) in Theorem 7 (and Theorem 8 below) lies in the fact that the CLTs continue to hold when the powers grow to infinity with the graph size n.

Results under local alternatives

Recall the definition of c and t in (4.1.2) and (4.1.4). Our next result gives the counterpart of Theorem 7 under any local alternative where c and hence t are finite.

Theorem 8. *Suppose that* $A \sim \mathbb{P}_{1,n}$ *and that as* $n \to \infty$*,* $np_{n,\mathrm{av}} \to \infty$ *while* t *in* (4.1.4) *remains a constant. For any fixed* $l \ge 1$*, we have:*

(i) If $1 \le k_1 < \ldots < k_l = o(\min(\log(np_{n,\mathrm{av}}), \sqrt{\log n}))$*, then for* $\Sigma = \Sigma_{2k_1+1,\ldots,2k_l+1}$ *in*

(4.3.1)

$$\Sigma^{-\frac{1}{2}}\left(\mathsf{Tr}(A_{\text{cen1}}^{2k_1+1})-\nu_{2k_1+1},\ldots,\mathsf{Tr}(A_{\text{cen1}}^{2k_l+1})-\nu_{2k_l+1}\right)' \xrightarrow{d} N_l(0,I_l) \tag{4.3.7}$$

where if $p_n > q_n$, for $\mu_r = t^r$, $r = 1, 2, \ldots,$

$$\nu_{2k_i+1} = \sum_{r=3:r\ odd}^{2k_i+1} f(2k_i+1, r)\frac{2k_i+1}{r}\mu_r.$$

If $p_n < q_n$, we set $\mu_r = (-t)^r$ for all $r \geq 1$.

(ii) If $1 \leq k_1 < \ldots < k_l = o\left(\min\left(\log(np_{n,\text{av}}), \sqrt{\log n}\right)\right)$, then (4.3.4) (resp. (4.3.5)) continues to hold when $p_{n,\text{av}} \to 0$ (resp. when $p_{n,\text{av}} \to p$) where the expectation in the definition of β_{2k_i} is now taken under $\mathbb{P}_{1,n}$ while the definition of $\widetilde{\Sigma}$ remains unchanged.

(iii) For any $k_i = o\left(\min\left(\log(np_{n,\text{av}}), \sqrt{\log n}\right)\right)$, (4.3.6) continues to hold.

(iv) If $1 \leq k_1 < \ldots < k_l = o\left(\min\left(\log(np_{n,\text{av}}), \sqrt{\log n}\right)\right)$, results in (i) and (iii) continue to hold when A_{cen1} is replaced with A_{cen2}.

Theorem 7, Proposition 3.4 in Banerjee (2018) and Le Cam's Third Lemma Le Cam (2012) jointly imply claims (i) and (iv) in Theorem 8 within the contiguous regime, i.e., when $c < 2(1-p)$ or equivalently $t < 1$. The significance of Theorem 8 is that the CLTs continue to hold in the singular regime as long as t is finite. It requires a dedicated proof.

When $p_{n,\text{av}} \to 0$, for traces of even powers of A_{cen1} and A_{cen2}, the second term on the right side of (4.3.2) dominates. This explains the result in (4.3.4). Indeed, one can further show that $\frac{\sqrt{p_{n,\text{av}}}\beta_{2k_i}}{\sqrt{2k_i}\psi_{2k_i}}$ is asymptotically the same as a rescaled version of the average degree of the graph when c (and hence t) is finite. On the other hand, it is more complicated to state the counterpart of claim (iii) in both Theorem 7 and Theorem 8 for traces of even powers and signed cycles of even lengths, which is our next focus.

Connection between traces of even powers and even signed cycles

Fix any integer $k \geq 2$. We first decompose $\mathsf{Tr}(A_{\text{cen1}}^{2k})$ as

$$\mathsf{Tr}(A_{\text{cen1}}^{2k}) = n\psi_{2k} - \binom{k+1}{2}\psi_{2k} + R_{1,2k} + R_{2,2k} + T_{2k}, \tag{4.3.8}$$

where

$$R_{1,2k} := k\psi_{2k}\left[\mathsf{Tr}\left(A_{\text{cen1}}^{2}\right) - n + 1\right], \ R_{2,2k} := \sum_{r=4:r \text{ even}}^{2k} f(2k,r)\frac{2k}{r}C_{n,r}(G), \tag{4.3.9}$$

and T_{2k} is the remainder term. Observe that under both $\mathbb{P}_{0,n}$ and local $\mathbb{P}_{1,n}$,

$$\sqrt{p_{n,\text{av}}}\left[\mathsf{Tr}\left(A_{\text{cen1}}^{2}\right) - n + 1\right] \xrightarrow{d} N(0,\sigma^2).$$

Here $\sigma^2 = 2\lim_{n\to\infty}\frac{\text{Var}_{0,n}\left[(x_{1,2}-\text{E}_{0,n}[x_{1,2}])^2\right]}{p_{n,\text{av}}(1-p_{n,\text{av}})^2}$. When $p_{n,\text{av}} \to 0$, the scaling $\sqrt{p_{n,\text{av}}}$ kills the mean shift in $R_{2,2k}$ which leads to the degeneracy in the asymptotics. However, this can be circumvented by working with the difference

$$\mathsf{Tr}(A_{\text{cen1}}^{2k}) - k\psi_{2k}\,\mathsf{Tr}\left(A_{\text{cen1}}^{2}\right).$$

One can show that under appropriate conditions, T_{2k} has negligible fluctuation around its mean. Therefore, this difference has a nontrivial asymptotic normal distribution without any further scaling, which is a direct consequence of the joint asymptotic normality of the signed cycles Banerjee (2018). This is described in more details in Theorem 9 below.

The following theorem characterizes the first and second moments of T_{2k}.

Theorem 9. *Suppose $np_{n,\text{av}}^2 \to \infty$ and $k = o(\min(\log(np_{n,\text{av}}^2), \sqrt{\log(n)}))$ as $n \to \infty$. Under* *both* $\mathbb{P}_{0,n}$ *and any local* $\mathbb{P}_{1,n}$*, if* $p_{n,\mathrm{av}} \to 0$*, one has for* $i \in \{0,1\}$

$$\mathrm{E}_{i,n}[T_{2k}] = (1+o(1))\left[\alpha_{1,2k} + \frac{\alpha_{2,2k}}{p_{n,\mathrm{av}}} + \frac{\alpha_{3,2k}}{np_{n,\mathrm{av}}^2} + \varepsilon_{i,2k}^{(1)}\right],$$
$$\mathrm{Var}_{i,n}[T_{2k}] = (1+o(1))\left[\frac{v_{2,2k}}{np_{n,\mathrm{av}}^2} + \frac{v_{1,2k}}{n^2 p_{n,\mathrm{av}}^3} + \varepsilon_{i,2k}^{(2)}\right],$$

where

$$\alpha_{1,2k} = 2^{2k-1} - \binom{2k}{k}\frac{5k+1}{2(k+1)} + \binom{k+1}{2}\psi_{2k} - 3\binom{2k}{k+2},\ \alpha_{2,2k} = \binom{2k}{k+2},$$

$\alpha_{3,2k} \le 2^{2k}(2k)^{12}$ *and* $v_{2,2k}, v_{1,2k} \le 2^{4k}(C_1 k)^{C_2}$ *for some numeric constants* $C_1, C_2 > 0$*. Moreover,* $\varepsilon_{i,2k}^{(2)} \to 0$ *and* $\varepsilon_{i,2k}^{(1)} \to 0$ *for* $i = 0,1$*.*

When $p_{n,\mathrm{av}} \to p > 0$*, we replace the multiplier* $\frac{1}{p_{n,\mathrm{av}}}$ *in the second term of* $\mathrm{E}_{i,n}[T_{2k}]$ *with* $\lim_{n\to\infty} \frac{\mathrm{Var}_{0,n}\left[(x_{1,2}-\mathrm{E}_{0,n}[x_{1,2}])^2\right]}{p_{n,\mathrm{av}}^2(1-p_{n,\mathrm{av}})^2}$*, while all the other conclusions remain the same.*

4.3.2. Approximation of signed cycles by LSSs

Theorems 7–9 suggest that signed cycles and linear spectral statistics of properly rescaled adjacency matrices are closely connected. In what follows, we further formalize this idea and demonstrate how one could approximate signed cycles of growing lengths with carefully chosen linear spectral statistics.

As an illustration, let

$$\overrightarrow{C}_{n,2k+1} := (C_{n,3}(G), C_{n,5}(G), \ldots, C_{n,2k+1}(G))', \quad \text{and}$$
$$\overrightarrow{\mathrm{Tr}}_{n,2k+1} := \left(\mathsf{Tr}(A_{\mathrm{cen2}}^3), \mathsf{Tr}(A_{\mathrm{cen2}}^5), \ldots, \mathsf{Tr}(A_{\mathrm{cen2}}^{2k+1})\right)'.$$

We proved in Theorems 7 and 8 that under both null and local alternatives, whenever $k = o\left(\min\left(\log(np_{n,\mathrm{av}}), \sqrt{\log n}\right)\right)$ and $np_{n,\mathrm{av}} \to \infty$, we have elementwise,

$$\mathbb{D}_{2k+1}\overrightarrow{C}_{n,2k+1} - \overrightarrow{\mathrm{Tr}}_{n,2k+1} \xrightarrow{p} 0.$$

Here $\mathbb{D}_{2k+1}$ is the $k \times k$ lower triangular matrix given by

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{5f(5,3)}{3} & 1 & 0 & \dots & 0 \\ \frac{7f(7,3)}{3} & \frac{7f(7,5)}{5} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \frac{(2k+1)f(2k+1,3)}{3} & \frac{(2k+1)f(2k+1,5)}{5} & \frac{(2k+1)f(2k+1,7)}{7} & \dots & 1 \end{pmatrix}. \tag{4.3.10}$$

Using the fact (Lang, 2000, Lemma 2) that

$$f(m,r)\frac{m}{r} = \binom{m}{\frac{m+r}{2}}, \tag{4.3.11}$$

one can prove that

$$(\mathbb{D}_{2k+1}^{-1})_{k,j} = P_{2k+1}[2j+1]$$

where $P_j[i]$ is the coefficient of z^i in the polynomial $P_j(z)$ defined in (4.2.8). See, for instance, (Lang, 2000, Equation (37)). An analogous result holds for signed cycles of even lengths, in which case one needs to take in account the random variables T_{2k} to offset the mean values of the even powers. Formally, we have the following result.

Theorem 10. *Suppose $np_{n,\mathrm{av}} \to \infty$ and $k = o(\min(\log(np_{n,\mathrm{av}}), \sqrt{\log(n)}))$ as $n \to \infty$. The following results hold under both $\mathbb{P}_{0,n}$ and local $\mathbb{P}_{1,n}$:*

(i) (Construction of odd signed cycles from LSS) We have

$$C_{n,2k+1}(G) - \mathsf{Tr}\left(P_{2k+1}(A_{\mathrm{cen2}})\right) \xrightarrow{p} 0. \tag{4.3.12}$$

(ii) (Construction of even signed cycles from LSS) Let $T_0 = T_2 = 0$. Then

$$C_{n,2k}(G) - \mathsf{Tr}\left(P_{2k}(A_{\mathrm{cen2}})\right) - \sum_{r=0:r\ \mathit{even}}^{2k} P_{2k}[r]\left[T_r - \binom{\frac{r}{2}+1}{2}\psi_r\right] \xrightarrow{p} 0. \tag{4.3.13}$$

If further $np_{n,\mathrm{av}}^2 \to \infty$ and $k = o(\min(\log(np_{n,\mathrm{av}}^2), \sqrt{\log(n)}))$, then we may replace T_r

in (4.3.13) *with* $\mathrm{E}_{0,n}[T_r]$.

For the third term on the left side of (4.3.13), we do not have other deterministic terms involved in (4.3.8) because of the following cancellation (see supplement Banerjee and Ma (2017b) for proofs)

$$\sum_{r=0}^{k} P_{2k}[2r]\psi_{2r} = 0, \quad \text{and} \quad \sum_{r=1}^{k} P_{2k}[2r]r\psi_{2r} = 0, \quad \text{for all } k \geq 2. \tag{4.3.14}$$

Remark 10. A careful examination of the proofs in Banerjee and Ma (2017b) shows that all the conclusions under local alternatives in Theorems 8 – 10 actually hold conditioning on the group assignments σ_i, $1 \leq i \leq n$. Thus, the approximation of signed cycles by LSSs of Chebyshev polynomials works for any group assignment configurations as long as t in (4.1.4) is finite.

4.3.3. Likelihood ratio tests

Recall the null and alternative hypotheses in (4.1.1) with the key index t defined as in (4.1.4). Banerjee (2018) showed that if $np_{n,\mathrm{av}} \to \infty$ as $n \to \infty$ and $p_n > q_n$, in the contiguous regime, i.e. $0 < t < 1$, the likelihood ratio test is asymptotically the same as the test that rejects for large values of

$$L_c := \sum_{r=3}^{\infty} \frac{t^r C_{n,r}(G)}{2r} \tag{4.3.15}$$

which has the following asymptotic distributions under the null and alternative:

$$L_c|\mathbb{P}_{0,n} \xrightarrow{d} N\left(0, \sigma(t)^2\right) \quad \text{and} \quad L_c|\mathbb{P}_{1,n} \xrightarrow{d} N\left(\sigma(t)^2, \sigma(t)^2\right), \tag{4.3.16}$$

where

$$\sigma(t)^2 = \frac{1}{2}\left[-\log(1-t^2) - t^2 - \frac{t^4}{2}\right]. \tag{4.3.17}$$

If $p_n < q_n$, we replace every t^r in (4.3.15) with $(-t)^r$ while everything else remains the same. Hence, at any given $t \in (0,1)$, the largest asymptotic power achievable by any level α test is

$$\Phi\left(-z_\alpha + \sigma(t)\right), \tag{4.3.18}$$

where Φ is the CDF of the standard normal distribution and $z_\alpha = \Phi^{-1}(1-\alpha)$. However, neither the exact likelihood ratio test nor the test in (4.3.15) based on sign cycles is computationally tractable.

Given Theorems 7–10, one of the key findings of the present paper is that when $np_{n,\mathrm{av}}^2 \to \infty$, we can achieve the exact asymptotic optimal power (4.3.18) by a test based on some linear spectral statistic, which is of $\widetilde{O}(n^3)$ time complexity. If we only have $np_{n,\mathrm{av}} \to \infty$, we propose a slightly different test based on another linear spectral statistic that has a smaller but nontrivial asymptotic power in the contiguous regime. In particular, we have the following theorem.

Theorem 11. *Suppose that as $n \to \infty$, t defined in* (4.1.4) *satisfies $t \in (0,1)$. Then the following results hold if $p_n > q_n$:*

(i) When $np_{n,\mathrm{av}}^2 \to \infty$ and $k_n = o(\min(\log(np_{n,\mathrm{av}}^2), \sqrt{\log(n)})) \to \infty$, then the test statistic

$$L_a = \sum_{r=3}^{k_n} \frac{t^r \,\mathsf{Tr}\left(P_r\left(A_{\mathrm{cen2}}\right)\right)}{2r} \tag{4.3.19}$$

satisfies

$$\begin{aligned} L_a - \mu_{n,p_{n,\mathrm{av}}}(t)|\mathbb{P}_{0,n} &\xrightarrow{d} N\left(0, \sigma(t)^2\right), \\ L_a - \mu_{n,p_{n,\mathrm{av}}}(t)|\mathbb{P}_{1,n} &\xrightarrow{d} N\left(\sigma(t)^2, \sigma(t)^2\right), \end{aligned} \tag{4.3.20}$$

where $\mu_{n,p_{n,\mathrm{av}}}(t)$ is a deterministic quantity depending only on $n, p_{n,\mathrm{av}}$ and t. Therefore, a level α test that rejects for large values of L_a achieves the exact asymptotic optimal power (4.3.18).

(ii) When $np_{n,\mathrm{av}} \to \infty$ *and* $k_n = o(\min(\log(np_{n,\mathrm{av}}), \sqrt{\log(n)})) \to \infty$*, then the test statistic*

$$L_o = \sum_{r=1}^{k_n} \frac{t^{2r+1}\,\mathsf{Tr}(P_{2r+1}(A_{\mathrm{cen2}}))}{2(2r+1)} \tag{4.3.21}$$

satisfies

$$L_o|\mathbb{P}_{0,n} \xrightarrow{d} N\left(0, \sigma_1(t)^2\right) \quad \text{and} \quad L_o|\mathbb{P}_{1,n} \xrightarrow{d} N\left(\sigma_1(t)^2, \sigma_1(t)^2\right) \tag{4.3.22}$$

where

$$\sigma_1(t)^2 = \frac{1}{4}\left[-\log\left(\frac{1-t^2}{1+t^2}\right) - 2t^2\right]. \tag{4.3.23}$$

Therefore, a level α *test that rejects for large values of* L_o *achieves an asymptotic power of* $\Phi(-z_\alpha + \sigma_1(t))$.

If $p_n < q_n$*, we replace every* t *in the definitions of* (4.3.19) *and* (4.3.21) *with* $-t$*, and the same conclusions hold.*

We conclude the section with a discussion on the quantity $\mu_{n,p_{n,\mathrm{av}}}(t)$. First suppose $p_{n,\mathrm{av}} \to 0$. When $np_{n,\mathrm{av}}^2 \to \infty$ and $k_n = o(\min(\log(np_{n,\mathrm{av}}^2), \sqrt{\log(n)})) \to \infty$, we have from Theorems 9 and 10 that

$$\mu_{n,p_{n,\mathrm{av}}}(t) = \frac{1}{2}\sum_{i=1}^{\infty} \frac{1}{2i}\left(\frac{t}{1+t^2}\right)^{2i}\left[\alpha_{1,2i} + \frac{\alpha_{2,2i}}{p_{n,\mathrm{av}}} - \binom{i+1}{2}\psi_{2i}\right] - \frac{1}{2}\log(1+t^2). \tag{4.3.24}$$

Here $\alpha_{1,2i}, \alpha_{2,2i}$ and $\alpha_{3,2i}$ have been defined in Theorem 9. Although we do not have explicit formula for $\alpha_{3,2i}$'s, we may estimate them by simulation under $\mathbb{P}_{0,n}$ with an estimated $\widehat{p}_{n,\mathrm{av}}$. To obtain (4.3.24), we have used the following generating function of Chebyshev polynomials

$$\sum_{i=1}^{\infty} \frac{t^i P_i(x)}{i} = \log\left(\frac{1}{1-tx+t^2}\right).$$

If further $np_{n,\mathrm{av}}^2 \to \infty$ and $k_n = o(\min(\log(np_{n,\mathrm{av}}^2), \sqrt{\log(n)})) \to \infty$, then we may replace the right side in (4.3.24) with the following explicit expression

$$\frac{1}{2}\sum_{i=1}^{\infty}\frac{1}{2i}\left(\frac{t}{1+t^2}\right)^{2i}\left[\alpha_{1,2i}+\frac{\alpha_{2,2i}}{p_{n,\mathrm{av}}}-\binom{i+1}{2}\psi_{2i}\right]-\frac{1}{2}\log(1+t^2)$$
$$=\frac{1}{2}\sum_{i=1}^{\infty}\frac{1}{2i}\left(\frac{t}{1+t^2}\right)^{2i}\left[2^{2i-1}-\binom{2i}{i}\left[\frac{5i+1}{2i+2}\right]+\binom{2i}{i+2}\left[\frac{1}{p_{n,\mathrm{av}}}-3\right]\right]-\frac{1}{2}\log(1+t^2). \tag{4.3.25}$$

When $p_{n,\mathrm{av}} \to p > 0$, we replace $1/p_{n,\mathrm{av}}$ in the term involving $\alpha_{2,2i}$ with

$$\lim_{n\to\infty}\frac{\mathrm{Var}_{0,n}\left[(x_{1,2}-\mathrm{E}_{0,n}[x_{1,2}])^2\right]}{p_{n,\mathrm{av}}^2(1-p_{n,\mathrm{av}})^2}$$

while the others are the same.

Remark 11. Suppose for simplicity $p_n > q_n$. One might observe that when $t < 1$, the analytic functions used in the LSSs in L_a and L_o in Theorem 11 have limits

$$f_a(x)=\sum_{i=3}^{\infty}\frac{t^iP_i(x)}{2i}=\frac{1}{2}\log\left(\frac{1}{1-tx+t^2}\right)-\frac{tx}{2}-\frac{t^2(x^2-2)}{4} \tag{4.3.26}$$

and

$$f_o(x)=\sum_{i=1}^{\infty}\frac{t^{2i+1}P_{2i+1}(x)}{4i+2}=\frac{1}{4}\log\left(\frac{1+tx+t^2}{1-tx+t^2}\right)-\frac{tx}{2}, \tag{4.3.27}$$

respectively. So it might be tempting to directly use LSSs of the foregoing limits directly as the test statistics in (4.3.19) and (4.3.21) respectively. However, this is not preferable for the following two reasons.

First, observe that given any $t < 1$, both f_a and f_o take finite values only in the open interval $\left(-\left(t+\frac{1}{t}\right), t+\frac{1}{t}\right)$. On the other hand, it is known that the spectral norm of A_{cen1} converges to 2 under the condition $p_{n,\mathrm{av}} \gg \frac{\log(n)^4}{n}$. See Vu (2005) for a reference and using Weyl's interlacing inequality it is easy to see that the same holds for the spectral norm of A_{cen2}. However, the result in (4.3.21) holds as long as $np_{n,\mathrm{av}} \to \infty$. So in this case the test statistic $\mathsf{Tr}\, f_o(A_{\mathrm{cen2}})$ will be undefined when $p_{n,\mathrm{av}} \ll \frac{\log(n)^4}{n}$ with a nontrivial probability. In an unreported simulation study, we find both $\mathsf{Tr}\, f_a(A_{\mathrm{cen2}})$ and $\mathsf{Tr}\, f_o(A_{\mathrm{cen2}})$ highly unstable for small values of $p_{n,\mathrm{av}}$.

4.4. Outline of proofs

In this section we give a brief outline of the proofs for Theorems 7–9. The other theorems and propositions are essentially corollaries of these core results. For conciseness, throughout this section, we focus on the assortative case of $p_n > q_n$ when discussing results under local alternatives.

4.4.1. Outline of proof for Theorem 7 (i)–(iii)

The fundamental idea here is to prove that $\mathsf{Tr}(A^k_{\mathrm{cen1}}) - \mathrm{E}_{0,n}(\mathsf{Tr}(A^k_{\mathrm{cen1}}))$ converges in distribution by using the method of moments and the limiting random variables satisfy Wick's formula and hence are Gaussian. We first state the method of moments.

Lemma 10. *Let $Y_{n,1}, \dots, Y_{n,l}$ be a random vector of dimension l. Then $(Y_{n,1}, \dots, Y_{n,l}) \xrightarrow{d} (Z_1, \dots, Z_l)$ if both of the following conditions are satisfied:*

(i) $\lim_{n\to\infty} \mathrm{E}[X_{n,1} \dots X_{n,m}]$ exists for any fixed m and $X_{n,i} \in \{Y_{n,1}, \dots, Y_{n,l}\}$ for $1 \le i \le m$.

(ii) (Carleman's Condition)Carleman (1926)

$$\sum_{h=1}^{\infty} \left(\lim_{n\to\infty} \mathrm{E}[X^{2h}_{n,i}] \right)^{-\frac{1}{2h}} = \infty \quad \forall\ 1 \le i \le l.$$

Further, $\lim_{n\to\infty} \mathrm{E}[X_{n,1} \dots X_{n,m}] = \mathrm{E}[X_1 \dots X_m]$. Here $X_{n,i} \in \{Y_{n,1}, \dots, Y_{n,l}\}$ for $1 \le i \le m$ and X_i is the in distribution limit of $X_{n,i}$.

Next we state Wick's formula for Gaussian random variables which was first proved by Isserlis (1918) and later on introduced by Wick (1950) in the physics literature.

Lemma 11 (Wick's formula Wick (1950))**.** *Let $(Y_1, \dots, Y_l)$ be a multivariate mean 0 random vector of dimension l with covariance matrix Σ (possibly singular). Then $(Y_1, \dots, Y_l)$ is* *jointly Gaussian if and only if for any positive integer m and $X_i \in \{Y_1, \ldots, Y_l\}$ for $1 \le i \le m$*

$$\mathrm{E}[X_1 \ldots X_m] = \begin{cases} \sum_\eta \prod_{i=1}^{\frac{m}{2}} \mathrm{E}[X_{\eta(i,1)} X_{\eta(i,2)}] & \text{for } m \text{ even} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{4.4.1}$$

Here η is a partition of $\{1, \ldots, m\}$ into $\frac{m}{2}$ blocks such that each block contains exactly 2 elements and $\eta(i,j)$ denotes the jth element of the ith block of η for $j = 1, 2$.

It is worth noting that the random variables $Y_1, \ldots, Y_l$ need not be distinct. When $Y_1 = \cdots = Y_l$, Lemma 11 provides a description of the moments of Gaussian random variables.

In what follows, we focus on odd powers to illustrate the main ideas. Detailed arguments for even powers can be found in the actual proof. We start with the following identity.

$$\mathsf{Tr}(A_{\text{cen1}}^k) = \left(\frac{1}{np_{n,\text{av}}(1 - p_{n,\text{av}})} \right)^{\frac{k}{2}} \sum_w [X_w] \, . \tag{4.4.2}$$

Here any w is an ordered tuple of indices (not necessarily distinct) $(i_0, \ldots, i_k)$ with $i_0 = i_k$ where $i_j \in \{1, 2, \ldots, n\}$ for $0 \le j \le k$ and we define

$$X_w := \prod_{j=0}^{k-1} \left(x_{i_j, i_{j+1}} - p_{n,\text{av}} \right) .$$

As we shall formally define in Section 4.5.1 below, these w's are called closed word of length $k+1$.

We at first prove when k is odd, most of the random variables X_w have mean 0. As a consequence, one doesn't need a centering for $\mathsf{Tr}(A_{\text{cen1}}^k)$ when k is odd. This is not the case for even k though. The next step is to prove $\lim_{n\to\infty} \mathrm{E}[R_{n,1} \ldots R_{n,m}]$ exists for any fixed m where $R_{n,i} \in \{\mathsf{Tr}(A_{\text{cen1}}^{2k_1+1}), \ldots, \mathsf{Tr}(A_{\text{cen1}}^{2k_l+1})\}$ and to prove the limit $\lim_{n\to\infty} \mathrm{E}\left[R_{n,1} \ldots R_{n,m}\right]$

satisfy the Wick's formula (4.4.1). To this end, observe that

$$\begin{aligned}\mathrm{E}\left[\mathsf{Tr}(A_{\mathrm{cen1}}^{l_1})\dots\mathsf{Tr}(A_{\mathrm{cen1}}^{l_m})\right] &= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{\sum l_i}{2}} \sum_{w_1,\dots,w_i} \mathrm{E}\left[X_{w_1}\dots X_{w_m}\right] \\ &= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{\sum l_i}{2}} \sum_{a} \mathrm{E}\left[X_{w_1}\dots X_{w_m}\right].\end{aligned}$$

Here w_i is a closed word of length l_i+1 and any sentence a is an ordered collection of words $[w_i]_{i=1}^m$. Then we verify that $\mathrm{E}\left[X_{w_1}\dots X_{w_m}\right]=0$ unless the corresponding sentence $a=[w_i]_{i=1}^m$ is a weak CLT sentence Anderson and Zeitouni (2006) (see also Def. 22 in supplement). We then show that among weak CLT sentences,

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{\sum l_i}{2}} \sum_{a} \mathrm{E}\left[|X_{w_1}\dots X_{w_m}|\right] \le \frac{\Psi(l_1,\dots,l_m)}{np_{n,\mathrm{av}}} \to 0$$

as $np_{n,\mathrm{av}}\to\infty$ unless a is a CLT sentence in which all the involved random variables are naturally paired Anderson and Zeitouni (2006) (cf. Def. 22 and Prop. 13 in the supplement). This natural pairing is closely related to the partition η introduced in Lemma 11 which essentially proves the CLT in part (i) of Theorem 7. Here $\Psi(\cdot)$ is an implicit function depending on the values l_i, and we develop a careful upper bound on the number of weak CLT sentences (Lemma 16) to ensure that the convergence to zero in the last display happens whenever $\max_i(l_i)=o(\log(np_n))$. This completes the proof of asymptotic normality under the condition of of the theorem.

The variance formula (4.3.1) is derived using the concepts of Füredi–Komlós sentences and unicyclic graphs introduced in Sections 4.5.2 and 4.5.3. Here the basic idea is similar to that in Anderson and Zeitouni (2006). The major difference is that we shall also develop Proposition 16 and Lemma 21 which enable us to calculate the covariance between the traces and the signed cycles in addition to calculating Σ in (4.3.1).

The proof of part (iii) is completed by calculating the co-variance between the signed cycles and traces and hence showing the variance of the random variable in (4.3.6) goes to 0.

4.4.2. Outline of proof Theorem 8 (i)–(iii)

This proof is based on the second moment argument. All expectation and variance calculation is under $\mathbb{P}_{1,n}$ conditioning on group assignments σ_i for $1 \le i \le n$. The subscript is thus omitted. As before, we focus on odd powers to illustrate the main idea. Observe that when the data is generated from $\mathcal{G}_2(n, p_n, q_n)$, the matrix A_{cen1} is *not* properly centered, i.e., $\mathrm{E}\left[A_{\text{cen1}}\right] \neq 0$. Here we write

$$\begin{aligned}\mathsf{Tr}(A_{\text{cen1}}^k) &= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{k}{2}} \sum_w [X_w] \\ &= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{k}{2}} \sum_w \left[\prod_{j=0}^{k-1} \left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right) + V'_{n,w}\right]\end{aligned}$$

where $w = (i_0, \ldots, i_k)$ is a generic closed word of length $k+1$ and $p_{i_j,i_{j+1}} = \mathrm{E}[x_{i_j,i_{j+1}} | \sigma_{i_j}, \sigma_{i_{j+1}}]$. Here $V'_{n,w}$ is obtained by expanding X_w for any w and considering all the remaining terms apart from $\prod_{j=0}^{k-1} \left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right)$. Using arguments similar to those in the proof of part (i) in Theorem 7 one can prove that the random variable

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{k}{2}} \sum_w \prod_{j=0}^{k-1} \left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right)$$

converges to a Gaussian random variable with mean 0 and variance same as the null case irrespective of the group assignments σ_i, $i = 1, \ldots, n$.

The main task of the proof is then to prove that

$$\mathrm{E}\left[\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{k}{2}} \sum_w V'_{n,w}\right] \to \sum_{r=3:r \text{ odd}}^{k} f(k,r) \frac{k}{r}\, t^r$$

and

$$\mathrm{Var}\left[\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{k}{2}} \sum_w V'_{n,w}\right] \to 0$$

under suitable growth condition of k. The level of technicality here is increased due to the complicated form of the $V'_{n,w}$s, while the key ideas underlying the proof are still Füredi–Komlós sentences and unicyclic graphs.

We mention that the arguments in this particular proof are new and cannot be obtained by modifying the arguments in Anderson and Zeitouni (2006). In particular, we will be able to show that

$$\begin{aligned}
&\mathrm{E}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{k}{2}}\sum_{w}V'_{n,w}\right]\\
&\quad=(1+o(1))\sum_{r=3:r\text{ odd}}^{k}f(k,r)\frac{k}{r}\,t^r+O\left(\frac{2^{C_1k}\mathrm{poly}(k)}{np_{n,\mathrm{av}}}+\frac{(C_2k)^{C_3k}}{n}\right)
\end{aligned}$$

and

$$\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{k}{2}}\sum_{w}V'_{n,w}\right]=O\left(\frac{(C_4k)^{C_5k}}{n}\right).$$

Here C_i's are positive numeric constants and $\mathrm{poly}(k)$ is a known polynomial of k. Note that when $k=o(\min(\log(np_{n,\mathrm{av}}),\sqrt{\log n}))$, $\frac{2^{Ck}\mathrm{poly}(k)}{np_{n,\mathrm{av}}}\to 0$ for any fixed C and $\frac{(Ck)^{Dk}}{n}\to 0$ for any fixed C and D. This gives part (i) of Theorem 8. The proofs of parts (ii) and (iii) here rely on similar ideas to those used in the proofs of their counterparts in Theorem 7.

4.4.3. Outline of proof for Theorem 7 (iv) and Theorem 8 (iv)

We focus on the null case and the proof for the alternative is similar. All the expectation and variance taken below are with respect to $\mathbb{P}_{0,n}$. Recall that when A_{cen2} is considered, the matrix is centered by the sample estimate $\widehat{p}_{n,\mathrm{av}}$ instead of the actual parameter $p_{n,\mathrm{av}}$. Here we write

$$\begin{aligned}
\mathsf{Tr}(A^k_{\mathrm{cen2}})&=\left(\frac{1}{n\widehat{p}_{n,\mathrm{av}}(1-\widehat{p}_{n,\mathrm{av}})}\right)^{\frac{k}{2}}\sum_{w}[X_w]\\
&=\left(\frac{1}{n\widehat{p}_{n,\mathrm{av}}(1-\widehat{p}_{n,\mathrm{av}})}\right)^{\frac{k}{2}}\sum_{w}\left[\prod_{j=0}^{k-1}\left(x_{i_j,i_{j+1}}-p_{n,\mathrm{av}}\right)+E_{n,w}\right].
\end{aligned}$$

Observe that $\text{Var}[\widehat{p}_{n,\text{av}}] = O(p_{n,\text{av}}/n^2)$. As a consequence, one might expect that $|\widehat{p}_{n,\text{av}} - p_{n,\text{av}}| \leq \sqrt{p_{n,\text{av}}}/n^{\delta}$ for some $\delta \in (\frac{1}{2}, 1)$ with very high probability. We call the indicator random variable corresponding to this high probability event Ev. When Ev, $|p_{n,\text{av}} - \widehat{p}_{n,\text{av}}| \ll (p_n - q_n) = O\left(\frac{\sqrt{p_{n,\text{av}}}}{\sqrt{n}}\right)$. We do the analysis of

$$\mathrm{E}\left[\text{Ev}\left(\left(\frac{1}{n\widehat{p}_{n,\text{av}}(1-\widehat{p}_{n,\text{av}})}\right)^{\frac{k}{2}} \sum_{w} E_{n,w}\right)^2\right].$$

Note that $\text{Ev}(p_n - \widehat{p}_{n,\text{av}})$ is a random variable, not a constant like $(p_n - q_n)$. So, many random variables that had mean 0 in the proof of part (i) of Theorem 8 no longer have mean 0 due to dependence. To tackle the additional dependence, a more careful combinatorial analysis will be carried out and we obtain

$$\begin{aligned}
&\mathrm{E}\left[\text{Ev}\left(\left(\frac{1}{n\widehat{p}_{n,\text{av}}(1-\widehat{p}_{n,\text{av}})}\right)^{\frac{k}{2}} \sum_{w} E_{n,w}\right)^2\right] \\
&\quad \leq (C_1 k)^{C_2 k} \frac{1}{\sqrt{n}} + n^{C_3 k} \exp(-n^{C_4}) + \frac{(C_5 k)^{C_6 k}}{n^{\delta - \frac{1}{2}}} \to 0.
\end{aligned}$$

Here the C_i's are positive numeric constants. This completes the proof.

4.4.4. Outline of proof for Theorem 9

We start with the random variable

$$\begin{aligned}
&\mathsf{Tr}(A_{\text{cen1}}^{2k}) - k\psi_{2k}\,\mathsf{Tr}\left(A_{\text{cen1}}^2\right) \\
&\quad = \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w} X_w - k\psi_{2k}\,\mathsf{Tr}\left(A_{\text{cen1}}^2\right).
\end{aligned} \tag{4.4.3}$$

In this case, we break the collection of words in (4.4.3) into four subgroups as follows

$$\begin{aligned}
\mathsf{Tr}(A_{\text{cen1}}^{2k}) &= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w} X_w \\
&= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \left[\sum_{w \in \mathcal{W}_1} X_w + \sum_{w \in \mathcal{W}_2} X_w + \sum_{w \in \mathcal{W}_3} X_w + \sum_{w \in \mathcal{W}_4} X_w\right].
\end{aligned}$$

Here $\mathcal{W}_1$ corresponds to the set of Wigner words, $\mathcal{W}_2$ stands for the set of all weak Wigner words (cf. Def. 20), $\mathcal{W}_3$ is $\cup_r \mathfrak{W}_{2k+1,r,k+r/2}$ which collects all unicyclic graphs (cf. Prop. 16) and $\mathcal{W}_4$ is the complement of $\mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3$. Using Lemma 14 in the supplement, one can ignore the class $\mathcal{W}_4$.

We first show that under both $\mathbb{P}_{0,n}$ and local $\mathbb{P}_{1,n}$ (conditioning on group assignment σ_i for $1 \le i \le n$),

$$\mathrm{E}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w\right] = n\psi_{2k} - \binom{k+1}{2}\psi_{2k} + o(1)$$

and

$$\mathrm{Cov}\left(\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w, \mathsf{Tr}\left(A_{\mathrm{cen}1}^2\right)\right) = 2k\psi_{2k}\frac{\mathrm{Var}\left[(x_{1,2}-\mathrm{E}[x_{1,2}])^2\right]}{p_{n,\mathrm{av}}^2(1-p_{n,\mathrm{av}})^2} + o(1).$$

On the other hand it can be shown that

$$\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w\right] = 2k^2\psi_{2k}^2\frac{\mathrm{Var}\left[(x_{1,2}-\mathrm{E}[x_{1,2}])^2\right]}{p_{n,\mathrm{av}}^2(1-p_{n,\mathrm{av}})^2} + O\left(\frac{2^{C_1 k}\mathrm{Poly}(k)}{np_{n,\mathrm{av}}^2}\right)$$

and

$$\mathrm{Var}\left[\mathsf{Tr}[A_{\mathrm{cen}1}^2]\right] = 2\frac{\mathrm{Var}\left[(x_{1,2}-\mathrm{E}[x_{1,2}])^2\right]}{p_{n,\mathrm{av}}^2(1-p_{n,\mathrm{av}})^2}.$$

Hence, $\mathrm{Var}((\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})})^k \sum_{w\in\mathcal{W}_1} X_w - k\psi_{2k}\,\mathsf{Tr}(A_{\mathrm{cen}1}^2)) \to 0$. Next, arguments similar to the proof of Theorem 8 will show that

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_3} X_w - \sum_{r=4:r\ \text{even}} f(2k,r)\frac{2k}{r}C_{n,r}(G) \xrightarrow{p} 0.$$

So our final focus is on the random variable

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2} X_w = T_{2k} + o_p(1).$$

We again break $\mathcal{W}_2$ into two further groups depending on whether the graph G_w corresponding to a word w is a tree or not. It can be proved that under both $\mathbb{P}_{0,n}$ and local $\mathbb{P}_{1,n}$

$$\mathrm{E}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2:G_w\neq\mathrm{tree}} X_w\right] = (1+o(1))\alpha_{1,2k} + o(1),$$

and

$$\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2:G_w\neq\mathrm{tree}} X_w\right] \to 0.$$

Finally, to get the leading terms in the expectation and variance expressions, among the words $w \in \mathcal{W}_2$ we only need to focus on those where G_w is a tree with k or $k-1$ nodes. Call the collection of these words $\mathcal{W}_{2,1}$ and $\mathcal{W}_{2,2}$, and the corresponding sums $W_{2,1}$ and $W_{2,2}$. It can be shown then under $\mathbb{P}_{0,n}$,

$$\mathrm{E}[W_{2,1}] = (1+o(1))\frac{\alpha_{2,2k}}{p_{n,\mathrm{av}}} + o(1), \quad \mathrm{Var}[W_{2,1}] = (1+o(1))\frac{v_{2k}}{n^2p_{n,\mathrm{av}}^3} + \tilde{\varepsilon}_{2,1},$$
$$\mathrm{E}[W_{2,2}] = (1+o(1))\frac{\alpha_{3,2k}}{np_{n,\mathrm{av}}^2}, \quad \mathrm{Var}[W_{2,2}] = \tilde{\varepsilon}_{2,2},$$

and $\tilde{\varepsilon}_{2,i} \to 0$ for $i = 1, 2$ when $n^2p_{n,\mathrm{av}}^3 \to \infty$ and $k = o(\min(\log(n^2p_{n,\mathrm{av}}^3), \sqrt{\log n}))$. On the other hand, under $\mathbb{P}_{1,n}$ (conditioning on group assignment) by arguments similar to the proof of part (i) of Theorem 8 that for $i = 1, 2$,

$$W_{2,i} \overset{d}{=} \tau_i + \Xi_i$$

where τ_i has the same asymptotic distribution as $W_{2,i}|\mathbb{P}_{0,n}$ and

$$\mathrm{E}[\Xi_i] = O\left(2^{C_1k}\frac{\mathrm{Poly}(k)}{\sqrt{np_{n,\mathrm{av}}}} + (\frac{(C_3k)^{C_4k}}{n})\right) \quad \text{and} \quad \mathrm{Var}[\Xi_i] = \left(\frac{(C_1'k)^{C_2'k}}{n}\right).$$

for some universal constants C_1, C_1', C_1', C_2'. The proof of this step is very similar to that of part (i) of Theorem 8. This completes the proof of Theorem 9.

4.5. Preliminary combinatorics results

Section 4.5.1 is dedicated to building up the preliminary ideas about words, sentences, CLT sentences and state a few important lemmas required in the proofs. In Section 4.5.2 we present the ideas about Füredi–Komlós sentences and related topics. These ideas will be in the center of our proofs. In Section 4.5.3 we study unicyclic graphs. These results play a fundamental role in finding out the exact formula for the covariances of the linear spectral statistics and signed cycles. Most of the definitions and preliminary results in this section can be found in Anderson and Zeitouni (2006) and Anderson et al. (2010), which we include here mainly for the proofs to be self-contained. The new results here are Lemma 16 (which first appeared in Banerjee (2018)), Proposition 16 and Lemma 21. Informed readers may focus on these new results only while skipping the rest of this section.

4.5.1. Words, sentences and their equivalence classes

In this part we give a very brief introduction to words, sentences and their equivalence classes essential for the combinatorial analysis of random matrices. The definitions are taken from Anderson et al. (2010) and Anderson and Zeitouni (2006). For more general information, see (Anderson et al., 2010, Chapter 1) and Anderson and Zeitouni (2006).

Definition 18 ($\mathcal{S}$ words)**.** Given a set $\mathcal{S}$, an $\mathcal{S}$ letter s is simply an element of $\mathcal{S}$. An $\mathcal{S}$ word w is a finite sequence of letters $s_1 \dots s_k$, at least one letter long. An $\mathcal{S}$ word w is *closed* if its first and last letters are the same. In this paper $\mathcal{S} = \{1, \dots, n\}$ where n is the number of nodes in the graph.

Two $\mathcal{S}$ words w_1, w_2 are called *equivalent*, denoted $w_1 \sim w_2$, if there is a bijection on $\mathcal{S}$ that maps one into the other. For any word $w = s_1 \dots s_k$, we use $l(w) = k$ to denote the *length* of w, define the *weight* $wt(w)$ as the number of distinct elements of the set $s_1, \dots, s_k$ and the *support* of w, denoted by $\mathrm{supp}(w)$, as the set of letters appearing in w. With any word w we may associate an undirected graph, with $wt(w)$ vertices and at most $l(w) - 1$ edges, as follows.

Definition 19 (Graph associated with a word)**.** Given a word $w = s_1 \dots s_k$, we let $G_w = (V_w, E_w)$ be the graph with set of vertices $V_w = \mathrm{supp}(w)$ and (undirected) edges $E_w = \{\{s_i, s_{i+1}\}, i = 1, \dots, k-1\}$.

The graph G_w is connected since the word w defines a path connecting all the vertices of G_w, which further starts and terminates at the same vertex if the word is closed. We note that equivalent words generate the same graphs G_w (up to graph isomorphism) and the same passage-counts of the edges. Given an equivalence class $\mathbf{w}$, we shall sometimes denote $\#E_{\mathbf{w}}$ and $\#V_{\mathbf{w}}$ to be the common number of edges and vertices for graphs associated with all the words in this equivalence class $\mathbf{w}$.

Definition 20 (Weak Wigner words)**.** Any word w will be called a *weak Wigner word* if the following conditions are satisfied:

1. w is closed.

2. w visits every edge in G_w at least twice.

Suppose now that w is a weak Wigner word. If $wt(w) = (l(w) + 1)/2$, then we drop the modifier "weak" and call w a *Wigner word*. (Every single letter word is automatically a Wigner word.) Except for single letter words, each edge in a Wigner word is traversed exactly twice. If $wt(w) = (l(w) - 1)/2$, then we call w a *critical weak Wigner word*.

We now move to definitions related to sentences.

Definition 21 (Sentences and corresponding graphs)**.** A sentence $a = [w_i]_{i=1}^m = [[\alpha_{i,j}]_{j=1}^{l(w_i)}]_{i=1}^m$ is an ordered collection of m words of length $(l(w_1), \dots, l(w_m))$ respectively. We define the graph $G_a = (V_a, E_a)$ to be the graph with

$$V_a = \mathrm{supp}(a), \quad E_a = \{\{\alpha_{i,j}, \alpha_{i,j+1}\} | i = 1, \dots, m; j = 1, \dots, l(w_i) - 1\}\} .$$

Definition 22 (Weak CLT sentences)**.** A sentence $a = [w_i]_{i=1}^m$ is called a *weak CLT sentence*, if the following conditions are satisfied:

1. All the words w_i's are closed.

2. Jointly the words w_i visit each edge of G_a at least twice.

3. For each $i \in \{1, \ldots, m\}$, there is another $j \neq i \in \{1, \ldots, m\}$ such that G_{w_i} and G_{w_j} have at least one edge in common.

Suppose now that a is a weak CLT sentence. If $wt(a) = \sum_{i=1}^{m} \frac{l(w_i)-1}{2}$, then we a a *CLT sentence*. If $m = 2$ and a is a CLT sentence, then we call a a *CLT word pair*.

We now introduce an additional notion regarding permutation which will be important in our computations.

Definition 23. Suppose we have a word $w = (\alpha_1, \ldots, \alpha_k)$ of length k and a permutation σ of the set $\{1, \ldots, k\}$, we define w^σ to be the word $(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(k)})$. If σ is a power of the cycle $(123 \ldots k)$, we call σ a cyclic permutation and the corresponding word w^σ to be a cyclic permutation of w.

We now state a few propositions and lemmas which will be used in our proof. These results, except for that of Lemma 16, can be found in Anderson and Zeitouni (2006). We at first state an elementary yet general lemma about a forest G and a word w admitting the interpretation of a walk on G.

Lemma 12 (The parity principle. Lemma 4.4 in Anderson and Zeitouni (2006))**.** *Let G be a forest and e be an edge of G. Let w be a word admitting the interpretation as a walk on G. Let w_* be the unique path in G with initial and terminal vertices coinciding with those of w. Then the word/walk w visits the edge e an odd number of times if and only if w_* visits e.*

The following facts about critical weak Wigner words are important for analyzing trace of even powers and for proving Theorem 9.

Proposition 12 (Proposition 4.8 in Anderson and Zeitouni (2006))**.** *Let w be a critical weak Wigner word and let $G = (V, E) = G_w = (V_w, E_w)$. Then the following hold:*

1. G is connected.

2. Either $\#V - 1 = \#E$ or $\#V = \#E$.

3. If $\#V - 1 = \#E$, then:

a) G is a tree.

b) With exactly one exception w visits each edge of G exactly twice.

c) But w visits the exceptional edge exactly four times.

4. If $\#V = \#E$, then:

a) G is not a tree.

b) w visits each edge of G exactly twice.

The following proposition is crucial for verifying Wick's formula for traces of powers which will in turn prove the CLTs in Theorem 7.

Proposition 13 (Proposition 4.9 in Anderson and Zeitouni (2006))**.** *Let $a = [w_i]_{i=1}^m$ be a weak CLT sentence containing m words. Then we have the following:*

1.

$$wt(a) \leq \sum_{i=1}^{m} \frac{l(w_i) - 1}{2}.$$

2. Suppose the equality holds i.e. a is a CLT sentence. Then the words of the sentence a are perfectly matched in the sense that for all i there exists unique j distinct from i such that w_i and w_j have at least one letter in common. In particular, m is even.

The proof of Proposition 13 in Anderson and Zeitouni (2006) is based on the next important Lemma.

Lemma 13 (Lemma 4.10 in Anderson and Zeitouni (2006))**.** *Let $a = [w_i]_{i=1}^m$ be weak CLT sentence containing m words. Put $G = G_a$. Let k be the number of connected components* *of G. Then*

1. $k \leq [\frac{m}{2}]$.

2. $wt(a) \leq k - m + \left[\frac{\sum_{i=1}^{n} l(w_i)}{2}\right]$.

Here $[y]$ denotes the largest integer less than or equal to y.

Lemma 14 below gives an explicit description of the structure of the CLT word pairs.

Lemma 14 (Proposition 4.12 in Anderson and Zeitouni (2006))**.** *Let $a = [w, x]$ be a CLT word pair and put $G = (V, E) = G_a = (V_a, E_a)$. For any $e \in E$ let $\nu(e, w)$ (respectively $\nu(e, x)$) denote the number of time the edge e is visited by the word w (respectively x). Then the following hold:*

1. *G is connected.*

2. *$\#V - 1 = \#E$ or $\#V = \#E$.*

3. *If $\#V - 1 = \#E$, then:*

   a) *G is a tree.*

   b) *For all $e \in E$, $\nu(e, w)$ and $\nu(e, x)$ are even.*

   c) *There is an unique $e_0 \in E$ such that $\nu(e_0, w) = \nu(e_0, x) = 2$.*

   d) *For all $e \in E \backslash \{e_0\}$, $\nu(e, w) + \nu(e, x) = 2$.*

   e) *Both w and x are Wigner words.*

4. *If $\#V = \#E$, then:*

   a) *G is not a tree.*

   b) *For all $e \in E$ we have $\nu(e, w) + \nu(e, x) = 2$.*

   c) *There is at least one edge $e \in E$ such that $\nu(e, w) = \nu(e, x) = 1$.*

In the next part we shall be able to enumerate all the CLT word pairs explicitly. We end the discussion of this part by Lemma 16 which will crucial for proving the CLTs when the power k slowly diverges to infinity. Its first proof can be found in Banerjee (2018). However, the embedding algorithm used in the proof will also be useful in the proof of Theorem 8. So we spell out the proof of Lemma 16 here to make the present paper self-contained. To begin with, we give an upper bound on the number of equivalence classes corresponding to weak Wigner words.

Lemma 15 (Lemma 2.1.23 in Anderson et al. (2010))**.** *Let $\mathcal{W}_{k,t}$ collect the equivalence classes corresponding to all weak Wigner words w of length $k+1$ with $wt(w)=t$. Then for $k \geq \min(2, 2t-2)$,*

$$\#\mathcal{W}_{k,t} \leq 2^k k^{3(k-2t+2)}.$$

We now state Lemma 16 and its proof.

Lemma 16. *Let $\mathcal{A} = \mathcal{A}^n_{m,t}(l_1, \ldots, l_m)$ be the set of weak CLT sentences $a = [w_i]_{i=1}^m$ such that the letter set is $\{1, \ldots, n\}$, $\#V_a = t$ and $l(w_i) = l_i$ for $i = 1, \ldots, m$. If $l_i \geq 3$ for $i = 1, \ldots, m$, then*

$$\#\mathcal{A} \leq n^t 2^l \left(C_1 l\right)^{C_2 m} l^{3(l-2t)} \tag{4.5.1}$$

where $l = \sum_{i=1}^m l_i$ and $C_1, C_2 > 0$ are numeric constants.

Proof. Let $a = [w_i]_{i=1}^m$ be a weak CLT sentence such that G_a have $\mathcal{C}(a)$ many connected components. At first we introduce a partition $\eta(a)$ in the following way. We put i and j in same block of $\eta(a)$ if G_{w_i} and G_{w_j} share an edge. At first we fix such a partition η and consider all the sentences such that $\eta(a) = \eta$. Let $\mathcal{C}(\eta)$ be the number of blocks in η. It is easy to observe that for any a with $\eta(a) = \eta$, we have $\mathcal{C}(\eta) = \mathcal{C}(a)$. From now on we denote $\mathcal{C}(\eta)$ by $\mathcal{C}$ for convenience.

Let a be any weak CLT sentence such that $\eta(a) = \eta$. We now propose an algorithm to embed a into $\mathcal{C}$ ordered closed words $(W_1, \ldots, W_{\mathcal{C}})$ such that the equivalence class of each W_i belongs to $\mathcal{W}_{L_i,t_i}$ for some numbers L_i and t_i.

A similar type of argument can be found in Claim 3 of the proof of Theorem 2.2 in Banerjee and Bose(2017) Banerjee and Bose (2016).

An embedding algorithm: Let $B_1, \ldots, B_{\mathcal{C}}$ be the blocks of the partition η ordered in the following way. Let $m_i = \min\{j : j \in B_i\}$ and we order the blocks B_i such that $m_1 < m_2 \ldots < m_{\mathcal{C}}$. Given a partition η this ordering is unique. Let

$$B_i = \{i(1) < i(2) < \ldots < i(l(B_i))\}.$$

Here $l(B_i)$ denotes the number of elements in B_i.

For each B_i we embed the sentence $a_i = [w_{i(j)}]_{1 \le j \le l(B_i)}$ into W_i sequentially in the following manner.

1. Let $S_1 = \{i(1)\}$ and $\mathfrak{w}_1 = w_{i(1)}$.
2. For each $1 \le c \le l(B_i) - 1$ we perform the following.
   - Consider $\mathfrak{w}_c = (\alpha_{1,c}, \ldots, \alpha_{l(\mathfrak{w}_c),c})$ and $S_c \subset B_i$. Let $ne \in B_i \backslash S_c$ be the minimum index such that the following two conditions hold.
     (a) $G_{\mathfrak{w}_c}$ and $G_{w_{ne}}$ shares at least one edge $e = \{\alpha_{\kappa_1,c}, \alpha_{\kappa_1+1,c}\}$.
     (b) κ_1 is minimum among all such choices.
   - Let $w_{ne} = (\beta_{1,c}, \ldots, \beta_{l(w_{ne}),c})$ and $\{\beta_{\kappa_2,c}, \beta_{\kappa_2+1,c}\}$ be the first time e appears in w_{ne}. As $\{\beta_{\kappa_2,c}, \beta_{\kappa_2+1,c}\} = \{\alpha_{\kappa_1,c}, \alpha_{\kappa_1+1,c}\}$, $\alpha_{\kappa_1,c}$ is either equal to $\beta_{\kappa_2,c}$ or $\beta_{\kappa_2,c}$. Let $\kappa_3 \in \{\kappa_2, \kappa_2 + 1\}$ such that $\alpha_{\kappa_1,c} = \beta_{\kappa_3,c}$. If $\beta_{\kappa_2,c} = \beta_{\kappa_2+1,c}$, then we simply take $\kappa_3 = \kappa_2$.
   - We now generate $\mathfrak{w}_{c+1}$ in the following way

$$\mathfrak{w}_{c+1} = (\alpha_{1,c}, \ldots, \alpha_{\kappa_1,c}, \beta_{\kappa_3+1,c}, \ldots, \beta_{l(w_{ne}),c}, \beta_{2,c}, \ldots, \beta_{\kappa_3,c}, \alpha_{\kappa_1+1,c}, \ldots, \alpha_{l(\mathfrak{w}_c),c}).$$

Let $\tilde{a}_c := (\mathfrak{w}_c, w_{ne})$. It is easy to observe by induction that all $\mathfrak{w}_c$'s are closed words and so are all the w_{ne}'s. Also all the edges in the graph $G_{\tilde{a}_c}$ are preserved along with their passage counts in $G_{\mathfrak{w}_{c+1}}$.

- Generate $S_{c+1} = S_c \cup \{ne\}$.

3. Return $W_i = \mathfrak{w}_{l(B_i)}$.

In the preceding algorithm we have actually defined a function f which maps any weak CLT sentence a into $\mathcal{C}$ ordered closed words $(W_1, \ldots, W_{\mathcal{C}})$ such that the equivalence class of each W_i belongs to $\mathcal{W}_{L_i,t_i}$ for some numbers L_i and t_i. Observe that given two words $\mathfrak{w}_1$ and $\mathfrak{w}_2$, application of step 2 gives rise to a closed word $\mathfrak{w}_3$ where $l(\mathfrak{w}_3) = l(\mathfrak{w}_1) + l(\mathfrak{w}_2) - 1$. So

$$\begin{aligned} L_i &= \sum_{j \in B_i} l(w_j) - (l(B_i) - 1) < \sum_{j \in B_i} l(w_j). \\ \Rightarrow L_i + 1 &\le \sum_{j \in B_i} l(w_j) \\ \Rightarrow L_i + 1 - 2t_i &\le \sum_{j \in B_i} l(w_j) - 2t_i. \end{aligned} \tag{4.5.2}$$

Unfortunately f is not an injective map. So given $(W_1, \ldots, W_{\mathcal{C}})$ we find an upper bound to the cardinality of the following set

$$f^{-1}(W_1, \ldots, W_{\mathcal{C}}) := \{a | f(a) = (W_1, \ldots, W_{\mathcal{C}})\}$$

We have argued earlier $\mathcal{C}$ is the number of blocks in η. However, in general $(W_1, \ldots, W_{\mathcal{C}})$ does neither specify the partition η nor the order in which the words are concatenated with in each block B_i of η. So we fix a partition η with $\mathcal{C}$ many blocks and an order of concatenation $\mathcal{O}$. Observe that

$$\mathcal{O} = (\sigma_1(\eta), \ldots, \sigma_{\mathcal{C}}(\eta))$$

where for each i, $\sigma_i(\eta)$ is a permutation of the elements in B_i. Now we give an uniform upper bound to the cardinality of the following set

$$f^{-1}_{\eta,\mathcal{O}}(W_1,\ldots,W_{\mathcal{C}}) := \{a|\eta(a)=\eta \ \ ;\mathcal{O}(a)=\mathcal{O} \ \ \& f(a)=(W_1,\ldots,W_{\mathcal{C}})\}.$$

According to the algorithm any word W_i is formed by recursively applying step 2 to $(\mathfrak{w}_c, w_{ne})$ for $1 \le c \le l(B_i)$. Given a word $\mathfrak{w}_3 = (\alpha_1,\ldots,\alpha_{l(\mathfrak{w}_3)})$, we want to find out the number of two word sentences $(\mathfrak{w}_1,\mathfrak{w}_2)$ such that applying step 2 of the algorithm on $(\mathfrak{w}_1,\mathfrak{w}_2)$ gives $\mathfrak{w}_3$ as an output. This is equivalent to choose three positions $i_1 < i_2 < i_3$ from the set $\{1,\ldots,l(\mathfrak{w}_3)\}$ such that $\alpha_{i_1} = \alpha_{i_3}$. Once these three positions are chosen, $(\mathfrak{w}_1,\mathfrak{w}_2)$ can be constructed uniquely in the following manner

$$\begin{aligned}\mathfrak{w}_1 &= (\alpha_1,\ldots,\alpha_{i_1},\alpha_{i_3+1},\ldots,\alpha_{l(\mathfrak{w}_3)})\\ \mathfrak{w}_2 &= (\alpha_{i_2},\ldots,\alpha_{i_3},\alpha_{i_1+1},\ldots,\alpha_{i_2}).\end{aligned}$$

Total number of choices $i_1 < i_2 < i_3$ is bounded by $l(\mathfrak{w}_3)^3 \le \left(\sum_{i=1}^m l(w_i)\right)^3$. For each block B_i, step 2 of the algorithm has been used $l(B_i)$ many times. So

$$f^{-1}_{\eta,\mathcal{O}}(W_1,\ldots,W_{\mathcal{C}}) \le \left(\sum_{i=1}^m l(w_i)\right)^{3\sum_{i=1}^{\mathcal{C}} l(B_i)} = \left(\sum_{i=1}^m l(w_i)\right)^{3m}.$$

On the other hand, there are at most m^m many η's and for each η there are at most $\prod_{i=1}^{\mathcal{C}} l(B_i)! \le m^m$ choices of $\mathcal{O}$. So

$$f^{-1}(W_1,\ldots,W_{\mathcal{C}}) \le m^{2m}\left(\sum_{i=1}^m l(w_i)\right)^{3m} \le \left(D_1\sum_{i=1}^m l(w_i)\right)^{D_2 m} \tag{4.5.3}$$

for some known constants D_1 and D_2. Now we fix the sequence (L_i, t_i) and find an upper bound to the number of $(W_1,\ldots,W_{\mathcal{C}})$. From Lemma 6 we know the number of choices of W_i is bounded by $2^{L_i-1}(L_i-1)^{L_i-2t_i+1}n^{t_i}$. So the total number of choices for $(W_1,\ldots,W_{\mathcal{C}})$

is bounded by

$$2^{\sum_{i=1}^{\mathcal{C}} L_i} \prod_{i=1}^{\mathcal{C}} (L_i - 1)^{3(L_i - 2t_i + 1)} n^{t_i} \leq 2^{\sum_{i=1}^{m} l(w_i)} n^t \left(\sum_{i=1}^{m} l(w_i) \right)^{3(\sum_{i=1}^{m} l(w_i) - 2t)}. \tag{4.5.4}$$

Now the number of choices (L_i, t_i) such that $\sum_{i=1}^{\mathcal{C}} L_i = \sum_{i=1}^{m} l(w_i) - \sum_{i=1}^{\mathcal{C}} (l(B_i) - 1)$ and $\sum_{i=1}^{\mathcal{C}} t_i = t$ are bounded by

$$\binom{\sum_{i=1}^{m} l(w_i) - \sum_{i=1}^{\mathcal{C}} (l(B_i) - 1) - 1}{\mathcal{C} - 1} \binom{t-1}{\mathcal{C}-1} \leq \binom{\sum_{i=1}^{m} l(w_i) - 1}{\mathcal{C} - 1} \binom{t-1}{\mathcal{C}-1} \leq \left(\sum_{i=1}^{m} l(w_i) \right)^{2m}. \tag{4.5.5}$$

Here the inequality follows since $\mathcal{C} \leq m$ and $t < \sum_{i=1}^{m} \frac{l(w_i) - 1}{2}$. Finally we using the fact that $1 \leq \mathcal{C} \leq m$ and combining (4.5.3), (4.5.4) and (4.5.5) we finally have

$$\begin{aligned} \#\mathcal{A} &\leq \left(D_1 \sum_{i=1}^{m} l(w_i) \right)^{D_2 m} \times 2^{\sum_{i=1}^{m} l(w_i)} n^t \left(\sum_{i=1}^{m} l(w_i) \right)^{3(\sum_{i=1}^{m} l(w_i) - 2t)} \times \left(\sum_{i=1}^{m} l(w_i) \right)^{2m} \\ \Rightarrow \#\mathcal{A} &\leq 2^{\sum_i l(w_i)} \left(C_1 \sum_i l(w_i) \right)^{C_2 m} \left(\sum_i l(w_i) \right)^{3(\sum_i l(w_i) - 2t)} n^t \end{aligned} \tag{4.5.6}$$

as required. □

4.5.2. Füredi–Komlós enumeration

We now introduce the notion of Füredi–Komlós sentences. It was the key idea underlying the proof of Lemma 15 (Lemma 2.1.23 in Anderson et al. (2010)), which in turn was crucial for the proof of Lemma 16. In addition, as we shall show below, it plays an important role in getting the exact enumeration of CLT word pairs and provides some general insight about the covariance structure between different linear spectral statistics. Most of the materials in this Subsection are borrowed from Section 7 of Anderson and Zeitouni (2006) and Chapter 1 of Anderson et al. (2010). The original idea of Füredi–Komlós sentences dates back to Füredi and Komlós (1981).

Definition 24 (FK sentences)**.** Let $a = [w_i]_{i=1}^m$ be a sentence consisting of m words. We say that a is an *FK sentence* under the following conditions:

1. G_a is a tree.

2. Jointly the words/walks w_i, $i = 1, \ldots, m$, visit no edge of G_a more than twice.

3. For $i = 1, \ldots, m-1$, the first letter of w_{i+1} belongs to $\cup_{j=1}^i \text{supp}(w_j)$.

We say that a is an *FK word* if $m = 1$.

By definition, any word admitting interpretation as a walk on a forest visiting no edge of the forest more than twice is automatically an FK word. The constituent words of an FK sentence are FK words. If an FK sentence is at least two words long, then the result of dropping the last word is again an FK sentence. If the last word of an FK sentence is at least two letters long, then the result of dropping the last letter of the last word is again an FK sentence.

Definition 25 (The stem of an FK sentence)**.** Given an FK sentence $a = [w_i]_{i=1}^m$, we define $G_a^1 = (V_a^1, E_a^1)$ to be the subgraph of $G_a = (V_a, E_a)$ with $V_a^1 = V_a$ and E_a^1 equal to the set of edges $e \in E_a$ such that the words/walks w_i, $i = 1, \ldots, m$, jointly visit e exactly once.

The following lemma characterizes the exact structure of an FK word.

Lemma 17 (Lemma 2.1.24 in Anderson et al. (2010))**.** *Suppose w is an FK word. Then there is exactly one way to write $w = w_1, \ldots, w_r$ where each w_i is a closed Wigner word and they are pairwise disjoint.*

Let α_i be the first letter of the word w_i, we declare the word $\alpha_1, \ldots, \alpha_r$ to be the *acronym* of the word w in Lemma 17. Since the counts of Wigner words are well known, one can explicitly enumerate the equivalence classes of all FK words as follows.

Proposition 14. *Let $F(m, r)$ be the set of equivalence classes of all FK words of length m*

with acronym of size $r \leq m$. Then

$$\#F(m, r) = f(m, r)$$

and $\sum_{r=1}^{m} f(m, r) \leq 2^{m-1}$. Here $f(m, r)$ is as defined in (4.2.5).

Proof. The proof follows immediately from the proof of Lemma 17 (See the proof of Lemma 2.1.24 in Anderson et al. (2010)). □

FK syllabification Our interest in FK sentences is mainly due to the fact that any word w can be parsed into an FK sentence sequentially. In particular, one declares a new word at each time when not doing so would prevent the sentence formed up to that point from being an FK sentence. Formally, we define the FK sentence corresponding to any given word w in the following way. Suppose the word w is formed by m letters. We declare any edge $e \in E_w$ to be new if $e = \{\alpha_i, \alpha_{i+1}\}$ and $\alpha_{i+1} \notin \{\alpha_1, \ldots, \alpha_i\}$ otherwise we declare e to be old. We now construct the FK sentence w' corresponding to the word w by breaking the word at each position of an old edge and the third and all subsequent positions of a new edge. Observe that any old edge gives rise to a cycle in G_w. As a consequence, by breaking the word at the old edge we remove all the cycles in G_w. On the other hand, all new edges are traversed at most twice as we break at their third and all subsequent occurrences. It is easy to see that the graph $G_{w'}$ remains connected since we are not deleting the first occurrence of a new edge. As a consequence, the graph $G_{w'}$ is a tree where every edge is traversed at most twice. Furthermore, by the definition of old and new edges, the first letter in the second and any subsequent word in w' belongs to the support of all the previous ones. Therefore, the resulting sentence w' is an FK sentence. Note that the FK syllabification preserves equivalence, i.e., if $w \sim x$ then the corresponding FK sentences $w' \sim x'$.

The discussion about FK syllabification shows that all words can be uniquely parsed into an FK sentence. Hence we can use the enumeration of FK sentences to enumerate words of specific structures of interest. The following lemma gives an upper bound to the number of ways one FK sentence b and one FK word c can be concatenated so that the sentence $[b, c]$ is again an FK sentence.

Lemma 18 (Lemma 7.6 in Anderson and Zeitouni (2006))**.** *Let $b = [w_i]_{i=1}^m$ be an FK sentence and c be an FK word such that the first letter in c is in $\mathrm{supp}(b)$. Let $\gamma_1, \ldots, \gamma_r$ be the acronym of c where $\gamma_1 \in \mathrm{supp}(b)$. Let l be the largest index such that $\gamma_l \in \mathrm{supp}(b)$ and write $d = \gamma_1, \ldots, \gamma_l$. Then the sentence $[b, c]$ is an FK sentence if and only if the following conditions are satisfied:*

1. *d is a geodesic in the forest G_b^1.*
2. *$\mathrm{supp}(b) \cap \mathrm{supp}(c) = \mathrm{supp}(d)$.*

Here a geodesic connecting $x, y \in G_b^1$ is a path of minimal length starting at x and terminating at y. Further, there are at most $(wt(b))^2$ equivalence classes $[x_i]_{i=1}^{m+1}$ such that $b \sim [x_i]_{i=1}^m$ and $c \sim x_{m+1}$.

The following two lemmas together give an upper bound on the number of equivalence classes corresponding to closed words via the corresponding FK sentences.

Lemma 19 (Lemma 7.7 in Anderson and Zeitouni (2006))**.** *Let $\Gamma(k, l, m)$ denote the set of equivalence classes of FK sentences $a = [w_i]_{i=1}^m$ consisting of m words such that $\sum_{i=1}^m l(w_i) = l$ and $wt(a) = k$. Then we have*

$$\#\Gamma(k, l, m) \leq 2^{l-m} \binom{l-1}{m-1} k^{2(m-1)}. \tag{4.5.7}$$

Lemma 20 (Lemma 7.8 in Anderson and Zeitouni (2006))**.** *For any FK sentence $a = [w_i]_{i=1}^m$, we have*

$$m = \#E_a^1 - 2wt(a) + 2 + \sum_{i=1}^m l(w_i). \tag{4.5.8}$$

4.5.3. Unicyclic graphs, CLT word pairs and their enumeration

As discussed earlier, enumeration of the CLT word pairs gives us the exact variance expression in Theorem 7. In this part we shall present a few more concepts and finally give an explicit enumeration of the CLT word pairs based on Füredi–Komlós enumeration. Most of the concepts in this part can be found in Section 8 in Anderson and Zeitouni (2006).

Definition 26 (Bracelets)**.** We say a graph $G = (V, E)$ is a *bracelet* if there is an enumeration $\alpha_1, \ldots, \alpha_r$ of V such that

$$E = \begin{cases} \{\{\alpha_1, \alpha_1\}\} & \text{if } r = 1 \\ \{\{\alpha_1, \alpha_2\}\} & \text{if } r = 2 \\ \{\{\alpha_1, \alpha_2\}, \{\alpha_2, \alpha_3\}, \ldots, \{\alpha_{r-1}, \alpha_r\}, \{\alpha_r, \alpha_1\}\} & \text{if } r \geq 3. \end{cases} \tag{4.5.9}$$

We call r the *circuit length* of the bracelet G.

In this paper we shall ignore the bracelets corresponding $r = 1$ since they correspond to the diagonal elements of A_{ceni}, $i = 1, 2$, which are all zeros.

Definition 27 (Unicyclic graphs)**.** A graph $G = (V, E)$ is called *unicyclic* if $\#V = \#E$.

In particular any bracelet of length $\neq 2$ is a unicyclic graph, while a bracelet of length 2 is a tree. The following proposition describes the structure of unicyclic graphs.

Proposition 15 (Proposition 8.2 in Anderson and Zeitouni (2006))**.** *Let $G = (V, E)$ be a unicyclic graph. For each edge $e \in E$ put $G \backslash e = (V, E \backslash \{e\})$. Let Z be the subgraph of G consisting of all $e \in E$ such that $G \backslash e$ is connected, along with all attached vertices. Let r be the number of edges of Z. Let F be the graph obtained from G by deleting all edges of Z. The following statements hold:*

1. *F is a forest with exactly r connected components.*
2. *If G has a degenerate edge, then $r = 1$.*
3. *If G has no degenerate edge, then $r \geq 3$.*

4. Z meets each connected component of F in exactly one vertex.

5. Z is a bracelet of circuit length r.

6. For all $e \in E$ the following conditions are equivalent:

(a) $G\backslash e$ is connected.

(b) $G\backslash e$ is a tree.

(c) $G\backslash e$ is a forest.

We call Z the bracelet of G. We call r the circuit length of G, and each of the connected components of F a pendant tree.

We shall see in the proof of Theorem 7 that we need to consider closed words w such that G_w is unicyclic. We call such words *uniwords*. We now provide an explicit way to enumerate uniwords. This part of the proof deviates from Anderson and Zeitouni (2006). The argument presented here gives us a unified way to calculate the covariances in (4.3.1) and (4.3.2), and to calculate the covariance between the signed cycles and the LSSs.

Proposition 16 (Enumeration of uniwords)**.** *Let $\mathfrak{W}_{m+1,r,t}$ denote the set of all closed words with letters taken from $\{1, \dots, n\}$, such that for any $w \in \mathfrak{W}_{m+1,r,t}$, $l(w) = m+1$, $wt(w) = t$ so that $2t - r = m$ and G_w is a unicyclic graph with circuit length $r \geq 3$. Then for any $m = o(\sqrt{n})$,*

$$\frac{\#\mathfrak{W}_{m+1,r,t}}{n^t} = (1 + o(1))\frac{mf(m, r)}{r}. \tag{4.5.10}$$

Here $f(m, r)$ is as defined in (4.2.5) which by Proposition 14 is the number of equivalence classes of FK words of length m with acronyms of length r.

Proof. The proof will be done by creating a multivalued map χ from the set of FK words of length m having acronym of length r to $\mathfrak{W}_{m+1,r,t}$. We shall enumerate exactly the cardinality of the forward and inverse image of every element. In this proof every FK word will be denoted by w_{FK} and any closed word will be denoted by w to make the distinction.

First start with any FK word w_{FK} of length m having acronym of length r. We at first construct the "base" $w_{\text{FK}}^B \in \mathfrak{W}_{m+1,r,t}$ of w_{FK}. Let $(\alpha_1, \ldots, \alpha_r)$ be the acronym of w_{FK}. From the proof of Lemma 17, it is easy to see that the first and the last letter of w_{FK} is given by α_1 and α_r respectively. We take

$$w_{\text{FK}}^B = (w_{\text{FK}}, \alpha_1). \tag{4.5.11}$$

Observe that w_{FK}^B is a closed word of length $m+1$ and it has a bracelet of length r formed by the acronym and removing the bracelet we are left with a forest where each edge in the forest has been traversed exactly twice by w_{FK}^B. So the graph $G_{w_{\text{FK}}^B}$ is unicyclic hence w_{FK}^B is a uinword. It is also easy to check that $2t - r = m$. So $w_{\text{FK}}^B \in \mathfrak{W}_{m+1,r,t}$.

Before going into the construction of the map χ, we introduce a useful notation. Let $w = [\alpha_i]_{i=1}^{l(w)}$ be a closed word. We denote $\breve{w}$ to be the word dropping the last letter of w. Now we construct the multivalued map χ as follows:

$$\chi(w_{\text{FK}}) = \{w \in \mathfrak{W}_{m+1,r,t} |\ \exists\ \sigma \text{ so that } \breve{w}^{\sigma} \alpha_{\sigma(1)} = w_{\text{FK}}^B\} \tag{4.5.12}$$

Here σ is a cyclic permutation of $\{1, \ldots, m\}$ and for any word w we have defined w^σ in Definition 23. As there are m cyclic permutations σ of $\{1, \ldots, m\}$, $\#\chi(w_{\text{FK}}) = m$.

Now we shall prove that for any given $w = [\alpha_i]_{i=1}^{m+1} \in \mathfrak{W}_{m+1,r,t}$, $\#\chi^{-1}(w) = r$. We start with any word $w \in \mathfrak{W}_{m+1,r,t}$ and drop the last letter to get $\breve{w}$. Now consider the bracelet Z_w in G_w. There are r many vertices in Z_w from the assumption and let the corresponding letters be $\{\beta_1, \ldots, \beta_r\}$. The set $\{\beta_1, \ldots, \beta_r\}$ has cardinality r from the definition of Z_w. As a consequence, the letters β_i, $1 \le i \le r$ are distinct. Now consider the cyclic permutation σ_i, $1 \le i \le r$ such that $\sigma_i(1) = \beta_i$ and $\sigma_i(2)$ is a vertex in the pendent tree meeting Z_w at position β_i if it is not empty. However if the pendent tree meeting Z_w at position β_i is empty we take σ_i to be such that $\sigma_i(1) = \beta_i$ and $\sigma_i(2) \in \{\beta_1, \ldots, \beta_r\}$. From the condition $2t - r = m$ it follows that every edge in Z_w has been traversed exactly once and every edge in the forest F_w has been traversed exactly twice by the word w_i. As a consequence, there are exactly r such permutations $\{\sigma_i\}_{1\le i\le r}$. Consider the closed word $w_i = \check{w}^{\sigma_i}\alpha_{\sigma_i(1)}$. Let $\beta^{(i,j)}$ be the j th appearing letter of $\{\beta_1,\ldots,\beta_r\}$ in the permutation σ_i. Observe that $\beta^{(i,1)} = \beta_i$. The construction of σ_i compels the word w_i to be of the following form

$$w_i = w^{(i,1)}w^{(i,2)}\ldots w^{(i,r)}\beta_i.$$

Here $w^{(i,j)}$, $1\le j\le r$ is a Wigner word corresponding to the pendent tree meeting Z_w at position $\beta^{(i,j)}$. Now given any such w_i consider the unique FK word

$$W_{\mathrm{FK},i} = w^{(i,1)}w^{(i,2)}\ldots w^{(i,r)}.$$

All these words $W_{\mathrm{FK},i}$'s are distinct since their starting points are distinct. As a consequence, $\#\chi^{-1}(w) = r$.

There are $f(m,r)$ many equivalence classes corresponding to FK words of length m having acronym of length r. So the total number of words corresponding to this class is

$$n(n-1)\ldots(n-t+1)f(m,r).$$

Observing $t\le m$, we get

$$\frac{\#\mathfrak{W}_{m+1,r,t}}{n^t} = (1+o(1))\frac{mf(m,r)}{r}$$

for all $m = o(\sqrt{n})$ as declared. □

Now we state one more property about the CLT word pairs. Its proof is straightforward and follows from the discussion on p.32 of Anderson et al. (2010). We omit the details.

Proposition 17. *Let $a = [w,x]$ be a CLT word pair such that G_a is not a tree. Then*

1. *The graphs G_w and G_x are both unicyclic.*

2. *They have common bracelet Z.*

3. *Every edge in the bracelet Z is traversed exactly once by both w and x.*

4. *Let F_w and F_x be the forests corresponding to G_w and G_x. Then the common vertices between F_w and F_x are subset of Z. In other words, F_w and F_x can't have any common edge.*

5. *Each edge in F_w is traversed exactly twice by the word w and each edge it F_x is traversed exactly twice by the word x.*

Our last result fixes a uniword w and calculates the number of words x such that $a = [w, x]$ is a CLT word pair.

Lemma 21 (CLT word pairing)**.** *Fix a uniword $w \in \mathfrak{W}_{m+1,r,t}$. Let $S_{m',t'}(w)$ be the set of words x such that $x \in \mathfrak{W}_{m'+1,r,t'}$ and $a = [w, x]$ is a CLT word pair. Then for all $m' = o(\sqrt{n})$,*

$$\frac{\#S_{m',t'}(w)}{n^{t'-r}} = (1 + o(1))\, 2m' f(m', r).$$

Proof. Before going into proof at first we observe that as $2t' - r = m'$, given r and fixing m', t' automatically fixed.

Consider the graph G_w with bracelet Z_w. The word w admits a walk on the edges of G_w. Let $(\beta_1, \dots, \beta_r)$ be the vertices in Z_w *ordered* according to their exploration by the walk corresponding to w. We consider all FK words of length m' and weight t' w_{FK} such that $V_w \cap V_{w_{\text{FK}}} = V_{Z_w}$. Here V_{Z_w} is the set of vertices of the graph Z_w. We denote this set of FK words by $F_{m',t'}(w)$. There are total $f(m', r)$ *equivalence classes* of such words. Since we have fixed the acronym and the word w, the number of possible choices of such w_{FK} is given by

$$\#F_{m',t'}(w) = f(m', r)(n - t) \dots (n - t - t' + r + 1) = f(m', r)(1 + o(1))n^{t'-r}.$$

For any FK word w_{FK} let us recall its base w_{FK}^B in $\mathfrak{W}_{m'+1,r,t'}$ from (4.5.11). Let $w = [\alpha_i]_{i=1}^{l(w)}$

be a closed word. We denote $\breve{w}$ to be the word dropping the last letter of w. Now construct the set $S_{m',t'}(w)$ as follows

$$S_{m',t'}(w) = \cup_{w_{\mathrm{FK}\in F_{m',t'}(w)}}\{x \in \mathfrak{W}_{m'+1,r,t'} \mid \exists\, \sigma \text{ so that } \breve{w}^{\sigma}\alpha_{\sigma(1)} = w_{\mathrm{FK}}^{B}\}.$$

Here σ is either a cyclic permutation of $\{1,\ldots,m\}$ or its mirror image. It is easy to observe that there are $2m'$ such σ's.

We now prove that there is no over-counting in $S_{m',t'}$. This trivially follows from the proof of Proposition 16. Since the distinct FK words in the inverse image $\chi^{-1}(x)$ of any word $x \in S_{m',t'}(w)$ will have acronyms such that one is a non-trivial cyclic permutation of other. However we only considered FK words corresponding to fixed acronym $(\beta_1,\ldots,\beta_r)$. Now observe that these are the only possible choices of x so that $[w,x]$ is a CLT word pair. Hence the result is proved. □

4.6. Proofs of main results

We shall first prove part (i)–(iii) of Theorem 7. Then we shall prove part (i)–(iii) of Theorem 8 and finally we shall come back to prove part (iv) of Theorem 7 and part (iv) of Theorem 8. The proof of all the subsequent results are given after the completion of the proofs of Theorem 7 and Theorem 8. The proofs of Theorem 11 and Proposition **??** are omitted as they follow directly from Theorems 8–10. Before we proceed, we quote the following result on the joint asymptotic normality of signed cycles under both the null and local alternatives, which will be used repeatedly in the rest of this section. Throughout the rest of this section, we focus on the assortative case of $p_n > q_n$ when proving results under local alternatives, and the proofs are essentially the same for the disassortative case of $p_n < q_n$ due to the second part of the following proposition.

Proposition 4 (Banerjee (2018))**.** *Suppose that as $n \to \infty$, $np_{n,\mathrm{av}} \to \infty$ and c and t are constants. Then the following results hold:*

(i) Under $P_{0,n}$*, for any* $3 \le k_1 < \ldots < k_l = o(\log(np_{n,\mathrm{av}}))$*,*

$$\left(\frac{C_{n,k_1}(G)}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l}(G)}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l). \tag{4.6.1}$$

(ii) Under $P_{1,n}$*, for any* $3 \le k_1 < \ldots < k_l = o(\min(\log(np_{n,\mathrm{av}}), \sqrt{\log(n)}))$*, if* $p_n > q_n$*,*

$$\left(\frac{C_{n,k_1}(G) - \mu_1}{\sqrt{2k_1}}, \ldots, \frac{C_{n,k_l}(G) - \mu_l}{\sqrt{2k_l}}\right) \xrightarrow{d} N_l(0, I_l) \tag{4.6.2}$$

where $\mu_i = t^{k_i}$ *for* $1 \le i \le l$*. If* $p_n < q_n$*, the conclusion holds with* $\mu_i = (-t)^i$ *for all* i*.*

4.6.1. Proof of parts (i)–(iii) of Theorem 7

Throughout this subsection, all expectation and variance are taken under $\mathbb{P}_{0,n}$.

Proof of part (i)

We start with a generic $k = o(\log(np_{n,\mathrm{av}}))$. Observe that

$$\mathsf{Tr}(A_{\mathrm{cen1}}^{2k+1}) = \left(\frac{1}{np_{n,\mathrm{av}}(1 - p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \sum_{w: l(w) = 2k+2 \text{ \& } w \text{ closed}} [X_w]. \tag{4.6.3}$$

Here any word w is an ordered pair $(i_0, \ldots, i_{2k+1})$ where the numbers $i_j \in \{1, 2, \ldots, n\}$ for $0 \le j \le 2k+1$, $i_0 = i_{2k+1}$, and we define $X_w := \prod_{j=0}^{2k} \left(x_{i_j, i_{j+1}} - p_{n,\mathrm{av}}\right)$. The proof is divided into following two steps: (1) we figure out a subset of words in the summation which matters for the asymptotic distribution; (2) we apply the method of moments spelled out in Section 4.4 to summation over that subset.

Step 1: At first we prove that the random variable $\mathsf{Tr}(A_{\mathrm{cen1}}^{2k+1})$ does not require any additional centering. Observe that if $\mathrm{E}[X_w] \neq 0$, all the edges in $G_w = (V_w, E_w)$ have been traversed at least twice. Since $l(w) = 2k+2$, this will imply that

$$\#E_w \le \frac{2k+1}{2} \quad \Rightarrow \quad \#E_w \le k. \tag{4.6.4}$$

On the other hand from Lemma 12, G_w cannot be a tree as the total number of edge traversals on G_w by the word w is odd. So $\#E_w \geq \#V_w$. This forces $\#V_w \leq k$. Denote the set of such words by No_k. We shall prove that the contribution of these words is negligible. In particular,

$$\begin{aligned}
&\mathrm{E}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}}\sum_{w\in\mathrm{No}_k}X_w\right]^2\\
&=\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}\sum_{w,w'\in\mathrm{No}_k}\mathrm{E}[X_wX_{w'}] \\
&\leq\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}\sum_{a=[w,w']:w,w'\in\mathrm{No}_k}p_{n,\mathrm{av}}^{\#E_a}.
\end{aligned}\tag{4.6.5}$$

Here a is the two word sentence obtained by concatenating w and w'. Now there can be two cases.

Case 1: The words w_1 and w_2 share an edge. In this case, a is a weak CLT sentence. So we can apply Lemma 16 with $m=2$ and $l_1=l_2=2k+2$ to get that the sum in this case is bounded by

$$\begin{aligned}
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}\sum_{\zeta=1}^{2k}n^{\zeta}p_{n,\mathrm{av}}^{\zeta}2^{4k+4}(C_1k)^{2C_2}(4k+4)^{6(2k+2-\zeta)}\\
&\leq\left(\frac{1}{(1-p_{n,\mathrm{av}})}\right)^{2k+1}2^{4k+4}(C_1k)^{2C_2}(4k+4)^6\sum_{\zeta=1}^{2k}\left(\frac{(4k+4)^6}{np_{n,\mathrm{av}}}\right)^{2k+1-\zeta}
\end{aligned}\tag{4.6.6}$$

where C_1 and C_2 are known constants.

Observe that $2k+1-\zeta>1$. As a consequence, the R.S. of (4.6.6) is a geometric sum on $\left(\left(\frac{(4k+4)^6}{np_{n,\mathrm{av}}}\right)\right)^i$ with lowest index being 1. We also have $\left(\frac{(4k+4)^6}{np_{n,\mathrm{av}}}\right)\to 0$ by the assumption $k=o(\log(np_{n,\mathrm{av}}))$. As a consequence, the R.S. of (4.6.6) can be bounded by

$$\left(\frac{1}{(1-p_{n,\mathrm{av}})}\right)^{2k+1}2^{4k+4}(C_1k)^{2C_2}(4k+4)^6C_3\left(\frac{(4k+4)^6}{np_{n,\mathrm{av}}}\right).\tag{4.6.7}$$

Here C_3 is another known constant. It is easy to see (4.6.7) goes to zero when $k = o(\log(np_{n,\text{av}}))$.

Case 2: The words w_1 and w_2 don't share an edge. Let $wt(w_1) = \zeta_1$ and $wt(w_2) = \zeta_2$. We shall apply Lemma 15 in this case. Since both ζ_1 and ζ_2 are less than or equal to k. The equation $2k + 1 > 2\zeta - 2$ is trivially satisfied. Now from Lemma 15 a crude upper bound to the number of sentences $a = [w_1, w_2]$ such that w_1 and w_2 don't share an edge such that $wt(a) = \zeta$ is given by

$$\begin{aligned}\sum_{\zeta_1} \sum_{\zeta_2=\zeta-\zeta_1} & n^{\zeta} 2^{2k+1} (2k)^{3(2k+1-2\zeta_1+2)} \times 2^{2k+1} (2k)^{3(2k+1-2\zeta_2+2)} \\ &= \sum_{\zeta_1} \sum_{\zeta_2=\zeta-\zeta_1} n^{\zeta} 2^{4k} (2k)^{3(4k-2\zeta+4)} \leq n^{\zeta} \zeta^2 2^{4k+2} (2k)^{6(2k-\zeta+2)}.\end{aligned} \tag{4.6.8}$$

Here the factor ζ^2 comes due to the sum. Consequently, the sum in this case is bounded by

$$\begin{aligned}&\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k+1} \sum_{\zeta=1}^{2k} n^{\zeta} p_{n,\text{av}}^{\zeta} \zeta^2 2^{4k+2} (2k)^{6(2k-\zeta+2)} \\ &\leq \left(\frac{1}{1-p_{n,\text{av}}}\right)^{2k+1} \sum_{\zeta=1}^{2k} 2^{4k+2} k^2 (2k)^6 \left(\frac{(2k)^6}{np_{n,\text{av}}}\right)^{(2k+1-\zeta)}.\end{aligned} \tag{4.6.9}$$

Now (4.6.9) can be analyzed similarly as (4.6.6) to get that (4.6.9) goes to 0 also. This forces the first expression of (4.6.5) to go to 0. As a consequence, we can simply neglect the words in No_k. In particular, any limiting distribution (if exists) of $\mathsf{Tr}(A_{\text{cen1}}^{2k+1})$ is same as the limiting distribution of the following random variable

$$Y_{n,2k+1} = \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} \sum_{w \notin \text{No}_k} X_w. \tag{4.6.10}$$

Step 2: Now we prove the joint asymptotic normality of

$$\Sigma^{-\frac{1}{2}} \big(\mathsf{Tr}(A_{\text{cen1}}^{2k_1+1}), \ldots, \mathsf{Tr}(A_{\text{cen1}}^{2k_l+1}) \big).$$

In particular, we are to prove the following: There exists random variables $Z_1, \dots, Z_l$ such that for any fixed m

$$\lim_{n\to\infty} \mathrm{E}[R_{n,1}\dots R_{n,m}] \to \begin{cases} \sum_\eta \prod_{i=1}^{\frac{m}{2}} \mathrm{E}[Z_{\eta(i,1)}Z_{\eta(i,2)}] & \text{for } m \text{ even} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{4.6.11}$$

Here $R_{n,i} \in \{Y_{n,2k_1+1}, \dots, Y_{n,2k_l+1}\}$ and η is a partition of $\{1, 2\dots, m\}$ into $\frac{m}{2}$ blocks such that each block contains exactly two elements. First observe that (4.6.11) will simultaneously imply part (i) and (ii) of Lemma 10. Implication of (i) is obvious. However, for (ii) one can take $R_{n,i}$'s to be all equal and from Wick's formula (Lemma 11) the limiting distribution of $R_{n,i}$'s is normal. It is well known that normal random variables satisfy Carleman's condition.

Note that

$$\mathrm{E}[R_{n,1}\dots R_{n,m}] = \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{w_1\dots w_m} \mathrm{E}\left[X_{w_1}\dots X_{w_m}\right]. \tag{4.6.12}$$

Here w_i is a closed word with $l_i = l(w_i) = 2k+2$, not belonging to No_k if $R_{n,i} = Y_{n,2k+1}$. We start with any generic $X_{w_1}\dots X_{w_m}$. We at first prove

$$\mathrm{E}\left[X_{w_1}\dots X_{w_m}\right] = 0$$

if the sentence $a = [w_i]_{i=1}^m$ is not a weak CLT sentence. If a is not a weak CLT sentence, then there is at least one edge in G_a which is traversed exactly once by the sentence a. This means there is at least one random variable in the product $X_{w_1}\dots X_{w_m}$ which has appeared exactly once. Since $X_{w_1}\dots X_{w_m}$ is product of independent mean 0 random variable, we have

$$\mathrm{E}\left[X_{w_1}\dots X_{w_m}\right] = 0.$$

Now let $\mathcal{A}_{m,\zeta}$ be the set of weak CLT sentences obtained by concatenating m words such that the ith word has length l_i and $wt(a) = \zeta$ for any $a \in \mathcal{A}_{m,\zeta}$. Proposition 13 leads to

$$wt(a) \leq \sum_{i=1}^{m} \frac{l_i - 1}{2}.$$

As a consequence, we can write (4.6.12) as

$$\mathrm{E}[R_{n,1} \dots R_{n,m}] = \left(\frac{1}{np_{n,\mathrm{av}}(1 - p_{n,\mathrm{av}})} \right)^{\sum_{i=1}^{m} \frac{l_i - 1}{2}} \sum_{\zeta=1}^{\sum_{i=1}^{m} \frac{l_i - 1}{2}} \sum_{a=[w_i]_{i=1}^{m} \in \mathcal{A}_{m,\zeta}} \mathrm{E}\left[X_{w_1} \dots X_{w_m} \right]. \tag{4.6.13}$$

We now show that only CLT sentences matter (those s.t. $\zeta = \sum_{i=1}^{m} \frac{l_i - 1}{2}$) on the right side of the last display asymptotically. To this end, fix any weak CLT sentences $a \in \mathcal{A}_{m,\zeta}$. For any edge $e = \{i, j\}$ in the graph G_a, we shall denote the random variable $x_{i,j} - p_{n,\mathrm{av}}$ by x_e. Since $|x_{i,j} - p_{n,\mathrm{av}}| \leq 1$, we have for any power $b \geq 2$,

$$\mathrm{E}|x_{i,j} - p_{n,\mathrm{av}}|^b \leq \mathrm{E}|x_{i,j} - p_{n,\mathrm{av}}|^2 = p_{n,\mathrm{av}}(1 - p_{n,\mathrm{av}}).$$

As a consequence, we have

$$\mathrm{E}|X_{w_1} \dots X_{w_m}| \leq \left(p_{n,\mathrm{av}}(1 - p_{n,\mathrm{av}})\right)^{\#E_a} \leq \left(p_{n,\mathrm{av}}(1 - p_{n,\mathrm{av}})\right)^{\#V_a}.$$

Here V_a and E_a denote the vertex and the edge set of the graph G_a respectively. The second inequality follows from the fact $\#E_a \geq \#V_a$ as $l(w_i)$ is even for all $1 \leq i \leq m$. As a consequence,

$$\begin{aligned}
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{a\in\mathcal{A}_{m,\zeta}} \mathrm{E}\,|X_{w_1}\dots X_{w_m}| \\
&\le \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{a\in\mathcal{A}_{m,\zeta}} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^\zeta \qquad (4.6.14)\\
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^\zeta \,\#\mathcal{A}_{m,\zeta}.
\end{aligned}$$

Now we use Lemma 16 to get that

$$\#\mathcal{A}_{m,\zeta} \le n^\zeta 2^{\sum_i l_i} \left(C_1 \sum_i l_i\right)^{C_2 m} \left(\sum_i l_i\right)^{3(\sum_i l_i - 2\zeta)} .$$

As a consequence, the first expression in (4.6.14) is bounded by

$$\begin{aligned}
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \\
&\sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} (np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^\zeta \, 2^{\sum_i l_i} \left(C_1 \sum_i l_i\right)^{C_2 m} \left(\sum_i l_i\right)^{3(\sum_i l_i - 2\zeta)} \\
&\le \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \\
&\sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} (np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^\zeta \, 2^{2mk^*} (C_1 2mk^*)^{C_2 m} \, (2mk^*)^{3(\sum_i l_i - 2\zeta)} \\
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \\
&\qquad \times \sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} (np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^\zeta \, 2^{2mk^*} (C_1 2mk^*)^{C_2 m} (2mk^*)^{3m} \, (2mk^*)^{6(\sum_i \frac{l_i-1}{2} - \zeta)} \\
&= \sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} 2^{2mk^*} (C_3 mk^*)^{C_4 m} \left(\frac{(2mk^*)^6}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_i \frac{l_i-1}{2} - \zeta} .
\end{aligned} \qquad (4.6.15)$$

Here $k^* = \max_{1 \le i \le l}(k_i + 1)$ and C_1, C_2, C_3 and C_4 are positive numeric constants. Now ζ and $\sum_i (l_i - 1)$ are both integers so $\sum_i (l_i - 1) - 2\zeta \ge 1$. As a consequence, again by property of geometric series and the fact $k^* = o(\log(np_{n,\text{av}}))$, we have the last expression in (4.6.15) is bounded by

$$C_5 2^{2mk^*} (C_3 m k^*)^{C_4 m} \left(\frac{(2mk^*)^6}{np_{n,\text{av}}(1 - p_{n,\text{av}})} \right) \tag{4.6.16}$$

where C_5 is another numeric constant. Now we consider two cases when $p_{n,\text{av}}$ converges to 0 and when $p_{n,\text{av}}$ converges to some $p < 1$. In both the cases $(1 - p_{n,\text{av}})$ is asymptotically lower bounded by $\frac{1}{2}$ and $\frac{1-p}{2}$ respectively. So we shall not be concerned about the factor $\frac{1}{(1-p_{n,\text{av}})^3}$ in (4.6.16). Now ignoring $\frac{1}{(1-p_{n,\text{av}})^3}$ in (4.6.16) and taking logarithm of the rest we have

$$\log(C_5) + 2mk^* \log(2) + (C_4 m) \left(\log(k^*) + \log(m) + \log(C_3) \right) + \left(6 \log(2mk^*) - \log(np_{n,\text{av}}) \right). \tag{4.6.17}$$

For large value of k^* the dominant term with the positive sign in (4.6.17) is $2mk^* \log(2)$. However from our assumption

$$2mk^* \log(2) - \log(np_{n,\text{av}}) \to -\infty$$

for any fixed m. As a consequence, the first expression in (4.6.14) goes to 0.

So we can only focus on the words such that $wt(a) = \frac{\sum_{i=1}^m l_i - 1}{2}$. In this case the words w_i of the sentence a are perfectly matched in the sense that for any i there exists a unique j distinct from i such that w_i and w_j have at least one letter in common. In particular, m is even. Now given any such sentence a, we introduce a partition $\eta(a)$ of $\{1, \ldots, m\}$ in the following way. If i and j are in same block of the partition $\eta(a)$, then G_{w_i} and G_{w_j} have at least one edge in common. Observe that any such $\eta(a)$ is a partition of $\{1, \ldots, m\}$ such that each block contains exactly two elements. As a consequence, we can write the L.S. of

(4.6.13) as

$$
\begin{aligned}
&\mathrm{E}[R_{n,1}\dots R_{n,m}] \\
&= o(1) + \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^{m}\frac{l_i-1}{2}} \sum_{\eta}\sum_{a:\eta(a)=\eta} \mathrm{E}\left[X_{w_1}\dots X_{w_m}\right] \\
&= o(1) + \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^{m}\frac{l_i-1}{2}} \sum_{\eta}\sum_{a:\eta(a)=\eta} \prod_{i=1}^{\frac{m}{2}} \mathrm{E}\left[X_{w_{\eta(i,1)}}X_{w_{\eta(i,2)}}\right] \\
&= o(1) + \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^{m}\frac{l_i-1}{2}} \sum_{\eta}\sum_{a:\eta(a)=\eta} \prod_{i=1}^{\frac{m}{2}} \left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^{wt([w_{\eta(i,1)},w_{\eta(i,2)}])} \\
&= o(1) + \left(\frac{1}{n}\right)^{\sum_{i=1}^{m}\frac{l_i-1}{2}} \sum_{\eta}\prod_{i=1}^{\frac{m}{2}} (\#[w_{\eta(i,1)},w_{\eta(i,2)}]).
\end{aligned}
\tag{4.6.18}
$$

Here $[w_{\eta(i,1)}, w_{\eta(i,2)}]$ denotes a typical CLT word pair where $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$ are closed words of length $l_{\eta(i,1)}$ and $l_{\eta(i,2)}$ respectively and $\#[w_{\eta(i,1)}, w_{\eta(i,2)}]$ denotes the cardinality of such CLT word pairs. Note that second step in (4.6.18) follows from Proposition 13. The third step follows from Lemma 14, and since each w_i has an odd number of total edge visits, it has to be $\#V = \#E$ by Lemma 14. Now recalling Proposition 17 and applying Proposition 16 and Lemma 21 we get if the length of the common bracelet between $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$ is r then there are

$$n^{\zeta_1}(1+o(1))\frac{(l_{\eta(i,1)}-1)f(l_{\eta(i,1)}-1,r)}{r}$$

many choices of $w_{\eta(i,1)}$ and for any such $w_{\eta(i,1)}$, there are

$$n^{\zeta_2-r}(1+o(1))2(l_{\eta(i,2)}-1)f(l_{\eta(i,2)}-1,r)$$

many choices of $w_{\eta(i,2)}$. Here ζ_1 and ζ_2 are $wt(w_{\eta(i,1)})$ and $wt(w_{\eta(i,2)})$ respectively. Finally

$wt(w_{\eta(i,1)}, w_{\eta(i,2)}) = \zeta_1 + \zeta_2 - r$. So

$$\#[w_{\eta(i,1)}, w_{\eta(i,2)}] = (1+o(1))n^{wt(w_{\eta(i,1)}, w_{\eta(i,2)})} V_{\eta(i,1),\eta(i,2)}, \tag{4.6.19}$$

where

$$V_{i,j} := \sum_{r=3\,:\,r\text{ odd}}^{\min(l_i-1,l_j-1)} f(l_i-1,r)f(l_j-1,r)\frac{2(l_i-1)(l_j-1)}{r}. \tag{4.6.20}$$

Plugging in these values in (4.6.18), we get

$$\mathrm{E}[R_{n,1}\dots R_{n,m}] = o(1) + \sum_{\eta}\prod_{i=1}^{\frac{m}{2}} V_{\eta(i,1),\eta(i,2)}.$$

Finally taking $m = 2$ we get $V_{i,j}$ to be the asymptotic covariance between $Y_{n,2k_i+1}$ and $Y_{n,2k_j+1}$. This completes the proof of part (i).

Proof of part (ii)

We only prove the case when $p_{n,\mathrm{av}} \to p \in (0,1)$ here. The case when $p_{n,\mathrm{av}} \to 0$ is similar. Firstly,

$$\beta_{2k} = \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\,:\,l(w)=2k+1\ \&\ w\text{ closed}} \big(X_w - \mathrm{E}\left[X_w\right]\big). \tag{4.6.21}$$

For any fixed m, we shall again verify (4.6.13), but with $R_{n,i} \in \{\beta_{2k_1},\dots,\beta_{2k_l}\}$. We again have,

$$\mathrm{E}\left[R_{n,1},\dots,R_{n,m}\right] = \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{w_1\dots w_m} \mathrm{E}\left[(X_{w_1} - \mathrm{E}\left[X_{w_1}\right])\dots(X_{w_m} - \mathrm{E}\left[X_{w_m}\right])\right].$$

Repeating the arguments in the previous proof, it is easy to see that

$$\mathrm{E}\left[(X_{w_1} - \mathrm{E}\left[X_{w_1}\right])\dots(X_{w_m} - \mathrm{E}\left[X_{w_m}\right])\right] = 0$$

unless the sentence $a = [w_i]_{i=1}^m$ is a weak CLT sentence. So we keep our focus only on the weak CLT sentences. Let $\mathcal{A}_{m,\zeta}$ be the set of weak CLT sentences obtained from concatenating m words such that the ith word has length l_i and $wt(a) = \zeta$ for any $a \in \mathcal{A}_{m,\zeta}$. From Proposition 13, $\zeta \le \sum_{i=1}^m \frac{l_i-1}{2}$ where the equality holds only if a is a CLT sentence. Analysis similar to (4.6.14) and (4.6.15) shows that

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} \sum_{a\in\mathcal{A}_{m,\zeta}} \mathrm{E}\left|(X_{w_1}-\mathrm{E}\left[X_{w_1}\right])\ldots(X_{w_1}-\mathrm{E}\left[X_{w_1}\right])\right|$$
$$\le \left(\frac{1}{p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{m}{2}} \sum_{1\le\zeta<\sum_{i=1}^m \frac{l_i-1}{2}} 2^{\sum_{i=1}^m l_i}\left(C_3 m(2k^*+1)\right)^{C_4 m}\left(\frac{(m(2k^*+1))^6}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^m \frac{l_i-1}{2}-\zeta} \tag{4.6.22}$$

Here $k^* = \max_{1\le i\le l}(k_i)$ and C_1, C_2, C_3 and C_4 are positive numeric constants. The additional factor $\left(\frac{1}{p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{m}{2}}$ is due to the following fact. Here some of the connected components in the graph G_a for some $a \in \mathcal{A}_{m,\zeta}$ can be trees where the number of edges is one less than the number of vertices. On the other hand, by Lemma 13 the number of connected components in $G_a \le \frac{m}{2}$. This gives rise to the additional factor. In this context we mention that this additional factor also comes when $p_{n,\mathrm{av}} \to 0$, which is compensated by the scaling $\sqrt{p_{n,\mathrm{av}}}$ in the CLT of β_{2k}.

By the foregoing discussion, we still only need to consider the CLT sentences as in the odd power case. Let $a = [w_i]_{i=1}^m$ be a typical CLT sentence. Applying Proposition 13 we again have for any word w_i there is exactly one other word w_j such that G_{w_i} and G_{w_j} share an edge. As a consequence, the partition $\eta = \eta(a)$ is again a partition of $\{1,\ldots,m\}$ such that each block has exactly two elements. As a consequence,

$$
\begin{aligned}
&\mathrm{E}[R_{n,1}\dots R_{n,m}] \\
&= o(1) + \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^{m}\frac{l_i-1}{2}} \sum_{\eta}\sum_{a:\eta(a)=\eta} \mathrm{E}\left[(X_{w_1}-\mathrm{E}\left[X_{w_1}\right])\dots(X_{w_m}-\mathrm{E}\left[X_{w_m}\right])\right] \\
&= o(1)+ \\
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\sum_{i=1}^{m}\frac{l_i-1}{2}} \\
&\sum_{\eta}\sum_{a:\eta(a)=\eta}\prod_{i=1}^{\frac{m}{2}}\mathrm{E}\Big[\big(X_{w_{\eta(i,1)}}-\mathrm{E}\big[X_{w_{\eta(i,1)}}\big]\big)\big(X_{w_{\eta(i,2)}}-\mathrm{E}\big[X_{w_{\eta(i,2)}}\big]\big)\Big],
\end{aligned}
\tag{4.6.23}
$$

where $[w_{\eta(i,1)}, w_{\eta(i,2)}]$ is a typical CLT word pair. From Lemma 14, there are two possible cases. Firstly, the graph corresponding to the CLT word pair is unicyclic. The analysis is of this case is the same as in the odd power case and has been presented in the proof of part (i). We only do the analysis of the second case when the graph is a tree. Observe that in this case, from Lemma 14, both $w_{\eta(i,1)}$ and $w_{\eta(i,2)}$ are Wigner words, and there is a common edge between the $G_{w_{\eta(i,1)}}$ and $G_{w_{\eta(i,2)}}$. Up to a multiplicative factor of $1+o(1)$, there are $n^{\frac{l_{\eta(i,1)}-1}{2}+1}C_{(l_{\eta(i,1)-1})}$ many Wigner words of length $l_{\eta(i,1)}$. Once any such word is fixed there are $\frac{l_{\eta(i,1)}-1}{2}$ many choices for the edge in $w_{\eta(i,1)}$ which is shared by $w_{\eta(i,2)}$. Once a word $w_{\eta(i,1)}$ and a choice of this edge is fixed there are exactly two ways this edge can be traversed by $w_{\eta(i,2)}$ depending on which letter appears first. Finally, we again have $n^{\frac{l_{\eta(i,2)}-1}{2}+1-2}C_{(l_{\eta(i,2)-1})}$ many choices of $w_{\eta(i,2)}$ after fixing $w_{\eta(i,1)}$, the edge which is shared by $w_{\eta(i,1)}$ and the order of traversal of this edge by $w_{\eta(i,2)}$. So the total number of choices for the CLT word pair of this kind is, up to a multiplicative factor of $1+o(1)$,

$$2\,n^{\frac{l_{\eta(i,1)}-1}{2}+\frac{l_{\eta(i,2)}-1}{2}}\,\frac{(l_{\eta(i,1)}-1)(l_{\eta(i,2)}-1)}{4}C_{(l_{\eta(i,1)-1})}C_{(l_{\eta(i,2)-1})}.$$

Now observe that for any such word pair,

$$\begin{aligned} &\mathrm{E}\Big[\big(X_{w_{\eta(i,1)}} - \mathrm{E}\big[X_{w_{\eta(i,1)}}\big]\big)\big(X_{w_{\eta(i,2)}} - \mathrm{E}\big[X_{w_{\eta(i,2)}}\big]\big)\Big] \\ &= (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\frac{l_{\eta(i,1)}-1}{2}+\frac{l_{\eta(i,2)}-1}{2}-2}\,\mathrm{E}\left[(x_{1,2}-p_{n,\mathrm{av}})^4\right] \\ &\quad - (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\frac{l_{\eta(i,1)}-1}{2}+\frac{l_{\eta(i,2)}-1}{2}} \\ &= (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\frac{l_{\eta(i,1)}-1}{2}+\frac{l_{\eta(i,2)}-1}{2}-2}\,\mathrm{Var}\left[(x_{1,2}-p_{n,\mathrm{av}})^2\right]. \end{aligned} \tag{4.6.24}$$

The rest of the argument is the same as the proof of the odd power case and so we omit the details.

Proof of part (iii)

Now we give a proof of part (iii) of Theorem 7. We are supposed to show for any $k = o(\log(np_{n,\mathrm{av}}))$,

$$\mathsf{Tr}(A_{\mathrm{cen1}}^{2k+1}) - \sum_{r=3:r\text{ odd}}^{2k+1} f(2k+1,r)\frac{2k+1}{r}C_{n,r}(G) \xrightarrow{p} 0. \tag{4.6.25}$$

We prove this by showing the variance of the L.S. of (4.6.25) goes to 0. Recalling the expression of $C_{n,r}(G)$ from (4.2.2) that

$$\begin{aligned} C_{n,r}(G) &= \left(\frac{1}{\sqrt{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}\right)^r \sum_{i_0,i_1,\ldots,i_{r-1}} (x_{i_0,i_1}-p_{n,\mathrm{av}})\ldots(x_{i_{r-1}i_0}-p_{n,\mathrm{av}}) \\ &= \left(\frac{1}{\sqrt{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}\right)^r \sum_{w\in\mathrm{Sc}_r} X_w \end{aligned} \tag{4.6.26}$$

where $i_0,\ldots,i_{r-1}$ are all distinct. Here Sc_r is the class of closed words such that for any $w \in \mathrm{Sc}_r$, G_w is a cycle of length r. First observe that if $r_1 \neq r_2$, then $\mathrm{E}[X_{w_1}X_{w_2}] = 0$ for any $w_1 \in \mathrm{Sc}_{r_1}$ and $w_2 \in \mathrm{Sc}_{r_2}$ trivially. As a consequence, $\mathrm{Cov}(C_{n,r_1}(G), C_{n,r_2}(G)) = 0$ whenever $r_1 \neq r_2$. Now we evaluate

$$\mathrm{Cov}(\mathsf{Tr}(A_{\mathrm{cen1}}^{2k+1}), C_{n,r}(G))$$

for any odd number $r \leq 2k+1$. One can imitate the proof of part (i) to get that

$$\begin{aligned} E_r :=& \mathrm{Cov}\Bigg(\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})} \right)^{\frac{2k+1}{2}} \sum_{w \in \mathrm{No}_k} X_w, C_{n,r}(G) \Bigg) \\ &\leq 2^{2k+r+3} (C_1(2k+r+3))^{2C_2} (2k+r+3)^6 C_3 \left(\frac{(2k+r+3)^6}{np_{n,\mathrm{av}}} \right). \end{aligned} \tag{4.6.27}$$

Here C_1, C_2 and C_3 are known constants. Since $r \leq 2k+1$, summing the second expression in (4.6.27) we have

$$\sum_{r=3: r \text{ odd}}^{2k+1} E_r \leq k 2^{4k+4} (C_1(4k+4))^{2C_2} (4k+4)^6 C_3 \left(\frac{(4k+4)^6}{np_{n,\mathrm{av}}} \right) \to 0. \tag{4.6.28}$$

As a consequence, we only need to analyze the covariance between $Y_{n,2k+1}$ and $C_{n,r}(G)$ where $Y_{n,2k+1}$ was defined in (4.6.10). Now

$$\mathrm{Cov}(Y_{n,2k+1}, C_{n,r}(G)) = \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})} \right)^{\frac{2k+1+r}{2}} \sum_{w_1 \notin \mathrm{No}_k} \sum_{w_2 \in \mathrm{Sc}_r} \mathrm{E}[X_{w_1} X_{w_2}]. \tag{4.6.29}$$

It is easy to see that $\mathrm{E}[X_{w_1} X_{w_2}] = 0$ unless $[w_1, w_2]$ is a weak CLT sentence. Observe that if $a = [w_1, w_2]$ is a CLT word pair then then $wt(a) = \frac{2k+1+r}{2}$, this is an integer as r is taken to be odd. Again we consider $\mathcal{A}_{2,\zeta}$ the set of weak CLT sentences obtained by concatenating two words $w_1 \notin \mathrm{No}_k$ and $w_2 \in \mathrm{Sc}_r$ such that for any $a \in \mathcal{A}_{2,\zeta}$, $wt(a) = \zeta$. Observe that

$$\mathrm{Cov}(Y_{n,2k+1}, C_{n,r}(G)) = \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})} \right)^{\frac{2k+1+r}{2}} \sum_{\zeta=1}^{\frac{2k+1+r}{2}} \sum_{a \in \mathcal{A}_{2,\zeta}} \mathrm{E}[X_{w_1} X_{w_2}]. \tag{4.6.30}$$

By applying Lemma 16 again we have

$$\begin{aligned} \mathcal{T}_r :=& \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})} \right)^{\frac{2k+1+r}{2}} \sum_{1 \leq \zeta < \frac{2k+1+r}{2}} \sum_{a \in \mathcal{A}_{2,\zeta}} \mathrm{E}[|X_{w_1} X_{w_2}|] \\ &\leq C_5 2^{4(k+1)} (2C_3(k+1))^{2C_4} \left(\frac{(4(k+1))^6}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})} \right). \end{aligned} \tag{4.6.31}$$

Note that

$$\sum_{r=3:r \text{ odd}}^{2k+1} \mathcal{T}_r \leq 2^{C_T k} P_T(k) \left(\frac{1}{np_{n,\mathrm{av}}}\right) \to 0,$$

where C_T is a known constant and P_T is a known polynomial in k. The convergence occurs whenever $k = o(\log(np_{n,\mathrm{av}}))$ as $np_{n,\mathrm{av}} \to \infty$.

Now observe that any $w_2 \in \mathrm{Sc}_r$ is a uniword with bracelet length r and there are

$$n(n-1)\ldots(n-r+1) \geq (n-r+1)^r \tag{4.6.32}$$

such words. Now by applying Lemma 21 we have for each w_2, there are at least

$$\begin{aligned}
&(n-r)(n-r-1)\ldots\left(n-\frac{2k+1+r}{2}+1\right)2(2k+1)f(2k+1,r) \\
&\geq 2(2k+1)f(2k+1,r)\left(n-\frac{2k+1+r}{2}+1\right)^{\frac{2k+1-r}{2}}
\end{aligned} \tag{4.6.33}$$

many choices of w_1 such that $[w_1, w_2]$ is a CLT word pair. As a consequence,

$$\begin{aligned}
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1+r}{2}} \sum_{a\in\mathcal{A}_{2,\frac{2k+1+r}{2}}} \mathrm{E}[X_{w_1}X_{w_2}] \\
&\geq \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1+r}{2}} 2(2k+1)f(2k+1,r)\left(n-\frac{2k+1+r}{2}+1\right)^{\frac{2k+1-r}{2}} \\
&\qquad \times (n-r+1)^r (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\frac{2k+1+r}{2}} \\
&\geq 2(2k+1)f(2k+1,r)\left(1-\frac{2k-1+r}{2n}\right)^{\frac{2k+1+r}{2}}.
\end{aligned} \tag{4.6.34}$$

On the other hand, arguments similar to step 2 of the proof of part (i) of Theorem 7 gives us

$$\mathrm{Var}(Y_{n,2k+1}) \leq \sum_{r=3:r \text{ odd}}^{2k+1} 2f(2k+1,r)^2 \frac{(2k+1)^2}{r} + o(1), \tag{4.6.35}$$

and

$$\mathrm{Var}(C_{n,r}) \leq 2r\left(1+O\left(\frac{k^2}{n}\right)\right). \tag{4.6.36}$$

Plugging in the estimates in (4.6.31),(4.6.34), (4.6.35) and (4.6.36) and recalling the fact $\mathrm{Cov}(C_{n,r_1}, C_{n,r_2}) = 0$ for $r_1 \neq r_2$, we have

$$\begin{aligned}
&\mathrm{Var}\left(Y_{n,2k+1} - \sum_{r=3\,:\,r\text{ odd}}^{2k+1} \frac{(2k+1)f(2k+1,r)}{r} C_{n,r}\right) \\
&= \mathrm{Var}(Y_{n,2k+1}) + \\
&\mathrm{Var}\left(\sum_r \frac{(2k+1)f(2k+1,r)}{r} C_{n,r}\right) - 2\sum_r \mathrm{Cov}\left(Y_{n,2k+1}, \frac{(2k+1)f(2k+1,r)}{r} C_{n,r}\right) \\
&\leq o(1) + \sum_r 4f(2k+1,r)^2 \frac{(2k+1)^2}{r} - \sum_r 4f(2k+1,r)^2 \frac{(2k+1)^2}{r}\left(1 - \frac{2k-1+r}{2n}\right)^{\frac{2k+1+r}{2}} \\
&= \sum_r \left(1 - \left(1 - \frac{2k-1+r}{2n}\right)^{\frac{2k+1+r}{2}}\right) 4f(2k+1,r)^2 \frac{(2k+1)^2}{r} + o(1) \\
&= \sum_r O\left(\frac{(2k-1+r)^2}{n}\right) 4f(2k+1,r)^2 \frac{(2k+1)^2}{r} + o(1).
\end{aligned} \tag{4.6.37}$$

Here the last step follows from the elementary inequality $1-(1-x)^y \leq \frac{xy}{1-x}$ for any $0 < x < 1$ and $y > 0$. From Proposition 14 we know $f(2k+1,r) \leq 2^{2k}$. As a consequence, the first expression in (4.6.37) can be further bounded by

$$4(2k+1)^2 2^{4k} \sum_r O\left(\frac{(2k-1+r)^2}{n}\right) + o(1) \to 0 \tag{4.6.38}$$

whenever $k = o(\log(np_{n,\mathrm{av}}))$ as $np_{n,\mathrm{av}} \to \infty$. Recalling (4.6.27), we get

$$\mathrm{Var}\left(\mathsf{Tr}\left(A_{\mathrm{cen1}}^{2k+1}\right) - \sum_{r=3:r\text{ odd}}^{2k+1} \frac{(2k+1)f(2k+1,r)}{r} C_{n,r}\right) \to 0. \tag{4.6.39}$$

This completes the proof of part (iii). □

4.6.2. Proof of parts (i)–(iii) of Theorem 8

We focus on part (i) of Theorem 8. The arguments for parts (ii) and (iii) are similar. All expectation and variance in this part are taken under $\mathbb{P}_{1,n}$ conditioning on the group assignment σ_i, $1 \le i \le n$.

Before going into the proof we introduce some notations that will be useful in the proof. We define $\mathcal{E}_k = \{(0,1), \ldots, (k-1,k)\}$. In the proof we often denote $\mathcal{E}_{2k+1}$ by $\mathcal{E}$ for notational convenience. We shall deal with two disjoint subsets $\mathcal{E}_L$ and $\mathcal{E}_T$ of $\mathcal{E}$ such that $\mathcal{E}_L \cup \mathcal{E}_T = \mathcal{E}$. Let $w = (i_0, \ldots, i_{2k+1})$ be any word. Then for any $e = (j, j+1) \in \mathcal{E}$, we define

$$e(w) = (i_j, i_{j+1}).$$

For any word w, we consider the graph $G_w = (V_w, E_w)$ as defined in Section 4.5.1. Given the word w and a subset $\mathcal{E}' \subset \mathcal{E}$, we define $E(\mathcal{E}'(w)) := \{e(w) : e \in \mathcal{E}'\}$. Observe that $E(\mathcal{E}'(w))$ is the set of unique (undirected) edges traversed by $e(w)$, $e \in \mathcal{E}'$, in the graph G_w, and it does not take into account the number of passages of any of its elements.

Let $d = \frac{p_n - q_n}{2}$. In what follows, we focus on the case where $p_n > q_n$. If $p_n < q_n$, we simply need to replace every t with $-t$. Recall (4.6.3) to get

$$\mathsf{Tr}(A_{\text{cen1}}^{2k+1}) = \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} \sum_{w: l(w)=2k+2 \ \& \ w \text{ closed}} [X_w]. \tag{4.6.40}$$

Here for any word w is an ordered pair $(i_0, \ldots, i_{2k+1})$ where the numbers $i_j \in \{1, 2, \ldots, n\}$ for $0 \le j \le 2k+1$ and $X_w = \prod_{j=0}^{2k} \left(x_{i_j, i_{j+1}} - p_{n,\text{av}}\right)$. However, as the data is generated under the alternative, here $\mathrm{E}[x_{i,j}] = p_{i,j}$ where $p_{i,j} = p_n$ if $\sigma_i = \sigma_j$ and $p_{i,j} = q_n$ if $\sigma_i \ne \sigma_j$. As a consequence, for any $i \ne j$,

$$x_{i,j} - p_{n,\text{av}} = x_{i,j} - p_{i,j} + p_{i,j} - p_{n,\text{av}} = x_{i,j} - p_{i,j} + d\sigma_i\sigma_j, \tag{4.6.41}$$

and so

$$X_w = \prod_{j=0}^{2k} \left(x_{i_j,i_{j+1}} - p_{n,\text{av}}\right) = \prod_{j=0}^{2k} \left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}} + d\sigma_{i_j}\sigma_{i_{j+1}}\right). \tag{4.6.42}$$

At first we note that

$$\prod_{j=0}^{2k} \sigma_{i_j}\sigma_{i_{j+1}} = 1 \tag{4.6.43}$$

irrespective of the values of σ_{i_j}'s. This is due to the fact that $\sigma_{i_0} = \sigma_{i_{2k+1}}$ and so each σ_{i_j} is multiplied an even number of times in (4.6.43). Note that the foregoing argument depends only on the word being closed, regardless of whether its length is odd or even. Now we can write (4.6.42) as

$$X_w = \prod_{j=0}^{2k} \left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right) + d^{2k+1} + V_{n,w}. \tag{4.6.44}$$

Here $V_{n,w}$ comprises of all the cross terms. Plugging (4.6.44) in (4.6.40), we get

$$\begin{aligned}
&\mathsf{Tr}(A_{\text{cen1}}^{2k+1}) \\
&= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} \sum_{w:l(w)=2k+2\ \&\ w\ \text{closed}} \left(\prod_{j=0}^{2k}\left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right) + V_{n,w}\right) \\
&\qquad + \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} (nd)^{2k+1}. \\
&= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} \sum_{w} \left(\prod_{j=0}^{2k}\left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right) + V_{n,w}\right) + t_n^{2k+1}.
\end{aligned} \tag{4.6.45}$$

Here $t_n = \sqrt{\frac{c}{2(1-p_{n,\text{av}})}} \to t$ as $n \to \infty$.

The analysis of

$$D_{n,k} := \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} \sum_{w:l(w)=2k+2\ \&\ w\ \text{closed}} \prod_{j=0}^{2k}\left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right).$$

is same as the proof of Theorem 7 part (i). We only mention that the covariance structure of $\{D_{n,k_i}\}_{i=1}^{l}$ is the same as the covariance structure of $\{\mathsf{Tr}(A_{\text{cen1}}^{2k_i+1})\}_{i=1}^{l}$ due the fact that whenever $k = o(\log(np_{n,\text{av}}))$ both

$$\lim_{n\to\infty} \left(\frac{(p_{n,\text{av}} + d)(1 - p_{n,\text{av}} - d)}{p_{n,\text{av}}(1 - p_{n,\text{av}})} \right)^{\frac{2k+1}{2}} = 1 \tag{4.6.46}$$

and

$$\lim_{n\to\infty} \left(\frac{(p_{n,\text{av}} - d)(1 - p_{n,\text{av}} + d)}{p_{n,\text{av}}(1 - p_{n,\text{av}})} \right)^{\frac{2k+1}{2}} = 1. \tag{4.6.47}$$

It is easy to see that $\text{Cov}(D_{n,k_i}, D_{n,k_j})/V_{i,j}$ is sandwiched by the left sides of (4.6.46) and (4.6.47). Here $V_{i,j}$ is defined as in (4.6.20).

In the rest of this subsection, we complete the proof by analyzing the mean and variance of

$$\left(\frac{1}{np_{n,\text{av}}(1 - p_{n,\text{av}})} \right)^{\frac{2k+1}{2}} \sum_{w} V_{n,w}. \tag{4.6.48}$$

Analysis of the mean of (4.6.48). At first fix w and consider the graph $G = (V, E)$ corresponding to the word w. Now

$$\begin{aligned} V_{n,w} &= \sum_{\emptyset \subsetneq \mathcal{E}_T \subsetneq \mathcal{E}} \prod_{e\in\mathcal{E}_T} (\sigma_{e(w)} d) \prod_{e\in\mathcal{E}_L} (x_{e(w)} - p_{e(w)}) \\ &= \sum_{\emptyset \subsetneq \mathcal{E}_T \subsetneq \mathcal{E}} d^{\#\mathcal{E}_T} \prod_{e\in\mathcal{E}_T} \sigma_{e(w)} \prod_{e\in\mathcal{E}_L} (x_{e(w)} - p_{e(w)}). \end{aligned} \tag{4.6.49}$$

Here for any $e(w) = (i_j, i_{j+1})$, $\sigma_{e(w)} := \sigma_{i_j}\sigma_{i_{j+1}}$, $x_{e(w)} := x_{i_j,i_{j+1}}$ and $p_{e(w)} := \text{E}[x_{e(w)}]$. Observe that

$$\text{E}\left[\prod_{e\in\mathcal{E}_L} (x_{e(w)} - p_{e(w)}) \right] = 0 \tag{4.6.50}$$

unless all the random variables $x_{e(w)} - p_{e(w)}$, $e \in \mathcal{E}_L$, have been repeated at least twice and in this case

$$\text{E}\left[\prod_{e\in\mathcal{E}_L} |x_{e(w)} - p_{e(w)}| \right] \le (1 + o(1)) \left(p_{n,\text{av}}(1 - p_{n,\text{av}})\right)^{\#E(\mathcal{E}_L(w))}.$$

We now fix a typical set $\emptyset \subsetneq \mathcal{E}_L \subsetneq \mathcal{E}$ and an equivalence class $\mathbf{w}$ such that all the random variables on the L.S. of (4.6.50) is repeated at least twice. Fixing $\mathbf{w}$ automatically fixes the graph $G_{\mathbf{w}} = (V_{\mathbf{w}}, E_{\mathbf{w}}) = G = (V, E)$. Observe that

$$\begin{aligned}
&\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} \sum_{w\,:\,w\in\mathbf{w}} \mathrm{E}\left[\left| d^{\#\mathcal{E}_T} \prod_{e\in\mathcal{E}_T} \sigma_{e(w)} \prod_{e\in\mathcal{E}_L} \left(x_{e(w)} - p_{e(w)}\right)\right|\right] \\
&\le \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} n^{\#V} d^{\#\mathcal{E}_T} (1+o(1)) \left(p_{n,\text{av}}(1-p_{n,\text{av}})\right)^{\#E(\mathcal{E}_L(w))} \\
&= (1+o(1)) \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}} n^{\#V} \left(\frac{cp_{n,\text{av}}}{2n}\right)^{\frac{\#\mathcal{E}_T}{2}} \left(p_{n,\text{av}}(1-p_{n,\text{av}})\right)^{\#E(\mathcal{E}_L(w))} \\
&\le C^k \left(\frac{1}{np_{n,\text{av}}}\right)^{\frac{2k+1}{2}} n^{\#V} \left(\frac{p_{n,\text{av}}}{n}\right)^{\frac{\#\mathcal{E}_T}{2}} p_{n,\text{av}}^{\#E(\mathcal{E}_L(w))} \\
&= C^k \left(\frac{1}{n}\right)^{\frac{2k+1}{2} - \#V + \frac{\#\mathcal{E}_T}{2}} \left(\frac{1}{p_{n,\text{av}}}\right)^{\frac{2k+1}{2} - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2}} .
\end{aligned} \tag{4.6.51}$$

Here C is a deterministic constant depending on c and $(1-p)$ where $p = \lim_{n\to\infty} p_{n,\text{av}} \in [0,1)$. Since every edge in $E(\mathcal{E}_L)$ has been traversed at least twice, we have

$$2k+1 = \#\mathcal{E}_L + \#\mathcal{E}_T \ge 2\#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T \;\Rightarrow\; \frac{2k+1}{2} - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2} \ge 0. \tag{4.6.52}$$

Now

$$\begin{aligned}
&\left(\frac{2k+1}{2} - \#V + \frac{\#\mathcal{E}_T}{2}\right) - \left(\frac{2k+1}{2} - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2}\right) \\
&\qquad = \#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T - \#V \ge \#E - \#V \ge 0.
\end{aligned} \tag{4.6.53}$$

Here the last inequality in (4.6.53) holds due to Lemma 12. In what follows, we divide the arguments into three different cases, depending on whether the equalities in (4.6.53) and/or (4.6.52) hold.

Case 1: the equalities in both (4.6.53) and (4.6.52) hold. This case occurs if and only if the following conditions are satisfied:

1. The graph G is unicyclic (from (4.6.53)).

2. Every edge in $E(\mathcal{E}_T(w))$ has been traversed exactly once (from (4.6.53)).

3. $E(\mathcal{E}_T(w)) \cap E(\mathcal{E}_L(w)) = \emptyset$ (from (4.6.53)).

4. Every edge in $E(\mathcal{E}_L(w))$ has been traversed exactly twice (from (4.6.52)).

Observe that from Proposition 15 these properties are satisfied if and only if

$$w \in \mathfrak{W}_{2k+2,r,\zeta}$$

for some odd number r and $2\zeta - r = 2k+1$. Here $\mathfrak{W}_{2k+2,r,\zeta}$ is defined as in Proposition 16. From condition 2 above, we get $\mathcal{E}_T$ corresponds to the walk along the bracelet of the unicyclic graph G. Hence the collection $\mathcal{E}_T$ is actually a closed word. As a consequence, arguing as (4.6.43), we get $\prod_{e\in\mathcal{E}_T} \sigma_{e(w)} = 1$.

Using Proposition 16 for any r, there are $(1+o(1))f(2k+1,r)\frac{2k+1}{r}$ many equivalence classes of such words. Further, for each of these equivalence classes $\mathbf{w}$

$$\begin{aligned}
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \sum_{w\,:\,w\in\mathbf{w}} \mathrm{E}\left[d^{\#\mathcal{E}_T} \prod_{e\in\mathcal{E}_T} \sigma_{e(w)} \prod_{e\in\mathcal{E}_L} \left(x_{e(w)} - p_{e(w)}\right)\right] \\
&= (1+o(1)) \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} n^{\#V} \left(\frac{cp_{n,\mathrm{av}}}{2n}\right)^{\frac{\#\mathcal{E}_T}{2}} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E(\mathcal{E}_L(w))} \\
&= (1+o(1))t^r.
\end{aligned} \tag{4.6.54}$$

Here the second step follows from the fact that $\prod_{e\in\mathcal{E}_T} \sigma_{e(w)} = 1$ and every edge in $E(\mathcal{E}_L)$ has been traversed exactly twice. The third step follows from the equality in (4.6.52) and (4.6.53). So summing over the equivalence classes and the value of r, we get the total contribution of these words in the mean of (4.6.48) is, up to a $1+o(1)$ multiplier,

$$\sum_{r=3:r \text{ odd}}^{2k-1} f(2k+1,r)\frac{2k+1}{r}t^r. \tag{4.6.55}$$

We sum up to $2k-1$ due to the fact that $\mathcal{E}_T \neq \mathcal{E}$.

Case 2: the equality (4.6.53) is satisfied but that in (4.6.52) is violated. In this case, the graph is unicyclic. Let Z and F be the bracelet and the forest corresponding to G respectively. Using the parity principle (Lemma 12) we get that every edge in the forest F has been traversed an even number of times. So the edges traversed exactly once are a subset of the edges in the bracelet Z. Let r be the circuit length. Then $\#E^1 \leq r$. Let $a = [w_i']_{i=1}^m$ be the FK parsing of the word w. Then from Lemma 19 and Lemma 20 we have the number of equivalence classes corresponding to a given m is bounded by

$$\#\Gamma(\zeta, 2k+2, m) \leq 2^{2k+1-m}\binom{2k+1}{m-1}\zeta^{2(m-1)} \leq 2^{2k+1}(2k+1)^{3(m-1)}. \tag{4.6.56}$$

and

$$m = \#E_a^1 - 2wt(a) + 2 + (2k+2) \leq \#E^1 - 2\zeta + 2 + (2k+2). \tag{4.6.57}$$

Here $\zeta = wt(a) = wt(w)$.

As the equality in (4.6.53) is satisfied, we have $\#E(\mathcal{E}_T(w)) = \#E^1 = \#\mathcal{E}_T$. Observe that

$$\zeta = \#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T \quad \text{and} \quad \#\mathcal{E}_L + \#\mathcal{E}_T = 2k+1.$$

Let

$$m' := \#\mathcal{E}_L - 2\#E(\mathcal{E}_L(w))$$

where $m' \geq 1$ as the inequality in (4.6.52) is strict. Plugging in these values in (4.6.57) we have

$$m \le \#\mathcal{E}_T - 2(\#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T) + 2 + (2k+2) = -\#\mathcal{E}_T - \#\mathcal{E}_L + m' + 2 + (2k+2) = m' + 3.$$

On the other hand,

$$\frac{2k+1}{2} - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2} = \frac{2k+1}{2} - \frac{\#\mathcal{E}_L - m'}{2} - \frac{\#\mathcal{E}_T}{2} = \frac{m'}{2}.$$

Plugging in these estimates in (4.6.51) and summing over all equivalence classes $\mathbf{w}$ of current concern and summing over all such choices of $\mathcal{E}_T$ ($\le (2^{2k+1}-1)$), we have the contribution of these words in the expectation of (4.6.48) is bounded by

$$(2^{2k+1}-1)2^{2k+1}\sum_{m'=1}^{2k+1}(2k+1)^6\left(\frac{(2k+1)^6}{np_{n,\text{av}}}\right)^{\frac{m'}{2}} \to 0.$$

<u>Case 3: the equality in (4.6.53) is not satisfied.</u> In this case the graph is not unicyclic. As a consequence, $\#E - \#V \ge 1$. So for any equivalence class $\mathbf{w}$ of this type, we have from the rightmost side of (4.6.51) that

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{\frac{2k+1}{2}}\sum_{w:w\in\mathbf{w}}\mathrm{E}\left|d^{\#\mathcal{E}_T}\prod_{e\in\mathcal{E}_T}\sigma_e\prod_{e\in\mathcal{E}_L}x_{e(w)} - p_{e(w)}\right| \le \frac{C^k}{n}. \tag{4.6.58}$$

Consider any $w \in \mathbf{w}$ and let $a = [w_i']_{i=1}^m$ with $wt(a) = \zeta$ be the FK parsing of the word w. Then using Lemma 19 and Lemma 20 again we have the number of equivalence classes corresponding to a given m is bounded by

$$\#\Gamma(\zeta, 2k+2, m) \le 2^{2k+1-m}\binom{2k+1}{m-1}\zeta^{2(m-1)} \le 2^{2k+1}(2k+1)^{3(m-1)}. \tag{4.6.59}$$

and

$$m = \#E_a^1 - 2wt(a) + 2 + (2k+2) \le \#E^1 - 2\zeta + 2 + (2k+2) \le 4k+3. \tag{4.6.60}$$

Here the last step follows from $\#E^1 \leq 2k+1$ and $\zeta \geq 1$. So (4.6.59) can further be upper bounded by

$$2^{2k+1}(2k+1)^{3(4k+3)}.$$

Now taking the sum over $\mathbf{w}$ in the first expression of (4.6.58) and summing over all such choices of $\mathcal{E}_T$ ($\leq (2^{2k+1}-1)$), we get the contribution of theses words in (4.6.48) is bounded by

$$C^k(2^{2k+1}-1)2^{2k+1}(2k+1)^{3(4k+3)}\frac{1}{n}, \tag{4.6.61}$$

which converges to zero as $n \to \infty$ for all $k = o(\sqrt{\log n})$.

Combining all these results, we have

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \mathrm{E}\left[\sum_w V_{n,w}\right] - \sum_{r=3:r \text{ odd}}^{2k-1} f(2k+1,r)\frac{2k+1}{r}t^r \to 0.$$

Analysis of the variance of (4.6.48). Now we prove the variance of the random variable defined in (4.6.48) goes to 0. For any given word w and $\mathcal{E}_T \subset \mathcal{E}$ let us define

$$V_{n,w,\mathcal{E}_T} = \prod_{e\in\mathcal{E}_T}(\sigma_{e(w)}d)\prod_{e\in\mathcal{E}_L(w)}(x_{e(w)}-p_{e(w)}). \tag{4.6.62}$$

So

$$V_{n,w} - \mathrm{E}\left[V_{n,w}\right] = \sum_{\emptyset\subsetneq\mathcal{E}_T\subsetneq\mathcal{E}} V_{n,w,\mathcal{E}_T} - \mathrm{E}\left[V_{n,w,\mathcal{E}_T}\right].$$

As a consequence,

$$\begin{aligned}&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}\mathrm{Var}\left[\sum_w V_{n,w}\right]\\ &= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}\sum_w\sum_x\sum_{\emptyset\subsetneq\mathcal{E}_{T_1}\subsetneq\mathcal{E}}\sum_{\emptyset\subsetneq\mathcal{E}_{T_2}\subsetneq\mathcal{E}}\mathrm{Cov}\big(V_{n,w,\mathcal{E}_{T_1}},V_{n,x,\mathcal{E}_{T_2}}\big).\end{aligned} \tag{4.6.63}$$

Observe that if the graphs G_w and G_x corresponding to words w and x do not share any edge, the random variables $V_{n,w,\mathcal{E}_{T_1}}$ and $V_{n,x,\mathcal{E}_{T_2}}$ are independent no matter what $\mathcal{E}_{T_1}$ and $\mathcal{E}_{T_2}$ are. As a consequence,

$$\mathrm{Cov}\big(V_{n,w,\mathcal{E}_{T_1}}, V_{n,x,\mathcal{E}_{T_2}}\big) = 0$$

for these word pairs. So we consider the case when G_w and G_x share at least one edge.

As the first step, we bound the number of such word pairs by applying the embedding algorithm stated in the proof of Lemma 16. Here $m = 2$ and the partition $\eta = \{1, 2\}$. As a consequence, applying the embedding algorithm to any such pair (w, x) leads to a closed word $\mathfrak{w}$ of length $4k+3$ where at least one edge in the graph $G_{\mathfrak{w}}$ has been repeated at least twice. We call the function corresponding to the embedding algorithm f (i.e. $f(w, x) = \mathfrak{w}$). One can check that in this case

$$\#f^{-1}(\mathfrak{w}) \leq (4k+3)^3$$

for any closed word $\mathfrak{w}$. Now given any closed word $\mathfrak{w}$ with $wt(\mathfrak{w}) = \zeta$, we consider its FK parsing $a = [\mathfrak{w}_i]_{i=1}^m$ where $wt(a) = \zeta$. We again use Lemma 19 to get the number of equivalence classes of a with a fixed m and ζ is bounded by

$$\#\Gamma(\zeta, 4k+3, m) \leq 2^{4k+3}(4k+3)^{3(m-1)} \tag{4.6.64}$$

On the other hand, using Lemma 20 we get for any a

$$m = \#E_a^1 - 2wt(a) + 2 + (4k+3) \leq \#E_{\mathfrak{w}}^1 - 2\zeta + 2 + (4k+3) \leq 8k+5. \tag{4.6.65}$$

Here the last step follows from $\#E_{\mathfrak{w}}^1 \leq 4k+2$ and $\zeta \geq 1$. Plugging in the upper bound of m in the R.S. of (4.6.65) in (4.6.64) we get for any m,

$$\#\Gamma(\zeta, 4k+3, m) \leq 2^{4k+3}(4k+3)^{3(8k+4)}. \tag{4.6.66}$$

Now observe that for any $\mathcal{E}_{T_1}$ and $\mathcal{E}_{T_2}$, $V_{n,w,\mathcal{E}_{T_1}}$ and $V_{n,x,\mathcal{E}_{T_2}}$ are product of independent Bernoulli random variables multiplied with some deterministic constants. So $\mathrm{Cov}(V_{n,w,\mathcal{E}_{T_1}}, V_{n,x,\mathcal{E}_{T_2}}) = 0$ unless all the random variables in the product $V_{n,w,\mathcal{E}_{T_1}} V_{n,x,\mathcal{E}_{T_2}}$ are repeated at least twice. On the other hand, for all $(\mathcal{E}_{T_1}, \mathcal{E}_{T_2})$ where all the random variables in the product $V_{n,w,\mathcal{E}_{T_1}} V_{n,x,\mathcal{E}_{T_2}}$ are repeated at least twice, by Jensen's inequality $\mathrm{E}|V_{n,w,\mathcal{E}_{T_1}} V_{n,x,\mathcal{E}_{T_2}}| \geq \mathrm{E}|V_{n,w,\mathcal{E}_{T_1}}|\mathrm{E}|V_{n,x,\mathcal{E}_{T_2}}|$. As a consequence,

$$\left|\mathrm{Cov}\left(V_{n,w,\mathcal{E}_{T_1}}, V_{n,x,\mathcal{E}_{T_2}}\right)\right| \leq 2\mathrm{E}\left|V_{n,w,\mathcal{E}_{T_1}} V_{n,x,\mathcal{E}_{T_2}}\right|.$$

Recall that $\mathcal{E}_{L_i} = \mathcal{E} \backslash \mathcal{E}_{T_i}$ for $i = 1, 2$. Now, (4.6.63) can be upper bounded by

$$\begin{aligned}
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1} \mathrm{Var}\left[\sum_{w} V_{n,w}\right] \\
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1} \sum_{\mathfrak{w}} \sum_{(w,x)\in f^{-1}(\mathfrak{w})} \sum_{\mathcal{E}_{T_1},\mathcal{E}_{T_2}} \mathrm{Cov}\left(V_{n,w,\mathcal{E}_{T_1}}, V_{n,x,\mathcal{E}_{T_2}}\right) \\
&\leq \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1} \sum_{\mathfrak{w}} \sum_{(w,x)\in f^{-1}(\mathfrak{w})} \sum_{\mathcal{E}_{T_1},\mathcal{E}_{T_2}} 2\mathrm{E}\left|V_{n,w,\mathcal{E}_{T_1}} V_{n,x,\mathcal{E}_{T_2}}\right| \\
&= 2\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1} \\
&\quad \times \sum_{\mathfrak{w}} \sum_{(w,x)\in f^{-1}(\mathfrak{w})} \sum_{\mathcal{E}_{T_1},\mathcal{E}_{T_2}} d^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}} \mathrm{E}\left|\prod_{e\in\mathcal{E}_{L_1}} (x_{e(w)} - p_{e(w)}) \times \prod_{e\in\mathcal{E}_{L_2}} (x_{e(x)} - p_{e(x)})\right| \\
&\leq 2\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1} \sum_{\mathfrak{w}} \sum_{(w,x)\in f^{-1}(\mathfrak{w})} \sum_{\emptyset\subsetneq\mathcal{E}_T\subsetneq\mathcal{E}_{4k+2}} d^{\#\mathcal{E}_T} \mathrm{E}\left|\prod_{e\in\mathcal{E}_L} (x_{e(\mathfrak{w})} - p_{e(\mathfrak{w})})\right|.
\end{aligned} \tag{4.6.67}$$

Here $\mathcal{E}_{4k+2} := \{(0,1), \ldots, (4k+1, 4k+2)\}$ and $\mathcal{E}_T$ and $\mathcal{E}_L$ give a disjoint partition of $\mathcal{E}_{4k+2}$. Let $w = [\alpha_i]_{i=0}^{2k+1}$, $x = [\beta_i]_{i=0}^{2k+1}$ and $\mathfrak{w} = [\gamma_i]_{i=0}^{4k+2}$. Also let for any $e = (e_1, e_2) \in \mathcal{E}$, $\alpha_e = (\alpha_{e_1}, \alpha_{e_2})$ similarly define β_e and γ_e. The last expression follows from the construction of $\mathfrak{w}$ from (w, x). The most important observation here is, after fixing the words w and x, the function f is defined in such a way that for any $e^{(1)} \in \mathcal{E}_{L_1}$ and $e^{(2)} \in \mathcal{E}_{L_2}$ there are unique $e'^{(1)}, e'^{(2)} \in \mathcal{E}_{4k+2}$ such that $\alpha_{e^{(1)}} = \gamma_{e'^{(1)}}$ and $\beta_{e^{(2)}} = \gamma_{e'^{(2)}}$. Further, $e'^{(1)} \neq e'^{(2)}$.

Hence the last expression of (4.6.67) is justified.

It is easy to observe that $\mathrm{E}|\prod_{e\in\mathcal{E}_L}(x_{e(\mathfrak{w})}-p_{e(\mathfrak{w})})| \le (1+o(1))(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E(\mathcal{E}_L)}$ and plugging in the estimate $\#f^{-1}(\mathfrak{w}) \le (4k+3)^3$ we find the last expression in (4.6.67) is further bounded by

$$\begin{aligned}
&(1+o(1))2\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}(4k+3)^3\sum_{\mathfrak{w}}\sum_{\mathcal{E}_T\neq\emptyset} d^{\#\mathcal{E}_T}(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E(\mathcal{E}_L)}\\
&\le 4\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}(4k+3)^3\sum_{\mathbf{w}}\sum_{\mathcal{E}_T\neq\emptyset}\sum_{\mathfrak{w}\in\mathbf{w}} d^{\#\mathcal{E}_T}(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E(\mathcal{E}_L)}\\
&= 4\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}(4k+3)^3\sum_{\mathbf{w}}\sum_{\mathcal{E}_T\neq\emptyset} n^{\#V_{\mathbf{w}}} d^{\#\mathcal{E}_T}(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E(\mathcal{E}_L)}\\
&\le 4(4k+3)^3 C'^k\sum_{\mathbf{w}}\sum_{\mathcal{E}_T\neq\emptyset}\left(\frac{1}{n}\right)^{2k+1-\#V+\frac{\#\mathcal{E}_T}{2}}\left(\frac{1}{p_{n,\mathrm{av}}}\right)^{2k+1-\#E(\mathcal{E}_L)-\frac{\#\mathcal{E}_T}{2}}.
\end{aligned}\tag{4.6.68}$$

Here $\mathbf{w}$ is the equivalence class corresponding to $\mathfrak{w}$ and C' is a numeric constant. Arguing as (4.6.52) and (4.6.53) we again get

$$2k+1-\#E(\mathcal{E}_L)-\frac{\#\mathcal{E}_T}{2}\ge 0 \tag{4.6.69}$$

and

$$\#E(\mathcal{E}_L)+\#\mathcal{E}_T-\#V_{\mathbf{w}}\ge \#E_{\mathbf{w}}-\#V_{\mathbf{w}}\ge 0. \tag{4.6.70}$$

To see the last inequality in (4.6.70), notice that $\#E_w \ge \#V_w$ and $\#E_x \ge \#V_x$ due to the parity principle, and the embedding algorithm ensures that $\#E_w+\#E_x-\#E_{\mathbf{w}} \le \#V_w+\#V_x-\#V_{\mathbf{w}}$. Indeed we can further show that the inequality in (4.6.70) is always strict. Recall that if the equality in (4.6.70) holds, then $G_{\mathbf{w}}$ is a unicyclic graph and every edge in $E(\mathcal{E}_T)$ has been traversed exactly once. Let $\mathfrak{w}=f(w,x)\in\mathbf{w}$ be any word and $Z_{\mathfrak{w}}$ and $F_{\mathfrak{w}}$ be the bracelet and the forest in $G_{\mathfrak{w}}$ where r is the circuit length. As $\#E(\mathcal{E}_T)>0$, $\#E^1_{\mathfrak{w}}>0$. On the other hand f is defined in such a way that $V_{\mathfrak{w}}=V_w\cup V_x$ and $E_{\mathfrak{w}}=E_w\cup E_x$.

As $l(w) = 2k+2$ and $l(x) = 2k+2$, $\#E_w \geq \#V_w$ and $\#E_x \geq \#V_x$ by the parity principle. This forces both G_w and G_x to be unicyclic. This means $Z_w = Z_x = Z_{\mathfrak{w}}$ since $E_w \cap E_x \neq \emptyset$. This is a contradiction to the fact that at least one edge in $Z_{\mathfrak{w}}$ has been traversed exactly once by the word $\mathfrak{w}$. As a consequence the term inside the summand of the last expression in (4.6.68) is bounded by $\frac{1}{n}$ for any $\mathbf{w}$ and any $\mathcal{E}_T$. Plugging in this estimate and recalling that there are at most $2^{4k+3}(4k+3)^{3(8k+4)}$ many $\mathbf{w}$'s and at most 2^{4k+2} many $\mathcal{E}_T$, we come to the following final upper bound to the last expression in (4.6.68):

$$4(4k+3)^3 C'^k 2^{4k+2} 2^{4k+3} (4k+3)^{3(8k+4)} \frac{1}{n}. \tag{4.6.71}$$

Analysis similar to (4.6.61) will prove that (4.6.71) goes to 0. This completes the proof. □

4.6.3. Proof of part (iv) of Theorem 7 and part (iv) of Theorem 8

Here we focus on the proof of part (iv) of Theorem 7. The proof of part (iv) of Theorem 8 is similar. We first state two important Lemmas which will play important roles in the proof.

Lemma 22. *(Bernstein inequality) Let $\{X_i\}_{i=1}^m$ be independent mean 0 random variables such that $|X_i| \leq M$ for some fixed M. Then for any $s > 0$,*

$$\mathbb{P}\left[\sum_{i=1}^m X_i \geq s\right] \leq \exp\left(-\frac{\frac{1}{2}s^2}{\sum_i \mathrm{E}[X_i^2] + \frac{1}{3}Ms}\right). \tag{4.6.72}$$

In particular, if X_i's are i.i.d. centered Bernoulli p random variables, then

$$\mathbb{P}\left[\left|\sum_{i=1}^m X_i\right| \geq s\right] \leq 2 \times \begin{cases} \exp\left(-\frac{3s}{4}\right) & \text{if } 3mp(1-p) \leq s, \\ \exp\left(-\frac{s^2}{4mp(1-p)}\right) & \text{if } 3mp(1-p) > s. \end{cases} \tag{4.6.73}$$

The above inequality directly follows from plugging in $M = 1$ and using inequality (4.6.72) on X_i and $-X_i$ and taking the union bound.

This is a well known inequality in probability theory hence its proof will be omitted.

For any event E, let $\mathbb{I}_E$ stand for its indicator function.

Lemma 23. *1. Suppose A and B are any two random variables. Then*

$$\mathrm{E}\left[\left|AB\mathbb{I}_{|B|\leq s}\right|\right] \leq s\mathrm{E}[|A|]. \tag{4.6.74}$$

2. Let E be an event with $\mathbb{P}(E) \geq (1-c)$ and A be any random variable with $|A| \leq 1$. Then

$$|\mathrm{E}\left[A\mathbb{I}_E\right] - \mathrm{E}[A]| \leq c. \tag{4.6.75}$$

The proofs follow from direct application of the definition of expectation. We omit the details.

With Lemma 22 and Lemma 23 in hand, we now turn to the proof of part (iv) of Theorem 7. In the rest of this subsection, all expectation and variance are taken with respect to $\mathbb{P}_{0,n}$. The fundamental idea behind this proof is the following. As we estimate $\widehat{p}_{n,\mathrm{av}} = \frac{1}{n(n-1)}\sum X_{ij}$, $\mathrm{Var}(\widehat{p}_{n,\mathrm{av}}) = \frac{\sqrt{2p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}{\sqrt{n(n-1)}} \approx \frac{\sqrt{2p_{n,\mathrm{av}}}}{n}$. As a consequence, one expect that in a typical realization $\frac{\sqrt{p_{n,\mathrm{av}}}}{n} \ll |\widehat{p}_{n,\mathrm{av}} - p_{n,\mathrm{av}}| \ll \frac{\sqrt{p_{n,\mathrm{av}}}}{\sqrt{n}}$. Then one could imitate the proof of part (i) of Theorem 8. We now formalize these ideas.

At first we fix some $\delta \in (\frac{1}{2}, 1)$. Let

$$\mathrm{Ev} := \mathbb{I}_{|\widehat{p}_{n,\mathrm{av}} - p_{n,\mathrm{av}}| \leq \frac{\sqrt{p_{n,\mathrm{av}}}}{n^{\delta}}}. \tag{4.6.76}$$

Recall from (4.2.4), that

$$\mathsf{Tr}(A_{\mathrm{cen2}}^{2k+1}) = \left(\frac{1}{n\widehat{p}_{n,\mathrm{av}}(1-\widehat{p}_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \sum_{w: l(w) = 2k+2 \text{ \& } w \text{ closed}} \left[\hat{X}_w\right]. \tag{4.6.77}$$

Here for any word w, we define $\hat{X}_w := \prod_{j=0}^{2k}\left(x_{i_j,i_{j+1}} - \widehat{p}_{n,\mathrm{av}}\right)$. We now write the R.S. of

(4.6.77) in the following way:

$$\begin{aligned}
&\left(\frac{p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}{\widehat{p}_{n,\mathrm{av}}(1-\widehat{p}_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \sum_{w\,:\,l(w)=2k+2\ \&\ w \text{ closed}} \left[\hat{X}_w\right] \\
&= \mathrm{Ev}\left(\frac{p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}{\widehat{p}_{n,\mathrm{av}}(1-\widehat{p}_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \sum_{w\,:\,l(w)=2k+2\ \&\ w \text{ closed}} \left[\hat{X}_w\right] \\
&\qquad + (1-\mathrm{Ev})\left(\frac{1}{n\widehat{p}_{n,\mathrm{av}}(1-\widehat{p}_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}} \sum_{w\,:\,l(w)=2k+2\ \&\ w \text{ closed}} \left[\hat{X}_w\right].
\end{aligned} \tag{4.6.78}$$

Now we apply (4.6.73) with $m = \frac{n(n-1)}{2}$ and $s = \frac{m\sqrt{p_{n,\mathrm{av}}}}{n^\delta}$ to get that

$$\mathbb{P}\left[\mathrm{Ev}=0\right] \leq 2 \times \begin{cases} \exp\left(-\frac{3s}{4}\right) & \text{if } 3\sqrt{p_{n,\mathrm{av}}}(1-p_{n,\mathrm{av}}) \leq \frac{1}{n^\delta} \\ \exp\left(-\frac{s^2}{4mp_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right) & \text{if } 3\sqrt{p_{n,\mathrm{av}}}(1-p_{n,\mathrm{av}}) \geq \frac{1}{n^\delta}. \end{cases} \tag{4.6.79}$$

Since $m = O(n^2)$,

$$s = \frac{m\sqrt{p_{n,\mathrm{av}}}}{n^\delta} = O(n^{2-\delta}\sqrt{p_{n,\mathrm{av}}}) = O(n^{\frac{3}{2}-\delta}\sqrt{np_{n,\mathrm{av}}}) \gg \sqrt{n}.$$

On the other hand,

$$\frac{s^2}{mp_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})} = O\left(\frac{m^2 p_{n,\mathrm{av}}}{n^{2\delta} m p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right) = O\left(\frac{m}{n^{2\delta}}\right) = O(n^{2-2\delta}).$$

As a consequence, in either case there exists some $\eta > 0$ such that

$$\mathbb{P}\left[\mathrm{Ev}=0\right] \leq \exp(-n^\eta) \to 0.$$

So we can ignore the second term in the last expression of (4.6.78).

Now we analyze the first term of (4.6.78). Observe that when $\mathrm{Ev}=1$,

$$\frac{\widehat{p}_{n,\mathrm{av}}}{p_{n,\mathrm{av}}}=1+\frac{\widehat{p}_{n,\mathrm{av}}-p_{n,\mathrm{av}}}{p_{n,\mathrm{av}}}=1+O\left(\frac{\sqrt{p_{n,\mathrm{av}}}}{n^{\delta}p_{n,\mathrm{av}}}\right)=1+O\left(\frac{1}{n^{\delta-\frac{1}{2}}\sqrt{np_{n,\mathrm{av}}}}\right)$$

and

$$\frac{1-\widehat{p}_{n,\mathrm{av}}}{1-p_{n,\mathrm{av}}}=1-\frac{\widehat{p}_{n,\mathrm{av}}-p_{n,\mathrm{av}}}{1-p_{n,\mathrm{av}}}.$$

Now

$$\left|\left(1+\frac{\widehat{p}_{n,\mathrm{av}}-p_{n,\mathrm{av}}}{p_{n,\mathrm{av}}}\right)^{\frac{2k+1}{2}}-1\right|=O\left(\frac{2k+1}{2n^{\delta-\frac{1}{2}}\sqrt{np_{n,\mathrm{av}}}}\right)\to 0.$$

Hence $\left(\frac{p_{n,\mathrm{av}}}{\widehat{p}_{n,\mathrm{av}}}\right)^{\frac{2k+1}{2}}\to 1$. A similar argument proves that $\left(\frac{1-p_{n,\mathrm{av}}}{1-\widehat{p}_{n,\mathrm{av}}}\right)^{\frac{2k+1}{2}}\to 1$. Now

$$\begin{aligned}&\mathrm{Ev}\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}}\sum_{w\,:\,l(w)=2k+2\ \&\ w\ \text{closed}}\left[\hat{X}_w\right]\\&=\mathrm{Ev}\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}}\left(\sum_{w\,:\,l(w)=2k+2\ \&\ w\ \text{closed}}[X_w]+\sum_{w}E_{n,w}\right)\end{aligned}\tag{4.6.80}$$

where

$$E_{n,w}=\sum_{\mathcal{E}_T}(p_{n,\mathrm{av}}-\widehat{p}_{n,\mathrm{av}})^{\#\mathcal{E}_T}\prod_{e\in\mathcal{E}_L}(x_{e(w)}-p_{n,\mathrm{av}}).\tag{4.6.81}$$

Here $\mathcal{E}_T$ and $\mathcal{E}_L$ are as defined in the proof of part (i) of Theorem 8. As a consequence, in order to prove part (iv) of Theorem 7, it is enough to prove

$$\mathrm{E}\left[\left(\mathrm{Ev}\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}}\sum_{w}E_{n,w}\right)^2\right]\to 0.$$

To this end, first note

$$
\begin{aligned}
&\mathrm{E}\left[\left(\mathrm{Ev}\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}}\sum_{w}E_{n,w}\right)^2\right] \\
&= \mathrm{E}\left[\mathrm{Ev}\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}\sum_{w}\sum_{x}E_{n,w}E_{n,x}\right] \\
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}\times \\
&\quad \mathrm{E}\left[\mathrm{Ev}\sum_{w,x,\mathcal{E}_{T_1},\mathcal{E}_{T_2}}(p_{n,\mathrm{av}}-\widehat{p}_{n,\mathrm{av}})^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}}\prod_{e\in\mathcal{E}_{L_1}}(x_{e(w)}-p_{n,\mathrm{av}})\prod_{e\in\mathcal{E}_{L_2}}(x_{e(x)}-p_{n,\mathrm{av}})\right].
\end{aligned}
\tag{4.6.82}
$$

We divide the remaining arguments into two different cases, depending on whether $E(\mathcal{E}_{L_1}(w))\cup E(\mathcal{E}_{L_2}(x))$ has any edge that has been traversed only once.

Case 1: At least one edge in $E(\mathcal{E}_{L_1}(w))\cup E(\mathcal{E}_{L_2}(x))$ has been traversed exactly once. Let $\mathrm{def}(w,x)$ $(=\mathrm{def}_{L_1,L_2}(w,x))\geq 1$ be the total number of edges in $E(\mathcal{E}_{L_1}(w))\cup E(\mathcal{E}_{L_2}(x))$ which have been traversed exactly once. In this case,

$$
\mathrm{E}\left[\prod_{e\in\mathcal{E}_{L_1}}(x_{e(w)}-p_{n,\mathrm{av}})\prod_{e\in\mathcal{E}_{L_2}}(x_{e(x)}-p_{n,\mathrm{av}})\right]=0.
$$

It is easy to check that

$$
\left|\prod_{e\in\mathcal{E}_{L_1}}(x_{e(w)}-p_{n,\mathrm{av}})\prod_{e\in\mathcal{E}_{L_2}}(x_{e(x)}-p_{n,\mathrm{av}})\right|<1
$$

since each $|x_e-p_{n,\mathrm{av}}|<1$. We now expand the last expression of (4.6.82) in this case.

$$
\begin{aligned}
&\mathrm{E}\left[\mathrm{Ev}\,(p_{n,\mathrm{av}}-\widehat{p}_{n,\mathrm{av}})^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}}\prod_{e\in\mathcal{E}_{L_1}}(x_{e(w)}-p_{n,\mathrm{av}})\prod_{e\in\mathcal{E}_{L_2}}(x_{e(x)}-p_{n,\mathrm{av}})\right] \\
&= \mathrm{E}\left[\mathrm{Ev}\left(\frac{2}{n(n-1)}\right)^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}}\times\sum_{I_1,\ldots,I_{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}}}P_{w,x,I_j}(\mathcal{E}_{T_1},\mathcal{E}_{L_1},\mathcal{E}_{T_2},\mathcal{E}_{L_2})\right]
\end{aligned}
\tag{4.6.83}
$$

where for any $I_j \in \{(u,v) \mid 1 \leq u < v \leq n\}$

$$
\begin{aligned}
&P_{w,x,I_j}(\mathcal{E}_{T_1}, \mathcal{E}_{L_1}, \mathcal{E}_{T_2}, \mathcal{E}_{L_2}) \\
&\quad = \prod_{j=1}^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}} (x_{I_j} - p_{n,\mathrm{av}}) \prod_{e \in \mathcal{E}_{L_1}} (x_{e(w)} - p_{n,\mathrm{av}}) \prod_{e \in \mathcal{E}_{L_2}} (x_{e(x)} - p_{n,\mathrm{av}}).
\end{aligned} \tag{4.6.84}
$$

Subcase (a): Every random variable in (4.6.84) has been repeated at least twice. Clearly in this case $\#\mathcal{E}_{T_1} + \#\mathcal{E}_{T_2} \geq \mathrm{def}(w,x)$ since otherwise there are simply not enough random variables in the first product to match those that appear only once in the second and the third products combined. Let $l \geq \mathrm{def}(w,x)$ be the number of random variables common between

$$
\prod_{j=1}^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}} (x_{I_j} - p_{n,\mathrm{av}}) \tag{4.6.85}
$$

and

$$
\prod_{e \in \mathcal{E}_{L_1}} (x_{e(w)} - p_{n,\mathrm{av}}) \prod_{e \in \mathcal{E}_{L_2}} (x_{e(x)} - p_{n,\mathrm{av}}). \tag{4.6.86}
$$

Note that there are at most $\binom{\#\left(E(\mathcal{E}_{L_1}(w)) \cup E(\mathcal{E}_{L_2}(x))\right)}{l}$ ways these common random variables can be chosen from the product in (4.6.86). Once any such collection of random variables are fixed, we look at the positions occupied by these l common random variables in the product in (4.6.85). Let θ be the number of positions occupied by these l random variables. Clearly, $\theta \geq l$. There are $\binom{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}}{\theta}$ many choices of the positions. Once these positions are fixed, the chosen l random variables induces a partition of these θ positions into l blocks. There are at most l^θ many partitions of θ objects into l blocks. Finally one can permute the l random variables once such a partition is fixed. This further induces an additional $l!$ factor. Once all these are fixed, one is free to choose the rest of the positions in the product

(4.6.85), for those random variables which have not appeared in the product (4.6.86). Now

$$\begin{aligned}
&\sum_{j:j \text{ is not fixed}} \prod_{j=1}^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}} (x_{I_j} - p_{n,\text{av}}) \\
&= \prod_{j:j \text{ is fixed}} (x_{I_j} - p_{n,\text{av}}) \left(\sum_{j\,:\,j \text{ is not fixed}} (x_{I_j} - p_{n,\text{av}}) \right)^{\#\mathcal{E}_{T_1}+\mathcal{E}_{T_2}-\theta} \\
&= \prod_{j:j \text{ is fixed}} (x_{I_j} - p_{n,\text{av}}) \left(p_{n,\text{av}} - \widehat{p}_{n,\text{av}} + O\left(\frac{k}{n^2}\right)\right)^{\#\mathcal{E}_{T_1}+\mathcal{E}_{T_2}-\theta} \left(\frac{n(n-1)}{2}\right)^{\#\mathcal{E}_{T_1}+\mathcal{E}_{T_2}-\theta}
\end{aligned} \tag{4.6.87}$$

Observe that each of the quantities, $l!\binom{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}}{l} \leq (2k+1)^{(2k+1)}$, $\binom{\#\left(E(\mathcal{E}_{L_1}(w))\cup E(\mathcal{E}_{L_2}(x))\right)}{l}$ and θ^l are uniformly bounded by $(2k+1)^{2k+1}$. These estimates allow us to write R.S. of (4.6.83) in the following way

$$\mathrm{E}\left|(2k+1)^{6k+3}\mathrm{Ev} \sum_{l\geq \mathrm{def}(w,x)} \sum_{\theta \geq l} \left(\frac{2}{n(n-1)}\right)^{\theta} \left(p_{n,\text{av}} - \widehat{p}_{n,\text{av}} + O\left(\frac{k}{n^2}\right)\right)^{\#\mathcal{E}_{T_1}+\mathcal{E}_{T_2}-\theta} R_{l,\theta}\right|. \tag{4.6.88}$$

Here $R_{l,\theta}$ is a monomial of $\#\left(E(\mathcal{E}_{L_1}(w)) \cup E(\mathcal{E}_{L_2}(x))\right)$ many independent Bernoulli random variables such that each of the random variables appear more than once. Now

$$\mathrm{E}\left|R_{l,\theta}\right| \leq \left(p_{n,\text{av}}(1-p_{n,\text{av}})\right)^{\#E_a(L_1,L_2)}.$$

Here $E_a(L_1, L_2) := (E(\mathcal{E}_{L_1}(w)) \cup E(\mathcal{E}_{L_2}(x)))$ is a subset of edges in the graph G_a for the sentence $a = [w, x]$. Now applying Lemma 23 and using the fact when $\mathrm{Ev} = 1$, $|p_{n,\text{av}} -$

$\widehat{p}_{n,\mathrm{av}}| \leq \frac{\sqrt{p_{n,\mathrm{av}}}}{n^\delta}$ and $\frac{\sqrt{p_{n,\mathrm{av}}}}{n^\delta} \gg \frac{k}{n^2}$, we can bound (4.6.88) by

$$(2k+1)^{6k+3} \sum_{l \geq \mathrm{def}(w,x)} \sum_{\theta \geq l} \left(\frac{2}{n(n-1)}\right)^{\theta} \left(\frac{\sqrt{p_{n,\mathrm{av}}}}{n^\delta}\right)^{\#\mathcal{E}_{T_1} + \#\mathcal{E}_{T_2} - \theta}$$
$$\left[(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E_a(L_1,L_2)} + \exp(-n^\eta)\right]$$
$$\leq (2k+1)^{6k+3} (2p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E_a(L_1,L_2)} \left(\frac{\sqrt{p_{n,\mathrm{av}}}}{n^\delta}\right)^{\#\mathcal{E}_{T_1} + \#\mathcal{E}_{T_2}}$$
$$\sum_{l \geq \mathrm{def}(w,x)} \sum_{\theta \geq l} \left(\frac{2n^\delta}{n(n-1)\sqrt{p_{n,\mathrm{av}}}}\right)^{\theta}$$
$$\leq C(2k+1)^{6k+3} (2p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E_a(L_1,L_2)} \left(\frac{2}{n(n-1)}\right)^{\mathrm{def}(w,x)} \left(\frac{\sqrt{p_{n,\mathrm{av}}}}{n^\delta}\right)^{\#\mathcal{E}_{T_1} + \#\mathcal{E}_{T_2} - \mathrm{def}(w,x)} \tag{4.6.89}$$

form some numeric constant C. Here we have used the facts that $\frac{2n^\delta}{n(n-1)\sqrt{p_{n,\mathrm{av}}}} \to 0$ and that $(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{Ck} \gg \exp(-n^\eta)$ for any positive numeric constants C and η and $\#E_a(L_1, L_2) = O(k)$.

Now we look at the equivalence classes corresponding to the sentence $a = [w, x]$. Fixing any equivalence class $\mathbf{a}$, let $\#V_{\mathbf{a}}$ be the number of vertices in the graph $G_{\mathbf{a}}$. Summing the R.S. of the last expression of (4.6.89) over all $a \in \mathbf{a}$ and dividing the sum by $\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k+1}$, we have it is less than or equal to

$$D^{2k+1}(2k+1)^{6k+3} \left(\frac{1}{np_{n,\mathrm{av}}}\right)^{2k+1} n^{\#V_{\mathbf{a}}} (p_{n,\mathrm{av}})^{\#E_{\mathbf{a}}(L_1,L_2)} \left(\frac{1}{n^2}\right)^{\mathrm{def}(\mathbf{a})} \left(\frac{\sqrt{p_{n,\mathrm{av}}}}{n^\delta}\right)^{\#\mathcal{E}_{T_1} + \#\mathcal{E}_{T_2} - \mathrm{def}(\mathbf{a})} \tag{4.6.90}$$

Here $\#V_{\mathbf{a}}$, $\#E_{\mathbf{a}}(L_1, L_2)$ and $\mathrm{def}(\mathbf{a})$ are the common value of $\#V_a$, $\#E_a(L_1, L_2)$ and $\mathrm{def}(a)$ for any $a \in \mathbf{a}$. Also note that we have ignored the terms containing $(1 - p_{n,\mathrm{av}})$ since $\lim_n(1 - p_{n,\mathrm{av}}) > (1 - p)$ and $p \in [0, 1)$. Simplifying, we have the powers of $\frac{1}{n}$ and $\frac{1}{p_{n,\mathrm{av}}}$ in (4.6.90) are given by

$$2k + 1 - \#V_{\mathbf{a}} + 2\mathrm{def}(\mathbf{a}) + \delta(\#\mathcal{E}_{T_1} + \#\mathcal{E}_{T_2} - \mathrm{def}(\mathbf{a})) \tag{4.6.91}$$

and

$$2k+1-\#E_{\mathbf{a}}(L_1,L_2)-\frac{1}{2}(\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}-\mathrm{def}(\mathbf{a})). \tag{4.6.92}$$

Observe that

$$\begin{aligned} &2(\#E_{\mathbf{a}}(L_1,L_2)-\mathrm{def}(\mathbf{a}))+\mathrm{def}(\mathbf{a})\le \#\mathcal{E}_{L_1}+\#\mathcal{E}_{L_2}\\ &\Leftrightarrow \#E_{\mathbf{a}}(L_1,L_2)\le \frac{1}{2}(\#\mathcal{E}_{L_1}+\#\mathcal{E}_{L_2}+\mathrm{def}(\mathbf{a})). \end{aligned}$$

Plugging this estimate in (4.6.92) and using the fact $\#\mathcal{E}_{L_1}+\#\mathcal{E}_{L_2}+\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}=4k+2$,, we have (4.6.92) is always greater than or equal to 0. Now we prove the difference between (4.6.91) and (4.6.92) is always greater than or equal to $\frac{1}{2}$:

$$\begin{aligned} &-\#V_{\mathbf{a}}+2\mathrm{def}(\mathbf{a})+\delta(\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}-\mathrm{def}(\mathbf{a}))+\#E_{\mathbf{a}}(L_1,L_2)+\frac{1}{2}(\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}-\mathrm{def}(\mathbf{a}))\\ &\ge -\#V_{\mathbf{a}}+2\mathrm{def}(\mathbf{a})-(\delta+\frac{1}{2})\mathrm{def}(\mathbf{a})+\#E_{\mathbf{a}}(L_1,L_2)+\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}\\ &\ge \#E_{\mathbf{a}}-\#V_{\mathbf{a}}+\frac{1}{2}\mathrm{def}(\mathbf{a})\ge\frac{1}{2}\mathrm{def}(\mathbf{a})\ge\frac{1}{2}. \end{aligned} \tag{4.6.93}$$

Here we have used the fact $\frac{1}{2}<\delta<1$. Recall (4.6.64) and (4.6.60) to get that there are at most $(C'k)^{D'k}$ many equivalence classes $\mathbf{a}$ where C' and D' are some known numbers. So summing (4.6.90) over all the equivalence classes $\mathbf{a}$, we get the contribution of all terms in the current subcase in (4.6.82) is bounded by

$$D^{2k+1}(2k+1)^{4k+2}(C'k)^{D'k}\frac{1}{\sqrt{n}}\to 0. \tag{4.6.94}$$

Subcase (b): At least one random variable in the product (4.6.84) appears only once. Here we apply Lemma 23 to get

$$\sum_{I_1,\ldots,I_{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}}} \Bigg| \mathrm{E}\Bigg[\mathrm{Ev}\left(\frac{2}{n(n-1)}\right)^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}} \times \prod_{j=1}^{\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2}} (x_{I_j} - p_{n,\mathrm{av}}) \prod_{e\in\mathcal{E}_{L_1}} (x_{e(w)} - p_{n,\mathrm{av}}) \prod_{e\in\mathcal{E}_{L_2}} (x_{e(x)} - p_{n,\mathrm{av}})\Bigg]\Bigg| \le n^{2(\#\mathcal{E}_{T_1}+\#\mathcal{E}_{T_2})}\exp(-n^{\eta}).$$

One can again sum over all the equivalence classes **a** to get the contribution of the current subcase in (4.6.82) is bounded by

$$C_1^{2k+1}(C_2k)^{C_3k}n^{C_4k}\exp(-n^{\eta}) \le n^{C_5k}\exp(-n^{\eta}) \tag{4.6.95}$$

Here $C_1,\ldots,C_5$ are some known constants. Since $k = o(\min(\sqrt{\log n}, \log(np_{n,\mathrm{av}})))$, one gets $n^{C_5k}\exp(-n^{\eta}) \to 0$ for any $\eta > 0$. As a consequence, the contribution of the current subcase in (4.6.82) goes to 0.

Case 2: All the edges in $E(\mathcal{E}_{L_1}(w)) \cup E(\mathcal{E}_{L_2}(x))$ have been traversed at least twice. This case can be done by imitating the analysis of mean of (4.6.48) in the proof of part (i) of Theorem 8. In particular using arguments analogous to (4.6.52) and (4.6.53) for the sentence $a = [w, x]$ one can prove that the contribution of the present case in (4.6.82) is bounded by

$$\frac{(C_{\mathrm{r}}k)^{D_{\mathrm{r}}k}}{n^{\delta-\frac{1}{2}}} \to 0 \tag{4.6.96}$$

for some known C_{r} and D_{r}.

Summary Combining (4.6.94), (4.6.95) and (4.6.96) one gets

$$\mathrm{E}\left[\left(\mathrm{Ev}\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{\frac{2k+1}{2}}\sum_{w}E_{n,w}\right)^2\right] \le D^{2k+1}(2k+1)^{4k+2}(C'k)^{D'k}\frac{1}{\sqrt{n}} + n^{C_5k}\exp(-n^{\eta}) + \frac{(C_{\mathrm{r}}k)^{D_{\mathrm{r}}k}}{n^{\delta-\frac{1}{2}}} \to 0. \tag{4.6.97}$$

This completes the proof. $\square$

4.6.4. Proof of Theorem 9

In this proof we focus on proving results under both null and alternative for A_{cen1}. The proof of being able to use A_{cen2} instead of A_{cen1} is similar to the proof of part (iv) of Theorem 7, and hence is omitted.

Proof under the null

Throughout this part, all expectation and variance are taken with respect to $\mathbb{P}_{0,n}$. Observe that

$$\begin{aligned}\mathsf{Tr}\left(A_{\text{cen1}}^{2k}\right) &= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_w X_w \\ &= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \left[\sum_{w\in\mathcal{W}_1} X_w + \sum_{w\in\mathcal{W}_2} X_w + \sum_{w\in\mathcal{W}_3} X_w + \sum_{w\in\mathcal{W}_4} X_w\right].\end{aligned} \tag{4.6.98}$$

Here $\mathcal{W}_1$ corresponds to the set of Wigner words, $\mathcal{W}_2$ stands for the set of all weak Wigner words (Definition 20), $\mathcal{W}_3 = \cup_r \mathfrak{W}_{2k+1,r,k+r/2}$ (Proposition 16) collects all words corresponding to unicyclic graphs with a bracelet of length at least 4, and $\mathcal{W}_4$ is the complement of $\mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3$.

There are $\psi_{2k} n(n-1)\dots(n-k)$ many words in $\mathcal{W}_1$ and each of them have expectation $(p_{n,\text{av}}(1-p_{n,\text{av}}))^k$. As a consequence,

$$\begin{aligned}\mathrm{E}\left[\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w\right] &= n\psi_{2k} - \psi_{2k}\sum_{j=1}^k j + O\left(\frac{1}{n}\right) \\ &= n\psi_{2k} - \psi_{2k}\binom{k+1}{2} + O\left(\frac{1}{n}\right).\end{aligned}$$

We know that

$$\begin{aligned}
&\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w - k\psi_{2k}\,\mathsf{Tr}(A_{\mathrm{cen1}}^2)\right] \\
&= \mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w\right] \\
&- 2\mathrm{Cov}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w, k\psi_{2k}\,\mathsf{Tr}(A_{\mathrm{cen1}}^2)\right] \\
&\qquad + \mathrm{Var}\left[k\psi_{2k}\,\mathsf{Tr}(A_{\mathrm{cen1}}^2)\right].
\end{aligned} \tag{4.6.99}$$

Now we analyze each term of (4.6.99) separately. Observe that for each $w \in \mathcal{W}_1$, $\mathrm{E}[X_w] = \left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})^k\right)$. So

$$\begin{aligned}
&\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w\right] \\
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k} \sum_{w_1,w_2\in\mathcal{W}_1} \mathrm{E}\left[\left(X_{w_1} - (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^k\right)\left(X_{w_2} - (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^k\right)\right] \\
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k} \sum_{w_1,w_2\in\mathcal{W}_1 \,\mid\, \#(E_{w_1}\cap E_{w_2})\geq 1} \left[\mathrm{E}\left[X_{w_1,w_2}\right] - (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{2k}\right].
\end{aligned} \tag{4.6.100}$$

Observe that in the sum

$$\begin{aligned}
&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k} \sum_{w_1,w_2\in\mathcal{W}_1 \,\mid\, \#(E_{w_1}\cap E_{w_2})\geq 1} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{2k} \\
&= 2k^2\psi_{2k}^2\left(1 + O\left(\frac{k^2}{n}\right)\right) + O\left(2^{4k+2}\sum_{t=1}^{2k-1}(C_1(4k+2))^{2C_2}(4k+2)^6\left(\frac{(4k+2)^6}{n}\right)^{2k-t}\right) \\
&= 2k^2\psi_{2k}^2 + o(1).
\end{aligned} \tag{4.6.101}$$

Now

$$
\begin{aligned}
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k} \sum_{w_1,w_2\in\mathcal{W}_1 \;|\; \#(E_{w_1}\cap E_{w_2})\geq 1} \mathrm{E}\left[X_{w_1,w_2}\right] \\
&= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k} \sum_{w_1,w_2\in\mathcal{W}_1 \;|\; \#(E_{w_1}\cap E_{w_2})= 1} \mathrm{E}\left[X_{w_1,w_2}\right] + \mathbf{E} \\
&= \left(\frac{1}{p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2} 2k^2\psi_{2k}^2\left(1+O\left(\frac{k^2}{n}\right)\right)\mathrm{E}\left(x_{1,2}-p_{n,\mathrm{av}}\right)^4 + \mathbf{E} \\
&= \left(\frac{1}{p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2} 2k^2\psi_{2k}^2\mathrm{E}\left(x_{1,2}-p_{n,\mathrm{av}}\right)^4 + O\left(\frac{2k^4\psi_{2k}^2}{np_{n,\mathrm{av}}}\right) + \mathbf{E}.
\end{aligned} \tag{4.6.102}
$$

where $\mathbf{E}$ is an error term arising from the cases where $\#E_{w_1}\cap E_{w_2}\geq 2$. In this $\mathbf{E}$, the dominant term is the case when $\#E_{w_1}\cap E_{w_2}=2$ and the graph induced by the sentence $a=[w_1,w_2]$ is a tree. One can show that the terms

$$
-2\mathrm{Cov}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{k}\sum_{w\in\mathcal{W}_1} X_w, k\psi_{2k}\,\mathsf{Tr}(A_{\mathrm{cen1}}^2)\right] + \mathrm{Var}\left[k\psi_{2k}\,\mathsf{Tr}(A_{\mathrm{cen1}}^2)\right]
$$

cancels out the term

$$
\left(\frac{1}{p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2} 2k^2\psi_{2k}^2\mathrm{E}\left(x_{1,2}-p_{n,\mathrm{av}}\right)^4 - 2k^2\psi_{2k}^2
$$

however the term $\mathbf{E}$ remains. On the other hand when $k=o\left(\log\left(np_{n,\mathrm{av}}^2\right)\right)$ and $np_{n,\mathrm{av}}^2\to\infty$, $\mathbf{E}\to 0$. Hence

$$
\begin{aligned}
\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{k}\sum_{w\in\mathcal{W}_1} X_w - k\psi_{2k}\,\mathsf{Tr}(A_{\mathrm{cen1}}^2)\right] &= E + \frac{2^{C_1k}\mathrm{Poly}(k)}{np_{n,\mathrm{av}}} \\
&= \frac{v_{2,2k}}{np_{n,\mathrm{av}}^2}(say) + \frac{2^{C_1k}\mathrm{Poly}(k)}{np_{n,\mathrm{av}}} + O\left(\frac{2^{C_2k}\mathrm{Poly}(k)}{n}\right).
\end{aligned} \tag{4.6.103}
$$

Next, arguments similar to the proof of Theorem 7 part (i) and (iii) lead to

$$
\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{k}\sum_{w\in\mathcal{W}_3} X_w - \sum_{r=4:r\text{ even}}^{2k}\frac{f(2k,r)2k}{r}C_{n,r}(G) \xrightarrow{p} 0.
$$

Furthermore, by definition all the words in $\mathcal{W}_4$ have zero expectation and we have seen in Lemma 14 that these words do not contribute in the asymptotic variance of $\mathsf{Tr}\left(A_{\text{cen}_1}^{2k}\right)$, either. So we can simply ignore the words in $\mathcal{W}_4$. At this point it is clear that

$$T_{2k} - \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w \in \mathcal{W}_2} X_w \xrightarrow{p} 0.$$

So we turn to inspecting the words $w \in \mathcal{W}_2$.

The words in $\mathcal{W}_2$ can further be divided into the following two classes:

1. w is a critical weak Wigner word (Definition 20): $\#V_w = k$ for $G_w = (V_w, E_w)$.

2. w is not a critical weak Wigner word. In this case we call w to be a sub-critical weak Wigner words.

The computation related to critical weak Wigner words on dense graphs can be found in (Anderson and Zeitouni, 2006, pp.320-322). However, when the graph is sparse (i.e., when $p_{n,\text{av}} \to 0$) we have to be especially careful with the trees: Since the number of edges in a tree is one less than the number of vertices, one gets additional powers of $p_{n,\text{av}}$ in the denominator.

From Proposition 12, we know that whenever w is critical weak Wigner word, G_w is either a unicyclic graph or a tree. When G_w is a tree, there is one exceptional edge which is traversed four times and all the other edges are traversed twice. Also $\#V_w = k$ and $\#E_w = k-1$. So for any such w,

$$\mathrm{E}[X_w] = (p_{n,\text{av}}(1-p_{n,\text{av}}))^{k-2}\,\mathrm{E}\left[(x_{1,2} - p_{n,\text{av}})^4\right].$$

On the other hand, given any equivalence class there are $(1+o(1))n^k$ many such words. Let $\alpha_{2,2k}$ be the number of equivalence classes of such critical weak Wigner trees. (An exact enumeration can be found below.) So the total contribution of these critical weak Wigner trees to the expectation of T_{2k} is given by

$$\frac{\alpha_{2,2k}\mathrm{E}\left[(x_{1,2}-p_{n,\mathrm{av}})^4\right]}{\left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^2}=(1+o(1))\frac{\alpha_{2,2k}}{p_{n,\mathrm{av}}}$$

for vanishing $p_{n,\mathrm{av}}$. For the variance calculation, we need to consider the word pairs $[w_1, w_2]$ where both w_1 and w_2 are critical weak Wigner tree and the sentence $a = [w_1, w_2]$ is a weak CLT sentence. The leading term here comes from the case when a is a tree and w_1, w_2 share exactly one edge. There can be three possible sub-cases here. Firstly, one edge in a is repeated exactly eight times and all the other edges are repeated exactly twice. Secondly, three edges in a are repeated exactly four times and all the other edges are repeated exactly twice. Thirdly, one edge is repeated exactly six times, one edge is repeated exactly four times and all the other edges are repeated exactly twice. Now $\#V_a = 2k-2$ since w_1 and w_2 share exactly one edge this corresponds to two common vertices. Whenever $p_{n,\mathrm{av}} \to 0$, in all the aforesaid cases $\mathrm{E}[X_a] = (1+o(1))\left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^{2k-1}$. Let $v_{2k}n^{2k-2}$ be the number of such sentences. The contribution of these sentences in the variance of T_{2k} is

$$(1+o(1))v_{2k}\frac{n^{2k-2}\left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^{2k-1}}{n^{2k}\left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^{2k}}=(1+o(1))\frac{v_{1,2k}}{n^2p_{n,\mathrm{av}}^3}.$$

It can be proven that the variances of all the other random variables in T_{2k} are negligible with respect to $\frac{1}{n^2p_{n,\mathrm{av}}^3}$.

When w is a critical weak Wigner word and G_w is unicyclic, $\mathrm{E}[X_w] = \left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^k$. As a consequence, the total contribution of these critical weak Wigner unicyclic words in the expectation of T_{2k} is given by

$$\alpha_{1,2k}(1+o(1)).$$

Here $\alpha_{1,2k}$ is the number of equivalence classes of such words.

Now we consider the contributions of sub-critical weak Wigner words. In this case, we are only concerned about the trees. All the words w such that G_w is a tree and $\#V_w = k-1$ contribute jointly a term of $\frac{\alpha_{3,2k}}{np_{n,\mathrm{av}}^2}$ in the expectation of $\mathrm{E}[T_{2k}]$. Here $\alpha_{3,2k}$ is the number of equivalence classes of such words. The variance of these words is negligible compared to $\frac{1}{n^2p_{n,\mathrm{av}}^3}$. Unlike the previous two cases we do not know an easy way to calculate $\alpha_{3,2k}$ explicitly. There are also two cases arising here. Firstly, one edge in a is repeated exactly six times and all the other edges are repeated exactly twice. Secondly, two edges in a are repeated exactly four times and all the other edges are repeated exactly twice. Here one needs to consider multiple bracelets and the argument becomes tedious.

We conclude this part by deriving the expressions of $\alpha_{1,2k}$ and $\alpha_{2,2k}$. A generic recipe for evaluating $\alpha_{1,2k}$ and $\alpha_{2,2k}$ is given in equation (46) of Anderson and Zeitouni (2006), which was done for more general matrices. Simplifying all the results in (46) of Anderson and Zeitouni (2006) for Wigner matrices one gets

$$\begin{aligned}
\alpha_{1,2k} &= \sum_{r=3}^{k} f(2k,2r)\frac{k(r+1)}{r} = \sum_{r=3}^{k}(r+1)\binom{2k}{k+r} \\
&= \sum_{r=3}^{k}(r+1)\binom{2k}{k+r} - \binom{k+1}{2}\psi_{2k} + 3\binom{2k}{k+2} + \binom{k+1}{2}\psi_{2k} - 3\binom{2k}{k+2} \\
&= \sum_{r=1}^{k}\binom{2k}{k+r} - 2k\psi_{2k} + \binom{k+1}{2}\psi_{2k} - 3\binom{2k}{k+2} \\
&= 2^{2k-1} - \binom{2k}{k}\frac{5k+1}{2(k+1)} + \binom{k+1}{2}\psi_{2k} - 3\binom{2k}{k+2},
\end{aligned} \tag{4.6.104}$$

and

$$\alpha_{2,2k} = f(2k,4)\frac{k}{2} = \binom{2k}{k+2}. \tag{4.6.105}$$

Note that in order to derive the final expression in (4.6.104), we have used the following identity

$$\sum_{r=1}^{k}\binom{2k}{k+r} - 2k\psi_{2k} = \sum_{r=3}^{k}(r+1)\binom{2k}{k+r} - \binom{k+1}{2}\psi_{2k} + 3\binom{2k}{k+2}.$$

This can be verified by elementary calculation. Hence the proof is skipped. Using Lemma 15 and Lemma 16 one gets

$$\alpha_{3,2k} \leq 2^{2k}(2k)^{12} \tag{4.6.106}$$

and

$$v_{1,2k}, v_{2,2k} \leq 2^{4k}(C_1 k)^{C_2} \tag{4.6.107}$$

for some positive numeric constants C_1 and C_2. This completes the proof under the null.

Proof under the alternative

The proof under the alternative is very much similar in spirit to the proof of part (i) of Theorem 8. First of all we prove that we can simply ignore the words $w \in \mathcal{W}_2$ such that G_w is not a tree and all the words $w \in \mathcal{W}_4$.

If $w \in \mathcal{W}_4$, $V_w \leq E_w$ and there is at least one edge in $e \in E_w$ which is repeated exactly once. In this case under $\mathbb{P}_{0,n}\mathrm{E}[X_w] = 0$. However, under $\mathbb{P}_{1,n}$ it is not true that $\mathrm{E}_{\mathbb{P}_{1,n}}[X_w] = 0$. Let $e_1, \ldots, e_l$ be the edges which are traversed exactly once in G_w. Then

$$X_w = \prod_{m=1}^{l}(x_{e_m} - p_{n,\mathrm{av}}) \prod_{j\,:\,\{i_j, i_{j+1}\} \notin \{e_1,\ldots,e_l\}} \left(x_{i_j,i_{j+1}} - p_{n,\mathrm{av}}\right).$$

Further the random variables $x_{e_1}, \ldots, x_{e_l}$ are independent of the product

$$\prod_{j\,:\,\{i_j, i_{j+1}\} \notin \{e_1,\ldots,e_l\}} \left(x_{i_j,i_{j+1}} - p_{n,\mathrm{av}}\right).$$

Since all the other edges are repeated at least twice we have

$$l + 2(\#E_w - l) \leq 2k \Rightarrow \#E_w \leq k + l/2.$$

On the other hand $\#V_w \leq \#E_w$. However, the equality in the both cases leads us to $w \in \mathcal{W}_3$ which has been dealt before (requires a short proof. I am deliberately omitting). So at least one of the inequalities must be strict. On the other hand under the alternative

$$\begin{aligned}\left|\mathrm{E}_{\mathbb{P}_{1,n}}\left[X_w\right]\right| &\le \left|\prod_{m=1}^{l} \mathrm{E}_{\mathbb{P}_{1,n}}(x_{e_m} - p_{n,\mathrm{av}})\right| (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E_w - l} \\ &\le O\left(\sqrt{\frac{p_{n,\mathrm{av}}}{n}}\right)^{l} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E_w - l}\end{aligned} \tag{4.6.108}$$

Given any equivalence class $\mathbf{w}$, there are at most $n^{V_{\mathbf{w}}}$ many words. So we have

$$\left|\sum_{w\in\mathbf{w}} \mathrm{E}\left[X_w\right]\right| \le n^{\#V_{\mathbf{w}} - l/2} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E_{\mathbf{w}} - l/2}.$$

We now use the fact that either $\#E_{\mathbf{w}} - l/2 < k$ or $\#V_{\mathbf{w}} < \#E_{\mathbf{w}}$. In either cases

$$n^{\#V_{\mathbf{w}} - l/2} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#E_{\mathbf{w}} - l/2} \ll n^{k} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{k}.$$

We further divide the case into following two sub cases:

i)$\#V_{\mathbf{w}} < \#E_{\mathbf{w}}$: Now fixing the value of $\#E_{\mathbf{w}}$ a very crude upper bound of the corresponding words are given by $\#E_{\mathbf{w}}^{2k} \le (2k)^{2k}$. Since one can see the word $\mathbf{w}$ as a partition of the pairs $\{i_j, i_{j+1}\}$ into $\#E_{\mathbf{w}}$ blocks. As a consequence,

$$\sum_{\mathbf{w}} \sum_{w\in\mathbf{w}} \left|\mathrm{E}\left[X_w\right]\right| \le O\left(\frac{(2k)^{2k}}{n}\right) \to 0$$

ii) $\#V_{\mathbf{w}} = \#E_{\mathbf{w}}$: We fix the value of $\#V_{\mathbf{w}} = \zeta$. As argued earlier, we have in this case $\zeta < k - l/2$. In this case let $a = [w_i']_{i=1}^{m}$ be the FK parsing of any word $w \in \mathbf{w}$. Then by Lemma 19, we have

$$\#\Gamma(\zeta, 2k+2, m) \le 2^{2k-m}\binom{2k}{m-1}\zeta^{2(m-1)} \le 2^{2k}(2k)^{3(m-1)}$$

where

$$m = \#E_a^1 - 2wt(a) + 2 + (2k+1) \le l - 2\zeta + 2k + 3.$$

Here $\Gamma(\zeta, 2k+2, m)$ corresponds to the number of equivalence classes with $\#V_{\mathbf{w}} = \zeta$. As a consequence,

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{\mathbf{w}} \sum_{w \in \mathbf{w}} |\mathrm{E}[X_w]| \leq \sum_{l=1}^{2k} \sum_{\zeta=1}^{k-l/2-1} 2^{2k}(2k)^6 \left(\frac{(2k)^6}{np_{n,\text{av}}}\right)^{k-\zeta+l/2} \to 0. \tag{4.6.109}$$

Now our next task is to prove

$$\mathrm{E}_{\mathbb{P}_{1,n}}\left[\left(\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_w X_w\right)^2\right] \to 0.$$

We write

$$\mathrm{E}_{\mathbb{P}_{1,n}}\left[\left(\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_w X_w\right)^2\right] = \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k} \sum_{w,x} \mathrm{E}[X_w X_x]. \tag{4.6.110}$$

There can be two sub cases:

i) The sentence $a = [w, x]$ has at least one edge repeated exactly once.

ii) The sentence $a = [w, x]$ has all the edges repeated at least twice.

In the sub case $i)$ if the graph G_a is unicyclic, then the graphs G_w and G_x are disjoint from each other. Otherwise if G_a is unicyclic and G_w and G_x share an edge, they will have the common bracelet. This is a contradiction to the fact that G_a has at least one edge repeated exactly once. With this knowledge we can analyze

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k} \sum_{w,x} \mathrm{E}[X_w X_x]$$

in this case exactly similarly as the analysis of the mean

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_w \mathrm{E}[X_w].$$

Now we come to the analysis of sub case *ii*). In this case since all the edges appearing exactly once in w and x are paired, we have the number of edges appearing in w and x to be the same and both equal to l. Now

$$\#E_a \le (\#E_w - l) + (\#E_x - l) + l = \#E_w + \#E_x - l \le 2k + l - l = 2k$$

and

$$\#V_a \le \#E_a.$$

The equality in both the above equation makes $w, x \in \mathcal{W}_3$. So at least one of the inequality is strict. Now by Lemma 16 fixing a value of $\#V_a = t$, we have the number of sentences belonging to this class is less than or equal to $n^t 2^{4k}(4C_1k)^{2C_2}(4k)^{3(4k-2t)}$

Now

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k} \sum_a \mathrm{E}\left[X_w X_x\right] \le \sum_{t=1}^{2k-1} 2^{4k}(4C_1k)^{2C_2}\left(\frac{(4k)^6}{np_{n,\text{av}}}\right)^{2k-t} \to 0. \quad (4.6.111)$$

We now consider $w \in \mathcal{W}_2$ such that G_w is not a tree. In this case $V_w \le E_w$. We have proved earlier that under $\mathbb{P}_{0,n}$,

$$\mathrm{Var}\left[\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_w X_w\right] \to 0.$$

Now for a typical w in this class

$$\mathrm{E}_{\mathbb{P}_{1,n}}\left[|X_w|\right] \le \left(p_{n,\text{av}} + \frac{\sqrt{p_{n,\text{av}}}}{\sqrt{n}}\right)^{\#E_w} \le \left(1 + \frac{1}{\sqrt{np_{n,\text{av}}}}\right)^k p_{n,\text{av}}^{\#E_w}.$$

So fixing any equivalence class $\mathbf{w}$ we have

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathbf{w}} \mathrm{E}_{\mathbb{P}_{1,n}}\left[|X_w|\right] \le \left(1+\frac{1}{\sqrt{np_{n,\mathrm{av}}}}\right)^k \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k n^{\#V_{\mathbf{w}}} p_{n,\mathrm{av}}^{\#E_w}.$$

Further from Lemma 15 we have fixing $\#V_{\mathbf{w}} = t$, the cardinality of $\mathbf{w}$'s are bounded by $2^{2k}(2k)^{3(2k-2t+1)}$. If $\#E_w < k$ implying $\#V_w < k$, then the sum in this case

$$\sum_{t=1}^{k-1} \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathbf{w}\ |\ \#V_{\mathbf{w}}=t} \mathrm{E}_{\mathbb{P}_{1,n}}\left[|X_w|\right] \le \sum_{t=1}^{k-1} 2^{2k}(2k)^3 \left(\frac{(2k)^6}{np_{n,\mathrm{av}}}\right) \to 0. \quad (4.6.112)$$

Finally when $\#E_w = \#V_w = k$ which implies every edge is repeated exactly twice we have

$$\mathrm{E}_{\mathbb{P}_{0,n}}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w\right] = \frac{1}{n^k}.$$

So

$$\sum_{w\in\mathbf{w}} \mathrm{E}_{\mathbb{P}_{0,n}}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w\right] = 1.$$

Since each random variable is repeated exactly twice we have

$$\left(\left(p_{n,\mathrm{av}} - \frac{\sqrt{p_{n,\mathrm{av}}}}{n}\right)(1-p_{n,\mathrm{av}}-d)\right)^k \le \mathrm{E}_{\mathbb{P}_{1,n}}\left[X_w\right] \le \left((p_{n,\mathrm{av}}+d)\,(1-p_{n,\mathrm{av}}+d)+d^2\right)^k$$

So

$$\left|\sum_{w\in\mathbf{w}} \mathrm{E}_{\mathbb{P}_{0,n}}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w\right] - \sum_{w\in\mathbf{w}} \mathrm{E}_{\mathbb{P}_{1,n}}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w\right]\right| = o(1).$$

In particular, one can prove that the error term in above equation is actually of order $\left(\frac{k^2}{\sqrt{np_{n,\mathrm{av}}}}\right)$. Further there are at most $2^{2k}(2k)^3$ many such words As a consequence, under

$\mathbb{P}_{1,n}$ we can write

$$\sum_w \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w - \mathrm{E}_{\mathbb{P}_{0,n}}\left[\sum_w \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w\right]$$
$$= \sum_w \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w - \mathrm{E}_{\mathbb{P}_{1,n}}\left[\sum_w \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w\right] + O\left(\frac{2^{2k}(2k)^5}{\sqrt{np_{n,\mathrm{av}}}}\right). \tag{4.6.113}$$

The error term converges to 0. Now by the same argument as the null case, we can prove

$$\mathrm{Var}_{\mathbb{P}_{1,n}}\left[\sum_w \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k X_w\right] \to 0.$$

In what follows, we shall only focus on any word w such that G_w is a tree and shall only give the analysis of the mean. All the other cases follow from the arguments similar to the proof of part (i) of Theorem 8 with suitable modifications for the trees described here. We will also use notations defined at the beginning of the proof of part (i) of Theorem 8. Throughout this part, all expectation and variance are taken with respect to $\mathbb{P}_{1,n}$ conditioning on the group assignment σ_i for $1 \le i \le n$.

We at first fix any word $w \in \mathcal{W}_2$ such that G_w is a tree. Recall that for a word $w = (i_0, \ldots, i_{2k})$ we have

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2: G_w \text{ is a tree}} X_w$$
$$= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2: G_w \text{ is a tree}} \prod_{j=0}^{2k-1} (x_{i_j,i_{j+1}} - p_{n,\mathrm{av}})$$
$$= \left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2: G_w \text{ is a tree}} \left[\prod_{j=0}^{2k-1} (x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}) + d^{2k} + V_{n,w}\right]$$

where

$$V_{n,w} = \sum_{\emptyset \subsetneq \mathcal{E}_T \subsetneq \mathcal{E}_{2k}} \prod_{e \in \mathcal{E}_T} (\sigma_{e(w)} d) \prod_{e \in \mathcal{E}_L} (x_{e(w)} - p_{e(w)})$$
$$= \sum_{\emptyset \subsetneq \mathcal{E}_T \subsetneq \mathcal{E}_{2k}} d^{\#\mathcal{E}_T} \prod_{e \in \mathcal{E}_T} \sigma_{e(w)} \prod_{e \in \mathcal{E}_L} (x_{e(w)} - p_{e(w)}).$$

Since the graph corresponding to any word $w \in \mathcal{W}_2$ has the number of vertices less than or equal to k, it is easy to see that

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w \in \mathcal{W}_2 : G_w \text{ is a tree}} d^{2k} \to 0.$$

Arguing as before, we have $\mathrm{E}\left[V_{n,w}\right] \neq 0$ only if each edge in $E(\mathcal{E}_L(w))$ has been repeated at least twice by the exploration $e \in \mathcal{E}_L$. We shall only focus on this case. Now fix $\emptyset \subsetneq \mathcal{E}_T \subsetneq \mathcal{E}_{2k}$ and an equivalence class $\mathbf{w} \subset \mathcal{W}_2$ corresponding to graph $G = (V, E)$ such that for any $w \in \mathbf{w}$, G_w is a tree. Arguing as (4.6.50)-(4.6.51) one arrives at the following upper bound:

$$\begin{aligned}&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w : w \in \mathbf{w}} \mathrm{E}\left[\left| d^{\#\mathcal{E}_T} \prod_{e \in \mathcal{E}_T} \sigma_{e(w)} \prod_{e \in \mathcal{E}_L} \left(x_{e(w)} - p_{e(w)}\right)\right|\right] \\ &\leq C^k \left(\frac{1}{n}\right)^{k - \#V + \frac{\#\mathcal{E}_T}{2}} \left(\frac{1}{p_{n,\mathrm{av}}}\right)^{k - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2}}.\end{aligned} \tag{4.6.114}$$

Here $w \in \mathbf{w}$ is any word. Observe that in this case G_w is a tree. So we require a slight modification of (4.6.52) and (4.6.53) in the current scenario. Firstly

$$\begin{aligned}&\left(k - \#V + \frac{\#\mathcal{E}_T}{2}\right) - \left(k - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2}\right) \\ &= \#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T - \#V\end{aligned} \tag{4.6.115}$$

as before. Now

$$\#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T = \sum_{\gamma \in E_w} \left[\mathbb{I}_{\gamma \in E(\mathcal{E}_L(w))} + \sum_{e \in \mathcal{E}_T} \mathbb{I}_{\gamma = e(w)}\right]. \tag{4.6.116}$$

Since $\emptyset \subsetneq \mathcal{E}_T$ and $\mathcal{E}_L \cup \mathcal{E}_T = \mathcal{E}_{2k}$, we have for all γ

$$\mathbb{I}_{\gamma \in E(\mathcal{E}_L(w))} + \sum_{e \in \mathcal{E}_T} \mathbb{I}_{\gamma = e(w)} \geq 1 \tag{4.6.117}$$

and there exists at least one γ such that (4.6.117) is greater than equal to 2. As a consequence, (4.6.116) is greater than or equal to $\#E_w + 1$. So the final expression of (4.6.115) is always greater than equal to 0. Observe that the equality happens only if $\mathcal{E}_T$ is either exactly equal to both the traversal of an edge γ traversed exactly twice or $\#\mathcal{E}_T = 1$ and the corresponding edge has been traversed at least four times.

Now

$$2k = \#\mathcal{E}_L + \#\mathcal{E}_T = \sum_{\gamma \in E_w} \left[\sum_{e \in \mathcal{E}_L} \mathbb{I}_{\gamma = e(w)} + \sum_{e \in \mathcal{E}_T} \mathbb{I}_{\gamma = e(w)} \right]$$

Arguing similarly as (4.6.52) we always have

$$k - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2} \geq 0. \tag{4.6.118}$$

However here we prove that (4.6.115) and (4.6.118) can not be 0 simultaneously. Observe that as $\mathbf{w} \subset \mathcal{W}_2$, all the edges in G is traversed at least twice and at least one edge is traversed at least four times. As a consequence, we have for all γ we have

$$\sum_{e \in \mathcal{E}_L} \mathbb{I}_{\gamma = e(w)} + \sum_{e \in \mathcal{E}_T} \mathbb{I}_{\gamma = e(w)} \geq 2$$

and there exists at least one γ such that the above sum is greater than equal to 4. Now consider the cases when (4.6.115) is 0. In this case we have

$$\begin{aligned} \#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T - \#V &= 0 \\ \#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T &= \#E + 1. \end{aligned} \tag{4.6.119}$$

There can be two cases either, $E\left(\mathcal{E}_L(w)\right) \cap E\left(\mathcal{E}_T\right) = \emptyset$ or $E\left(\mathcal{E}_L(w)\right) \cap E\left(\mathcal{E}_T\right) \neq \emptyset$. If $E\left(\mathcal{E}_L(w)\right) \cap E\left(\mathcal{E}_T\right) = \emptyset$, then $\#\mathcal{E}_T = 2$ and there exists a single edge γ

$$\sum_{e \in \mathcal{E}_T} \mathbb{I}_{\gamma = e(w)} = 2.$$

However in this case

$$\#E\left(\mathcal{E}_L(w)\right) + \frac{\#\mathcal{E}_T}{2} = \#E < k.$$

Now consider the second case where $E\left(\mathcal{E}_L(w)\right) \cap E\left(\mathcal{E}_T(w)\right) \neq \emptyset$. If $E\left(\mathcal{E}_T(w)\right) \neq E\left(\mathcal{E}_L(w)\right) \cap E\left(\mathcal{E}_T(w)\right)$, then there exists at least one edge $\gamma \in E\left(\mathcal{E}_L(w)\right)$ such that

$$\sum_{e \in \mathcal{E}_T} \mathbb{I}_{\gamma = e(w)} \geq 1$$

and

$$\begin{aligned} &\sum_{\gamma \in E(\mathcal{E}_T(w)) \setminus E(\mathcal{E}_L(w))} \mathbb{I}_{\gamma = e(w)} \\ &\geq 2\#\left(E\left(\mathcal{E}_T(w)\right) \setminus E\left(\mathcal{E}_L(w)\right)\right) \\ &\geq \#\left(E\left(\mathcal{E}_T(w)\right) \setminus E\left(\mathcal{E}_L(w)\right)\right) + 1 \end{aligned} \tag{4.6.120}$$

This leads to a contradiction to the fact that

$$\#E(\mathcal{E}_L(w)) + \#\mathcal{E}_T = \#E + 1. \tag{4.6.121}$$

So we have

$$E\left(\mathcal{E}_T(w)\right) = E\left(\mathcal{E}_L(w)\right) \cap E\left(\mathcal{E}_T(w)\right).$$

So

$$E\left(\mathcal{E}_L(w)\right) = E_w.$$

Hence $\#\mathcal{E}_T(w) = 1$. Since $\#E\left(\mathcal{E}_L(w)\right) < k$ we have

$$E\left(\mathcal{E}_L(w)\right) + \frac{1}{2}\#\mathcal{E}_T(w) \leq k - \frac{1}{2}.$$

Hence

$$k - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2} \geq \frac{1}{2}.$$

In the case of equality in (4.6.115), we analyze the terms one by one. Let

$$k - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2} = l.$$

Then either $\#\mathcal{E}_T = 1$ or 2. In the first case $E(\mathcal{E}_L(w)) = E(w)$. Hence

$$k - \frac{1}{2} - l = \#E(w).$$

Since G_w is a tree, we have $\#V(w) = k + \frac{1}{2} - l$. By Lemma 15 we have the number of equivalence class of this kind is bounded by $2^{2k+1}(2k+1)^{3(l+2)}$. Now given an equivalence class the number of choices of $\mathcal{E}_T$ is bounded by $2k$. Hence the mean is bounded by

$$\sum_{l \geq \frac{1}{2}} 2^{2k+1} 2k \frac{(2k+1)^{6(l+1)}}{\left(np_{n,\text{av}}\right)^l} \to 0.$$

Now consider the case when $\#\mathcal{E}_T = 2$. Here $\#E(\mathcal{E}_L(w)) + \frac{\#\mathcal{E}_T}{2} = \#E(w)$. Let

$$k - \#E(\mathcal{E}_L(w)) - \frac{\#\mathcal{E}_T}{2} = l$$

then

$$\#V(w) = k - l + 1.$$

Also given an equivalence class the number of choices of $\#\mathcal{E}_T \leq (2k)^2$. Hence similar calculation gives the mean corresponding to this case is bounded by

$$\sum_{l\geq 1} 2^{2k+1}(2k)^2 \frac{(2k+1)^{3(2l+3)}}{(np_{n,\mathrm{av}})^l} \to 0.$$

In all the other cases the contribution is bounded by $\frac{C_1 k^{C_2}}{n}$ for some deterministic constants C_1, C_2.

For the variance part arguing similarly as

Thus, we always have the right side of (4.6.114) converging to 0 as $np_{n,\mathrm{av}} \to \infty$. This proves

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2\,:\,G_w\ G_w \text{ is a tree}} \mathrm{E}\left[V_{n,w}\right] \to 0.$$

Proving

$$\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_2\,:\,G_w\ G_w \text{ is a tree}} V_{n,w}\right] \leq \left(\frac{(C_1'k)^{C_2'k}}{n}\right)$$

is similar to the analysis of the mean. Here one needs to use the embedding algorithm to concatenate two words with at least one common edge. However in this case one can check that one gets always the power of n higher than power of $p_{n,\mathrm{av}}$ in the denominator. We omit the details.

Alternative analysis of $\sum_{w\in\mathcal{W}_1} X_w$ It was proved in **??** that under $\mathbb{P}_{0,n}$

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1}\left[X_w - \mathrm{E}_{\mathbb{P}_{0,n}}\left[X_w\right]\right] - k\psi_{2k}\left(\mathsf{Tr}\left[A^2_{\mathrm{cen1}}\right] - \mathrm{E}_{\mathbb{P}_{0,n}}\left[\mathsf{Tr}\left[A^2_{\mathrm{cen1}}\right]\right]\right) \xrightarrow{p} 0 \tag{4.6.122}$$

as long as $np^2_{n,\mathrm{av}} \to \infty$.

The main goal of this part of this paper is to justify (4.6.122) holds true even under $\mathbb{P}_{1,n}$. Observe that the constants in (4.6.122) remains unchanged even under the alternative. So there are two main parts to be proved. Firstly under $\mathbb{P}_{1,n}$

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} \left[X_w - \mathrm{E}_{\mathbb{P}_{1,n}}\left[X_w\right]\right] - k\psi_{2k}\left(\mathsf{Tr}\left[A^2_{\text{cen1}}\right] - \mathrm{E}_{\mathbb{P}_{1,n}}\left[\mathsf{Tr}\left[A^2_{\text{cen1}}\right]\right]\right) \xrightarrow{p} 0 \tag{4.6.123}$$

and then

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} \left[\mathrm{E}_{\mathbb{P}_{1,n}}\left[X_w\right] - \mathrm{E}_{\mathbb{P}_{0,n}}\left[X_w\right]\right] \to 0. \tag{4.6.124}$$

Like before we write

$$\begin{aligned}
&\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} X_w \\
&= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} \prod_{j=0}^{2k-1} \left(x_{i_j,i_{j+1}} - p_{n,\text{av}}\right) \\
&= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} \left[\prod_{j=0}^{2k-1} \left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right) + d^{2k} + V_{n,w}\right]
\end{aligned}$$

where

$$\begin{aligned}
V_{n,w} &= \sum_{\emptyset\subsetneq\mathcal{E}_T\subsetneq\mathcal{E}_{2k}} \prod_{e\in\mathcal{E}_T} \left(\sigma_{e(w)} d\right) \prod_{e\in\mathcal{E}_L} \left(x_{e(w)} - p_{e(w)}\right) \\
&= \sum_{\emptyset\subsetneq\mathcal{E}_T\subsetneq\mathcal{E}_{2k}} d^{\#\mathcal{E}_T} \prod_{e\in\mathcal{E}_T} \sigma_{e(w)} \prod_{e\in\mathcal{E}_L} \left(x_{e(w)} - p_{e(w)}\right).
\end{aligned}$$

First of all there are $(1+o(1))n^{k+1}$ many words in $\mathcal{W}_1$ and $d = O\left(\sqrt{\frac{p_{n,\text{av}}}{n}}\right)$. So

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} d^{2k} = \psi_{2k} O\left(\frac{n^{k+1}p^k_{n,\text{av}}}{n^{2k}p^k_{n,\text{av}}}\right) = \psi_{2k} O\left(\frac{1}{n^{k-1}}\right)$$

which goes to 0 as long as $k \geq 2$.

By arguments similar to Subsection 4.6.4 one can prove

$$\mathrm{Var}\left[\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathcal{W}_1} V_{n,w}\right] \to 0.$$

Now fix $\emptyset \subsetneq \mathcal{E}_T \subsetneq \mathcal{E}_{2k}$ and an equivalence class $\mathbf{w} \subset \mathcal{W}_1$. Arguing as (4.6.50)-(4.6.51) one arrives at the following upper bound:

$$\begin{aligned}&\left|\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w:w\in\mathbf{w}} \mathrm{E}\left[d^{\#\mathcal{E}_T}\prod_{e\in\mathcal{E}_T}\sigma_{e(w)}\prod_{e\in\mathcal{E}_L}\left(x_{e(w)}-p_{e(w)}\right)\right]\right| \\ &\le C^k\left(\frac{1}{n}\right)^{k-\#V+\frac{\#\mathcal{E}_T}{2}}\left(\frac{1}{p_{n,\mathrm{av}}}\right)^{k-\#E(\mathcal{E}_L(w))-\frac{\#\mathcal{E}_T}{2}}.\end{aligned} \tag{4.6.125}$$

In order to get nonzero expectation in the first expression of (4.6.125), one needs every edge in $E(\mathcal{E}_L(w))$ to be traversed at least twice by the word w. We shall only consider this case. Since every edge in $E(\mathcal{E}_L(w))$ to be traversed at least twice by the word w

$$k - \#E\left(\mathcal{E}_L(w)\right) - \frac{\#\mathcal{E}_T}{2} \ge 0.$$

Now arguing as Subsection 4.6.4 we again have

$$\#E\left(\mathcal{E}_L(w)\right) + \#\mathcal{E}_T(w) - \#V \ge 0.$$

Here the equality occurs if and only if there exists an unique edge $\gamma \in E_w$ such that

$$\sum_{e\in\mathcal{E}_T} \mathbb{I}_{\gamma=e(w)} = 2$$

and for all other $\gamma \in E_w$

$$\sum_{e\in\mathcal{E}_T} \mathbb{I}_{\gamma=e(w)} = 0.$$

Observe that in this case

$$k - \#V + \frac{\#\mathcal{E}_T}{2} = 0.$$

As a consequence, this gives a nontrivial contribution to the mean. Now given any $w \in \mathbf{w}$ there exists exactly k choices of $\mathcal{E}_T$ such that this condition hold true. Namely one needs to take both the instances when a single edge is traversed for each edge. Since each edge is traversed exactly twice and

$$\left(\frac{(p_{n,\text{av}}+d)(1-p_{n,\text{av}}+d)}{(p_{n,\text{av}})(1-p_{n,\text{av}})}\right)^k = \left(1+O(k/\sqrt{np_{n,\text{av}}})\right)$$

and

$$\left(\frac{(p_{n,\text{av}}-d)(1-p_{n,\text{av}}-d)}{(p_{n,\text{av}})(1-p_{n,\text{av}})}\right)^k = \left(1+O(k/\sqrt{np_{n,\text{av}}})\right),$$

we have

$$\mathrm{E}\left[\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1} V_{n,w}\right] = \psi_{2k}kt^2\left(1+O(k/\sqrt{np_{n,\text{av}}})\right)\left(1+O(k^2/n)\right).$$

We now fix an equivalence class $\mathbf{w}$. There are ψ_{2k} many of them. We now prove for any fixed equivalence class $\mathbf{w}$,

$$\begin{aligned}&\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathbf{w}}\left[\prod_{\gamma\in E_w}(p_\gamma(1-p_\gamma))\right]\\ &-\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathbf{w}}(p_{n,\text{av}}(1-p_{n,\text{av}}))^k + kt^2 \xrightarrow{p} 0.\end{aligned} \tag{4.6.126}$$

where for any $\gamma=\{i,j\}\in E_w$ $p_\gamma = p_{n,\text{av}}+d\sigma_i\sigma_j$. Here the randomness of σ is taken into consideration. Now for any w, we have

$$\begin{aligned}\prod_{\gamma\in E_w}(p_\gamma(1-p_\gamma)) &= \prod_{\gamma\in E_w}((p_{n,\text{av}}+d\sigma_\gamma)(1-p_{n,\text{av}}-d\sigma_\gamma))\\ &= \prod_{\gamma\in E_w}\left(p_{n,\text{av}}(1-p_{n,\text{av}})+d(1-2p_{n,\text{av}})\sigma_\gamma - d^2\right)\\ &= (p_{n,\text{av}}(1-p_{n,\text{av}}))^k - kd^2\left(p_{n,\text{av}}(1-p_{n,\text{av}})\right)^{k-1} + \mathrm{Er}(w).\end{aligned} \tag{4.6.127}$$

Here the second term comes from taking $-d^2$ for exactly one of the edges and $p_{n,\text{av}}(1-p_{n,\text{av}})$

along all the other edges and $\mathrm{Er}(w)$ is the sum of all the remaining terms after expanding the product. Using the final expression of (4.6.127), we need to prove

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathbf{w}} \mathrm{Er}(w) \xrightarrow{p} 0$$

in order to complete the proof of (4.6.126). Now given w we order the edges in E_w as $\gamma_1,\ldots,\gamma_k$

$$\mathrm{Er}(w) = \sum_{S_1,S_2,S_3\,:\,\#S_1\neq k;(\#S_1,\#S_3)\neq(k-1,1)} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#S_1} (-d^2)^{\#S_3} \prod_{j\in S_2} \left(d\sigma_{\gamma_j}\right). \tag{4.6.128}$$

Here S_1, S_2, S_3 are subsets of $\{1,\ldots,k\}$ We now analyze the r.h.s. of (4.6.128) case by case.
Case 1: $\#S_2 = 0$ In this case we have $\#S_3 \geq 2$. We know $d^2 = O\left(\frac{p_{n,\mathrm{av}}}{n}\right)$. As a consequence for any given S_1, S_3 of this kind we have

$$(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#S_1} (-d^2)^{\#S_3} = O\left(p_{n,\mathrm{av}}^k/n^{\#S_3}\right). \tag{4.6.129}$$

As there are $n^{k+1}\left(1+O(k^2/n)\right)$ many words in any equivalence class $\mathbf{w}$, we have

$$\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^k \sum_{w\in\mathbf{w}} \left|\frac{p_{n,\mathrm{av}}^k}{n^{\#S_3}}\right| \leq O\left(\frac{1}{n^{\#S_3-1}}\right) \leq O\left(\frac{1}{n}\right).$$

Finally there are at most 2^k choices of S_3 and once S_3 is chosen S_1 is automatically fixed in this case. Hence the contribution in this case is of order $\left(\frac{2^k}{n}\right)$.
Case 2: $\#S_2 > 0$ Since d^2 is a non-random quantity and $d^2 = O\left(\frac{p_{n,\mathrm{av}}}{n}\right) \ll p_{n,\mathrm{av}}$, we shall only focus on the case when $\#S_3 = 0$. Now for any word w;

$$\begin{aligned}&(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#S_1} \prod_{j\in S_2}\left(d\sigma_{\gamma_j}\right)\\ &= t^{\#S_2} (p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}))^{\#S_1} \left(\sqrt{\frac{p_{n,\mathrm{av}}}{n}}\right)^{\#S_2} \prod_{j=1}^{l(S_2)} \sigma_{i_j}.\end{aligned} \tag{4.6.130}$$

Here the number $l(S_2)$ depends on the set S_2 and for each j, $i_j \in V_w$. Now

$$\begin{aligned}\sum_{w\in\mathbf{w}}\prod_{j=1}^{l(S_2)}\sigma_{i_j} &= n^{k+1-l(S_2)}\sum_{i_1,\ldots,i_{l(S_2)}}\prod\sigma_{i_j} + O(n^{k-l(S_2)})\sum_{i_1,\ldots,i_{l(S_2)}} 1 \\ &= O_p\left(\sqrt{n}+k\right)^{l(S_2)} n^{k+1-l(S_2)}.\end{aligned} \tag{4.6.131}$$

Here we have used the fact that $\sum_{i=1}^{n}\sigma_i = O_P(\sqrt{n})$. Since G_w is a tree, we have $l(S_2) \geq 2$. As a consequence,

$$\begin{aligned}&\sum_{w\in\mathbf{w}}\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k t^{\#S_2}\left(p_{n,\text{av}}(1-p_{n,\text{av}})\right)^{\#S_1}\left(\sqrt{\frac{p_{n,\text{av}}}{n}}\right)^{\#S_2}\prod_{j=1}^{l(S_2)}\sigma_{i_j} \\ &= \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k t^{\#S_2}\left(p_{n,\text{av}}(1-p_{n,\text{av}})\right)^{\#S_1}\left(\sqrt{\frac{p_{n,\text{av}}}{n}}\right)^{\#S_2} O_P(n^k) \\ &= O_P\left(\frac{1}{np_{n,\text{av}}}\right)^{\#S_2/2}.\end{aligned} \tag{4.6.132}$$

Since there are at most 4^k possible choices of (S_1, S_2, S_3) corresponding to this case, we have the required result. The proof of the fact

$$\text{Var}\left[\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^k \sum_{w\in\mathcal{W}_1}\left[\prod_{j=0}^{2k-1}\left(x_{i_j,i_{j+1}} - p_{i_j,i_{j+1}}\right)\right] - k\psi_{2k}\,\mathsf{Tr}\left[A_{\text{cen1}}^2\right]\right] \to 0$$

as $np_{n,\text{av}}^2 \to \infty$ is same as the null case. Hence we omit the details. □

4.6.5. Proof of Theorem 10

Proof of part (i)

Recall that

$$f(m,r)\frac{m}{r} = \begin{cases}\binom{m}{\frac{m+r}{2}} & \text{whenever } m-r \text{ even}\\ 0 & \text{otherwise.}\end{cases}$$

Observe that $\binom{m}{\frac{m+r}{2}}$ is the coefficient of $(z^r + 1/z^r)$ in the expansion of $(z+1/z)^m$. Recall the definition of $\mathbb{D}_{2k+1}$ in (4.3.10). We further define

$$u_{2k+1} := \left(\binom{3}{2}, \binom{5}{3}, \ldots, \binom{2k+1}{k+1}\right)'.$$

Then

$$\begin{pmatrix} 1 & 0 \\ u_{2k+1} & \mathbb{D}_{2k+1} \end{pmatrix} \left(z + \frac{1}{z}, z^3 + \frac{1}{z^3}, \ldots, z^{2k+1} + \frac{1}{z^{2k+1}}\right)'$$
$$= \left(\left(z+\frac{1}{z}\right), \left(z+\frac{1}{z}\right)^3, \ldots, \left(z+\frac{1}{z}\right)^{2k+1}\right)'.$$

Hence

$$\left(z + \frac{1}{z}, z^3 + \frac{1}{z^3}, \ldots, z^{2k+1} + \frac{1}{z^{2k+1}}\right)'$$
$$= \begin{pmatrix} 1 & 0 \\ -\mathbb{D}_{2k+1}^{-1} u_{2k+1} & \mathbb{D}_{2k+1}^{-1} \end{pmatrix} \left(\left(z+\frac{1}{z}\right), \left(z+\frac{1}{z}\right)^3, \ldots, \left(z+\frac{1}{z}\right)^{2k+1}\right)'.$$

On the other hand, we have defined $P_{2k+1}(\cdot)$ to be such that

$$P_{2k+1}\left(z + \frac{1}{z}\right) = z^{2k+1} + \frac{1}{z^{2k+1}}.$$

As a consequence, we have

$$\mathbb{D}_{2k+1}^{-1}(k, j) = P_{2k+1}[2j+1].$$

The rest of the proof of part (i) is exactly similar to the proofs of part (iii) of Theorem 7 and part (iii) of Theorem 8. We thus omit the details.

Proof of part (ii)

Overall the proof here is similar to the proof of part (i). However, here we further prove $\mathrm{Var}\left(\sum_{r=0:r \text{ even}}^{2k} P_{2k}(r) T_r\right) \to 0$ when $np_{n,\mathrm{av}}^2 \to 0$ under both null and local alternative.

Firstly, under $\mathbb{P}_{0,n}$,

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{r=0:r \text{ even}}^{2k} P_{2k}(r)T_r\right) &\le (2k)^2 \sup_{r\le k} P_{2k}(2r)^2 \sup_{r\le k}\left(\frac{v_{1,2r}}{n^2 p_{n,\mathrm{av}}^3} + \frac{v_{2,2r}}{n p_{n,\mathrm{av}}^2}\right) \\
&\le (C_3' k)^{C_4'} (C_1')^{C_2' k} \frac{1}{n p_{n,\mathrm{av}}^2}
\end{aligned}
\tag{4.6.133}
$$

for some universal constants C_1', C_2', C_3' and C_4'. Here we have used the well known fact that $\sup_{r\le k} P_{2k}(2r) \le 4^{2k}$.

The proof under the alternative follows almost immediately with the help of the proof under the alternative of Theorem 9.

□

4.6.6. Proof of (4.3.14)

Consider the first identity first. Note that by definition for any integer $r \ge 0$,

$$\psi_{2r} = \int_{-2}^{2} x^{2r} \frac{\sqrt{4-x^2}}{2\pi}\, \mathrm{d}x.$$

Hence for any $k \ge 2$

$$
\begin{aligned}
\sum_{r=0}^{k} P_{2k}[2r]\psi_{2r} &= \int_{-2}^{2} \sum_{r=0}^{k} P_{2k}[2r] x^{2r} \frac{\sqrt{4-x^2}}{2\pi} \mathrm{d}x = \int_{-2}^{2} P_{2k}(x) \frac{\sqrt{4-x^2}}{2\pi} \mathrm{d}x \\
&= \int_{-2}^{2} P_{2k}(x)[P_0(x) - P_2(x)] \frac{1}{2\pi\sqrt{4-x^2}} \mathrm{d}x \\
&= 0.
\end{aligned}
$$

Here, the second equality holds since the odd power terms in P_{2k} are all zeros. The third equality holds since $P_0(x) = 2$ and $P_2(x) = x^2 - 2$. The last equality holds since the P_j's are mutually orthogonal with respect to the weight function $1/\sqrt{4-x^2}$ on $[-2, 2]$ by their definitions and the orthogonality of Chebyshev polynomials with respect to $1/\sqrt{1-x^2}$ on $[-1, 1]$.

Turn to the second identity. we have

$$\begin{aligned}\sum_{r=1}^{k} P_{2k}[2r](2r)\psi_{2r} &= \int_{-2}^{2} \left[\frac{\mathrm{d}}{\mathrm{d}x} P_{2k}(x)\right] x \frac{\sqrt{4-x^2}}{2\pi} \mathrm{d}x \\ &= P_{2k}(x) x \frac{\sqrt{4-x^2}}{2\pi}\Big|_{-2}^{2} - \int_{-2}^{2} P_{2k}(x) \frac{\mathrm{d}}{\mathrm{d}x}\left[x \frac{\sqrt{4-x^2}}{2\pi}\right] \mathrm{d}x \\ &= 0 + \int_{-2}^{2} P_{2k}(x) P_2(x) \frac{1}{\pi\sqrt{4-x^2}} \mathrm{d}x = 0.\end{aligned}$$

Here we have used the fact that $P_2(x) = x^2 - 2$. This completes the proof.

CHAPTER 5 : Non backtracking matrices and optimal hypothesis testing for planted partition models with growing degrees

5.1. overview

We revisit the problem of testing Erdős-Renyi graph model against a stochastic block model in the asymptotic regime where the average degree of the graph grows to infinity with the graph size n. We are interested in the stochastic block model with two clusters and asymptotically equal cluster sizes (the planted partition model) and the regime when the probability measures induced by the Erdős-Renyi graph model and the stochastic block model are mutually contiguous. We prove that in the contiguous regime there exists a test computable in polynomial time which achieves the optimal power whenever $np_{n,\mathrm{av}} \to \infty$. Here $p_{n,\mathrm{av}}$ is the average connection probability. This problem was studied in Banerjee and Ma (2017a) where a test based on linear spectral statistics was proposed. This test achieves the optimal power only when the graph is sufficiently dense i.e. $np_{n,\mathrm{av}}^2 \to \infty$. In this paper this technical difficulty is circumvented by doing spectral analysis of a variant of non backtracking matrix.

5.2. Introduction

Stochastic block model (SBM) Holland et al. (1983) is an active domain of modern research in statistics, computer science and many other related fields. A stochastic block model for random graphs encodes a community structure where a pair of nodes from the same community are expected to be connected in a different manner from those from different communities. This model, together with the related community detection problem, has drawn substantial attentions in statistics and machine learning. Throughout the paper, let $\mathcal{G}_1(n, p_n)$ denote the Erdős-Renyi graph with n nodes in which the edges are i.i.d. Bernoulli random variables with success probability p_n. In this paper we consider a special case of the stochastic block model, the planted partition model which is defined as follows:

Definition 28. For $n \in \mathbb{N}$, and $p, q \in [0, 1]$ let $\mathcal{G}(n, p, q)$ denote the model of random,±

labelled graphs in which each vertex u is assigned (independently and uniformly at random) a label $\sigma_u \in \{\pm 1\}$ and each edge between u and v are included independently with probability p if they have the same label and with probability q if they have different labels.

For any graph G distributed as $\mathcal{G}(n, p_n)$ or $\mathcal{G}(n, p_n, q_n)$, we denote $p_{n,\mathrm{av}}$ to be the average connection probability i.e. $p_{n,\mathrm{av}} = 1/\left(n(n-1)\right)\sum_{i,j} \mathrm{E}[x_{i,j}]$ where $x_{i,j}$ is the indicator random variable corresponding to the edge between i and j th node. A fundamental question related to stochastic block models is community detection where one aims to recover the partition of nodes into communities based on one instance of the random graph. There has been a huge literature in this regard. In the asymptotic regime of bounded degrees (i.e. np_n and nq_n remain constants as $n \to \infty$), the seminal papers by Mossel et al. Mossel et al. (2015); Mossel et al. (2013) and Massoulié (2013) established sharp threshold for $\mathcal{G}_2(n, p_n, q_n)$ on when it is possible and impossible to achieve a partial recovery of community labels that is strictly better than random guessing, which confirmed the conjecture in Decelle et al. (2011). A lot of outstanding research have been done on the community detection problem after that in the sparse regime where the problems of multiple blocks, different types of recovery problems have been addressed. We will not be able to cite all these research since our primary interest is the dense regime where $np_{n,\mathrm{av}} \to \infty$ and we are interested in the hypothesis testing problem rather that community detection. However, one might have a look at the survey paper by Abbe (2017) for a survey of the recent literature.

The fundamental focus of this paper is the following testing problem:

$$H_0 : A \sim \mathcal{G}_1\left(n, \frac{p_n + q_n}{2}\right) \quad \text{vs.} \quad H_1 : A \sim \mathcal{G}_2(n, p_n, q_n) \tag{5.2.1}$$

when the average degree of the random graph grows to infinity with the graph size. The parameters in the hypotheses are so chosen that the expected numbers of edges match under null and alternative. Let $a_n = np_n$ and $b_n = nq_n$. Our primary interest lies in the cases where the signal-to-noise ratio

$$c := \frac{(a_n - b_n)^2}{a_n + b_n} \tag{5.2.2}$$

is a constant, and we call any such alternative a *local* one. For such cases, one has growing average degree if and only if $np_n \to \infty$. In what follows, we denote the null and alternative hypotheses in (5.2.1) by $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ respectively.

This problem also has received some amount of attention in the literature. Bickel and Sarkar (2016) proposed to recursively apply the largest eigenvalue test for partitioning the nodes and for determining κ. Here κ is the block size in a multi-community stochastic block model. The proposal was based on the GOE Tracy–Widom limit Tracy and Widom (1996) of the largest eigenvalue distribution for adjacency matrices of Erdős-Renyi graphs when the average degree grows linearly with n. Lei (2016) extended it to a procedure based on sequential largest eigenvalue tests in the regime where exact recovery can be achieved. See also Le and Levina (2015) for another spectral method for choosing κ.

In the dense regime the results in Mossel et al. (2015) were generalized by Banerjee (2018) when the average degrees grow to infinity (i.e. $np_{n,\mathrm{av}} \to \infty$). One of the fundamental results in Banerjee (2018) is to prove the probability models induced by Erdős-Renyi graph with n and that of a planted partition model are asymptotically mutually contiguous whenever $(a_n - b_n)^2 < 2(1-p)(a_n + b_n)$ and they are mutually singular when $(a_n - b_n)^2 > 2(1-p)(a_n + b_n)$ asymptotically. Further it is impossible to achieve a partial recovery of community labels that is strictly better than random guessing whenever $(a_n - b_n)^2 < 2(1-p)(a_n + b_n)$. Here $a_n := p_n/n$ and $b_n := q_n/n$ and one assumes $p_{n,\mathrm{av}} \to p \in [0, 1)$. The proof of mutual contiguity in Banerjee (2018) is based on proving an asymptotic decomposition result of the log likelihood ratio which can be written as a linear combination of a collection of statistics called the signed cycles(To be defined later). This allows one to construct the likelihood ratio test in this scenario provided the signed cycles statistics can be computed efficiently. This problem was partially solved in Banerjee and Ma (2017a) where the authors proved that the signed cycles statistics can be approximated by some special linear spectral statistics provided the graph is dense enough (i.e. $np_{n,\mathrm{av}}^2 \to \infty$).

The main focus of this paper is to solve this problem for the sparsity regime $np_{n,\mathrm{av}} \to \infty$.

The approach in this paper is to study the spectral properties of a special variant of the non backtracking matrix which will be defined in the next section.

5.3. Preliminaries

In this section we at first define some preliminary notations. We first introduce some preliminary definitions and notation to be used throughout the paper. We let $\mathrm{E}_{i,n}$ and $\mathrm{Var}_{i,n}$ denote expectation and variance under $\mathbb{P}_{i,n}$ for $i = 0$ and 1. For any random graph G, its adjacency matrix will be denoted by A and $x_{i,j}$ (instead of $a_{i,j}$) will be used to denote the indicator random variable corresponding to an edge between the nodes i and j. We denote the expected average connection probability and its sample counterpart by

$$p_{n,\mathrm{av}} = \frac{1}{n(n-1)} \sum_{i \neq j} \mathrm{E}_{0,n}[x_{i,j}], \quad \text{and} \quad \widehat{p}_{n,\mathrm{av}} = \frac{1}{n(n-1)} \sum_{i \neq j} x_{i,j}. \tag{5.3.1}$$

Under out settings, $p_{n,\mathrm{av}}$ remains unchanged if we replace $\mathrm{E}_{0,n}$ with $\mathrm{E}_{1,n}$ in its definition. The signed cycle of length k of the graph G is defined to be

$$C_{n,k}(G) = \left(\frac{1}{\sqrt{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}} \right)^k \sum_{i_0, i_1, \ldots, i_{k-1}} (x_{i_0,i_1} - p_{n,\mathrm{av}}) \ldots (x_{i_{k-1}i_0} - p_{n,\mathrm{av}}) \tag{5.3.2}$$

where $i_0, i_1, \ldots, i_{k-1}$ are all distinct. We define the following centered and scaled versions of the adjacency matrix A. For any $n \in \mathbb{N}$, let $1_n = (1, \ldots, 1)' \in \mathbb{R}^n$ and I_n be the $n \times n$ identity matrix. Then

$$A_{\mathrm{cen1}} := \frac{A - p_{n,\mathrm{av}}(1_n 1_n' - I_n)}{\sqrt{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}\,, \tag{5.3.3}$$

and

$$A_{\mathrm{cen2}} := \frac{A - \widehat{p}_{n,\mathrm{av}}(1_n 1_n' - I_n)}{\sqrt{n\widehat{p}_{n,\mathrm{av}}(1-\widehat{p}_{n,\mathrm{av}})}}\,. \tag{5.3.4}$$

Note that A_{cen2} is completely data-driven. If A is a random instance of the Erdős-Renyi $\mathcal{G}_1(n, p_{n,\mathrm{av}})$, then A_{cen1} has zeros on the diagonal and the sub-diagonal entries (subject to symmetry) are i.i.d. with mean zero and variance $1/n$.

Now we define a set of rescaled Chebyshev polynomials. These polynomials are important for drawing the connection between signed cycles and the spectrum of adjacency matrix. The standard Chebyshev polynomial of first kind of degree m is denoted by $S_m(x)$ and can be defined by the identity

$$S_m\left(\cos(\theta)\right) = \cos(m\theta). \tag{5.3.5}$$

In this paper we use a slight variant of S_m, denoted by P_m and defined as

$$P_m(x) = 2S_m\left(\frac{x}{2}\right). \tag{5.3.6}$$

In particular, $P_m(2\cos(\theta)) = 2\cos(m\theta)$. It is easy to note that $P_m\left(z + z^{-1}\right) = z^m + z^{-m}$ for all $z \in \mathbb{C}$. One also notes that $P_m(\cdot)$ is even and odd whenever m is even or odd respectively.

5.3.1. Non backtracking matrices and their variants

Now we define the non backtracking matrix for a graph and one of its variant which is of the key interest in this paper. For any graph $H = (V, E)$, let $c : E \to \mathbb{R}$ be a set of possibly negative edge weights. Let E^o denote the set of directed edges $E^o = \{(i,j),(j,i) \ : \ \{i,j\} \in E\}$. We now define the weighted non-backtracking matrix as follows:

$$B_c((i,j),(i',j')) = \begin{cases} c(i,j,i',j') & \text{when } j = i' \text{ and } j' \neq i \\ 0 & \text{otherwise.} \end{cases} \tag{5.3.7}$$

Given any graph G, when $c(i,j) = 1$, B_c reduces to the traditional non-backtracking operator. However, we shall deal with the following variant in this case. For either null or alternative we consider H to be the complete graph and we define

$$c(i,j,i',j') = \sqrt{A_{\text{cen1}}(i,j)A_{\text{cen1}}(i',j')}. \tag{5.3.8}$$

This particular variant of non-backtracking matrices will be of use in this paper. We use the notation B_u to denote this matrix. When $p_{n,\mathrm{av}}$ is not known one can use $c(i,j) = \sqrt{A_{\mathrm{cen2}}(i,j)A_{\mathrm{cen2}}(i',j')}$ and denote the corresponding matrix by $\hat{B}_u$.

5.4. Results from Banerjee (2018) and Banerjee and Ma (2017a)

We now briefly introduce the results from Banerjee (2018) and Banerjee and Ma (2017a) which are the building blocks of this paper. As mentioned earlier, in the paper Banerjee (2018) a decomposition type result based on the signed cycles statistics of the log-likelihood for the testing problem (5.2.1) was given and it was proved in Banerjee and Ma (2017a) that the signed cycles can be approximated by linear spectral statistics when $np_{n,\mathrm{av}}^2 \to \infty$. We at first begin with the decomposition result from Banerjee (2018).

Theorem 12. *(Banerjee (2018)) Let us consider the testing problem* (5.2.1)*. When* $np_{n,\mathrm{av}} \to \infty$*,* $p_{n,\mathrm{av}} \to p \in [0,1)$ *and the signal to noise ratio* c_n *converges to* $c < 2(1-p)$*, we have*

$$\log(L_n) := \log\left(\frac{d\mathbb{P}_{1,n}}{d\mathbb{P}_{0,n}}\right)\Big|_{\mathbb{P}_{0,n}} \xrightarrow{d} \sum_{i=3}^{\infty} \frac{2t^i Z_i - t^{2i}}{4i} \tag{5.4.1}$$

where $Z_i \sim_{indep} N(0,2i)$ *which are the in distribution limit of the cycle statistics* $C_{n,i}$ *defined in* (5.3.2)*. Here*

$$t := \sqrt{\frac{c}{2(1-p)}}.$$

Now the fundamental question is to compute the statistics $C_{n,i}$ in polynomial time at least with high probability. Observe that from mutual contiguity of $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$, when $c < 2(1-p)$ it is enough to compute $C_{n,k}$ under $\mathbb{P}_{0,n}$. This problem was partially solved in Banerjee and Ma (2017a) where it was proved when $np_{n,\mathrm{av}}^2 \to \infty$, the statistics $C_{n,k}$ can be approximated by the Chebyshev polynomials. Following is the precise version:

Theorem 13. *Suppose* $n \to \infty$*,* t *is finite then the following hold:*

(i) If $n\hat{p} \to \infty$ *and* $k = o(\min(\log(n\hat{p}), \sqrt{\log n}))$*, then under both* $\mathbb{P}_{0,n}$ *and* $\mathbb{P}_{1,n}$*,*

$$C_{n,2k+1}(A_{\text{cen1}}) - \text{Tr}(P_{2k+1}(A_{\text{cen1}})) \overset{P}{\to} 0.$$

(ii) If $n\hat{p}^2 \to \infty$ *and* $k = o(\min(\log(n\hat{p}^2), \sqrt{\log n}))$*, then under both* $\mathbb{P}_{0,n}$ *and* $\mathbb{P}_{1,n}$*,*

$$C_{n,2k}(A_{\text{cen1}}) - \text{Tr}(P_{2k}(A_{\text{cen1}})) - \mu_n(\hat{p}) \overset{P}{\to} 0,$$

where

$$\mu_n(p_{n,\text{av}}) = \alpha_{1,2k} + \frac{\alpha_{2,2k}}{p_{n,\text{av}}}.$$

Here

$$\alpha_{1,2k} = 2^{2k-1} - \binom{2k}{k}\frac{5k+1}{2(k+1)} + \binom{k+1}{2}\psi_{2k} - 3\binom{2k}{k+2}$$

and

$$\alpha_{2,2k} = \binom{2k}{k+2}.$$

5.5. Our results

As it is mentioned earlier that the approximation results from Banerjee and Ma (2017a) hold true only when the graph is dense enough. This happens due to the following fact. When one takes the trace, one has to consider the walks on the trees (see Banerjee and Ma (2017a) for specific details) which is problematic when the graph is sparse. However when considers the non-backtracking variant this does not happen since one simply ignores the walks on the tree. This will be discussed more elaborately in the proof section. From now on we shall only consider the matrix B_u as defined in the previous section.

Theorem 14. *We have the following results under* $\mathbb{P}_{0,n}$*.*

(i) For any $1 \leq k = o\left(\log(np_{n,\mathrm{av}})\right)$ *(?), we have*

$$\mathsf{Tr}\left(B_u^{2k+1}\right) - \sum_{l=3\ :\ 2k+1-l\ \text{even}}^{2k+1} C_{n,l} \xrightarrow{p} 0.$$

(ii) For any $2 \leq k = o\left(\log(np_{n,\mathrm{av}})\right)$ *(?) and* k *even, we have*

$$\mathsf{Tr}\left(B_u^{2k}\right) - \sum_{l=4\ :\ 2k-l\ \text{even}}^{2k} C_{n,l} - (k-2) \xrightarrow{p} 0.$$

As a consequence, we have

1. *For any* $1 \leq k = o\left(\log(np_{n,\mathrm{av}})\right)$

$$C_{n,2k+1} - \left[\mathsf{Tr}\left(B_u^{2k+1}\right) - \mathsf{Tr}\left(B_u^{2k-1}\right)\right] \xrightarrow{p} 0.$$

2. *For* $2 \leq k = o\left(\log(np_{n,\mathrm{av}})\right)$ *(?), we have*

$$C_{n,2k} - \left[\mathsf{Tr}\left(B_u^{2k}\right) - \mathsf{Tr}\left(B_u^{2k-2}\right) - 1\right] \xrightarrow{p} 0.$$

When $p_{n,\mathrm{av}}$ is not known a natural estimate of $p_{n,\mathrm{av}}$ is given by $\widehat{p_{n,\mathrm{av}}}$. The results in Theorem 14 remains unchanged if we replace $p_{n,\mathrm{av}}$ by $\widehat{p_{n,\mathrm{av}}}$. In particular, if we use $\hat{B}_u$ instead of B_u.

The result under alternative follows from mutual contiguity of $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ when $t < 1$.

Corollary 2. *When* $t < 1$*, we have the following results under* $\mathbb{P}_{1,n}$.

1. *For any* $1 \leq k = o\left(\log(np_{n,\mathrm{av}})\right)$

$$C_{n,2k+1} - \left[\mathsf{Tr}\left(B_u^{2k+1}\right) - \mathsf{Tr}\left(B_u^{2k-1}\right)\right] \xrightarrow{p} 0.$$

2. For $2 \leq k = o\left(\log(np_{n,\text{av}})\right)$ (?) *and* k*, we have*

$$C_{n,k} - \left[\mathsf{Tr}\left(B_u^{2k}\right) - \mathsf{Tr}\left(B_u^{2k-2}\right) - 1\right] \xrightarrow{p} 0.$$

From now on we shall denote the polynomial $P_k(\cdot)$ as follows

$$P_k(x) = \begin{cases} x^k - x^{k-2} & \text{when } k \text{ is odd} \\ x^k - x^{k-2} - 1 & \text{when } k \text{ is even.} \end{cases} \tag{5.5.1}$$

Using approximation result stated in Theorem 14, we have the following computable test which achieves full power when $t < 1$.

Theorem 15. *Let us consider the test statistics*

$$L_n(t) = \sum_{k=3}^{m_n} \frac{2t^k\,\mathsf{Tr}\left(P_k\left(B_u\right)\right) - t^{2k}}{4k}$$

for any $m_n = o\left(\min\{\log(np_{n,\text{av}}), \sqrt{\log n}\}\right)$*. When* $t < 1$*, under null* $L_n \xrightarrow{d} N(-1/2\sigma(t)^2, \sigma(t)^2)$ *and under alternative* $L_n \xrightarrow{d} N(1/2\sigma(t)^2, \sigma(t)^2)$ *where*

$$\sigma(t)^2 = -\frac{1}{2}\left[-\log(1-t^2) - t^2 - \frac{t^4}{2}\right].$$

5.6. Proofs

5.6.1. Preliminaries

We start with defining some preliminary definitions and notations. We shall mainly follow the notations from Anderson and Zeitouni (2006) for the convenience of the proof.

Definition 29 ($\mathcal{S}$ words)**.** Given a set $\mathcal{S}$, an $\mathcal{S}$ letter s is simply an element of $\mathcal{S}$. An $\mathcal{S}$ word w is a finite sequence of letters $s_1 \ldots s_k$, at least one letter long. An $\mathcal{S}$ word w is *closed* if its first and last letters are the same. In this paper $\mathcal{S} = \{1, \ldots, n\}$ where n is the number of nodes in the graph.

Two $\mathcal{S}$ words w_1, w_2 are called *equivalent*, denoted $w_1 \sim w_2$, if there is a bijection on $\mathcal{S}$ that maps one into the other. For any word $w = s_1 \dots s_k$, we use $l(w) = k$ to denote the *length* of w, define the *weight* $wt(w)$ as the number of distinct elements of the set $s_1, \dots, s_k$ and the *support* of w, denoted by $\text{supp}(w)$, as the set of letters appearing in w. With any word w we may associate an undirected graph, with $wt(w)$ vertices and at most $l(w) - 1$ edges, as follows.

Definition 30 (Graph associated with a word)**.** Given a word $w = s_1 \dots s_k$, we let $G_w = (V_w, E_w)$ be the graph with set of vertices $V_w = \text{supp}(w)$ and (undirected) edges $E_w = \{\{s_i, s_{i+1}\}, i = 1, \dots, k-1\}$.

The graph G_w is connected since the word w defines a path connecting all the vertices of G_w, which further starts and terminates at the same vertex if the word is closed. We note that equivalent words generate the same graphs G_w (up to graph isomorphism) and the same passage-counts of the edges. Given an equivalence class $\mathbf{w}$, we shall sometimes denote $\#E_{\mathbf{w}}$ and $\#V_{\mathbf{w}}$ to be the common number of edges and vertices for graphs associated with all the words in this equivalence class $\mathbf{w}$.

Observe that any word w induces a walk on the graph G_w. This walk is directed.

Definition 31. (Closed Non Backtracking words) Any word w of length k is called non-backtracking if the walk induced by the word w on G_w doesn't back track. In particular, suppose $w = (i_1, \dots, i_k)$ it induces an walk $(i_1, i_2, \dots, i_k)$ on the graph G_w which is defined above. Here the word w will be called non backtracking if for any j, $i_j \neq i_{j+2}$. Any word w is called "closed non-backtracking" if it is closed and non backtracking at the same time. We shall denote this collection of words by $\mathcal{W}_{NB,k}$.

For any word $w = (i_1, i_2, \dots, i_k)$, we define

$$X_w := \prod_{j=1}^{k-1} A_{\text{cen1}}(i_j, i_{j+1}). \tag{5.6.1}$$

We now observe that the trace of the matrix B_u can be written in terms of the words. It is done in the following way:

$$\begin{aligned} \mathsf{Tr}\left(B_u^k\right) &= \sum_{(i_0,j_0),(i_1,j_1),\ldots,(i_{k-1},j_{k-1})} B_u\left((i_0,j_0),(i_1,j_1)\right)\ldots B_u\left((i_{k-1},j_{k-1}),(i_0,j_0)\right) \\ &= \sum_{w=(i_1,i_2,\ldots,i_{k-1},i_0,i_1)\ :\ w\in\mathcal{W}_{NB,k}} X_w. \end{aligned} \tag{5.6.2}$$

Here in the last expression of (5.6.2) we have used the fact that $B_u\left((i,j),(i',j')\right) \neq 0$ only if $j = i'$ and $j' \neq i$.

We now explore one of the important properties of the closed non backtracking words which is the key ingredient of the results in this paper.

Proposition 18. *Suppose $w \in \mathcal{W}_{NB,k}$, then G_w cannot be a tree.*

Proof. We prove by contradiction. Since w is closed, the walk induced by w returns to its starting point. Suppose G_w is a tree, then at some point of time a leaf is explored in the forward direction. However on any tree there is an unique path between any two vertices. As a consequence, in order the walk to be closed one must go back from a leaf to its parent. However on the other hand the walk reaches the leaf from its parent. By definition of a leaf in the tree these two steps must be subsequent to each other. Hence the result follows. □

As a trivial corollary of Proposition 18 we have if $w \in \mathcal{W}_{NB,k}$, then $\#E_w \geq \#V_w$. Proposition 18 is the fundamental reason why the matrix B_u works whereas the matrix A_{cen2} fails. Finally, in order to establish the rates in Theorem 14 we need the following lemma from Banerjee and Ma (2017a). In the statement of Lemma 24 we need to introduce the concept of sentences.

Definition 32 (Sentences and corresponding graphs)**.** A sentence $a = [w_i]_{i=1}^m = [[\alpha_{i,j}]_{j=1}^{l(w_i)}]_{i=1}^m$ is an ordered collection of m words of length $(l(w_1),\ldots,l(w_m))$ respectively. We define the graph $G_a = (V_a, E_a)$ to be the graph with

$$V_a = \text{supp}(a), \quad E_a = \left\{\{\alpha_{i,j},\alpha_{i,j+1}\} | i = 1,\ldots,m; j = 1,\ldots,l(w_i)-1\right\}.$$

Definition 33 (Weak CLT sentences)**.** A sentence $a = [w_i]_{i=1}^m$ is called a *weak CLT sentence*, if the following conditions are satisfied:

1. All the words w_i's are closed.
2. Jointly the words w_i visit each edge of G_a at least twice.
3. For each $i \in \{1, \ldots, m\}$, there is another $j \neq i \in \{1, \ldots, m\}$ such that G_{w_i} and G_{w_j} have at least one edge in common.

We are now ready to state the lemma.

Lemma 24. *Let $\mathcal{A} = \mathcal{A}^n_{m,t}(l_1, \ldots, l_m)$ be the set of weak CLT sentences $a = [w_i]_{i=1}^m$ such that the letter set is $\{1, \ldots, n\}$, $\#V_a = t$ and $l(w_i) = l_i$ for $i = 1, \ldots, m$. If $l_i \geq 3$ for $i = 1, \ldots, m$, then*

$$\#\mathcal{A} \leq n^t 2^l \left(C_1 l\right)^{C_2 m} l^{3(l-2t)} \tag{5.6.3}$$

where $l = \sum_{i=1}^m l_i$ and $C_1, C_2 > 0$ are numeric constants.

The proof of this Lemma can be found in Banerjee and Ma (2017a). However the main idea of the proof of this result comes from Lemma 2.1.23 in Anderson et al. (2010).

Lemma 25 (Lemma 2.1.23 in Anderson et al. (2010))**.** *Let $\mathcal{W}_{k,t}$ collect the equivalence classes corresponding to all weak Wigner words w of length $k+1$ with $wt(w) = t$. Then for $k \geq \min(2, 2t-2)$,*

$$\#\mathcal{W}_{k,t} \leq 2^k k^{3(k-2t+2)}.$$

Since closed words contains closed non backtracking words, we shall use Lemma 24 and Lemma 25 for closed non backtracking words with out further mentioning.

5.6.2. Proofs of the main results

With these backgrounds in hand we are now ready to prove Theorem 14.

Proof of Theorem 14:

Before going into the proof we at first state a simple result which will be used repeatedly in the proof.

Lemma 26. *Suppose w is any closed word with $l(w) = k+1$ and recall the definition of X_w. Then* $\mathrm{E}\left[|X_w|\right] \le \left(\sqrt{\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}\right)^k 2^k p_{n,\mathrm{av}}^{\#E_w}$.

Proof. Suppose $w = (i_0, i_1, \ldots, i_{k-1}, i_0)$, then

$$\begin{aligned} X_w &= \left(\frac{1}{\sqrt{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}\right)^k \prod_{j=0}^{k-1} \left(x_{i_j,i_{j+1}} - p_{n,\mathrm{av}}\right) \\ &= \left(\frac{1}{\sqrt{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}}\right)^k \prod_{e\in E_w} \left(x_e - p_{n,\mathrm{av}}\right)^{N_e} \end{aligned} \tag{5.6.4}$$

where for any edge $e \in E_w$, N_e denotes the number of times the edge e is traversed by the walk induced by w. Since $|x_e - p_{n,\mathrm{av}}| \le 1$, we have $\mathrm{E}\left|x_e - p_{n,\mathrm{av}}\right|^{N_e} \le \mathrm{E}\left|x_e - p_{n,\mathrm{av}}\right| \le p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}}) + (1-p_{n,\mathrm{av}})(p_{n,\mathrm{av}}) \le 2p_{n,\mathrm{av}}$. □

We shall only prove part (ii) of the Theorem. Proof of part (i) is simpler and the main arguments are covered in the proof of part (ii).

Proof of part (ii): We at first use (5.6.2) to observe

$$\mathsf{Tr}\left(B_u^{2k}\right) = \sum_{w\ :\ w\in\mathcal{W}_{NB,2k}} X_w.$$

We now divide the class $\mathcal{W}_{NB,2k}$ into several further sub classes as follows: $\mathcal{W}_{NB,2k} = \mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3$ where

$$\begin{aligned} \mathcal{W}_1 &= \{w\ :\ \#V_w = \#E_w \text{and each edge in the cycle of } G_w \text{ is traversed exactly once}\}. \\ \mathcal{W}_2 &= \{w\ :\ \#V_w = \#E_w \text{and each edge in the cycle of } G_w \text{ is traversed twice or more}\}. \\ \mathcal{W}_3 &= \{w\ :\ \#E_w > \#V_w\}. \end{aligned} \tag{5.6.5}$$

We shall analyze each class separately.

Analysis of the term containing $\mathcal{W}_3$: Our first goal is to prove

$$\sum_{w\in\mathcal{W}_3} X_w \xrightarrow{p} 0.$$

This is done by proving

$$\mathrm{E}\left(\sum_{w\in\mathcal{W}_3} X_w\right)^2 \to 0$$

with help of Lemma 24. First observe that

$$\begin{aligned}\mathrm{E}\left(\sum_{w\in\mathcal{W}_3} X_w\right)^2 &= \sum_{w,x\in\mathcal{W}_3} \mathrm{E}\left[X_w X_x\right] \\ &= \sum_{a=(w,x)\in\mathcal{W}_3^2} \mathrm{E}\left[X_w X_x\right].\end{aligned} \tag{5.6.6}$$

Now observe that if there exist at least one edge in G_a which is traversed exactly once, then $\mathrm{E}\left[X_w X_x\right] = 0$. Now there can be two cases

Case 1: The words w and x share an edge. In this case, a is a weak CLT sentence. So we can apply Lemma 24 with $m = 2$ and $l_1 = l_2 = 2k+1$ to get that the sum in this case is bounded by

$$\begin{aligned}&\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k} \sum_{\zeta=1}^{2k-1} n^{\zeta} p_{n,\mathrm{av}}^{\zeta} 2^{4k+2} (C_1(4k+2))^{2C_2} (4k+2)^{6(2k-\zeta+1)} \\ &\leq \left(\frac{1}{(1-p_{n,\mathrm{av}})}\right)^{2k} 2^{4k+2} (C_1(4k+2))^{2C_2} (4k+2)^6 \sum_{\zeta=1}^{2k-1} \left(\frac{(4k+2)^6}{np_{n,\mathrm{av}}}\right)^{(2k-\zeta)}\end{aligned} \tag{5.6.7}$$

where C_1 and C_2 are known constants.

Observe that $2k - \zeta > 1$. As a consequence, the R.S. of (5.6.7) is a geometric sum on $\left((\frac{4k+2}{np_{n,\mathrm{av}}})^6\right)^i$ with lowest index being 1. We also have $(\frac{(4k+2)^6}{np_{n,\mathrm{av}}}) \to 0$ by the assumption $k = o(\log(np_{n,\mathrm{av}}))$. As a consequence, the R.S. of (5.6.7) can be bounded by

$$\left(\frac{1}{(1-p_{n,\mathrm{av}})}\right)^{2k+1} 2^{4k+2} (C_1(4k+2))^{2C_2} (4k+2)^6 C_3 \left(\frac{(4k+2)^6}{np_{n,\mathrm{av}}}\right). \tag{5.6.8}$$

Here C_3 is another known constant. It is easy to see (5.6.8) goes to zero when $k = o(\log(np_{n,\text{av}}))$.

Case 2: The words w_1 and w_2 don't share an edge. Let $wt(w_1) = \zeta_1$ and $wt(w_2) = \zeta_2$. We shall apply Lemma 25 in this case. Since both ζ_1 and ζ_2 are less than or equal to $k-1$. The equation $2k > 2\zeta_j - 2$ is trivially satisfied for $j \in \{1, 2\}$. Now from Lemma 25 a crude upper bound to the number of sentences $a = [w_1, w_2]$ such that w_1 and w_2 don't share an edge such that $wt(a) = \zeta$ is given by

$$\begin{aligned}&\sum_{\zeta_1}\sum_{\zeta_2=\zeta-\zeta_1} n^{\zeta}2^{2k}(2k)^{3(2k-2\zeta_1+2)} \times 2^{2k}(2k)^{3(2k-2\zeta_2+2)}\\ &\quad=\sum_{\zeta_1}\sum_{\zeta_2=\zeta-\zeta_1} n^{\zeta}2^{4k}(2k)^{3(4k-2\zeta+4)} \le n^{\zeta}\zeta^2 2^{4k}(2k)^{6(2k-\zeta+2)}.\end{aligned} \tag{5.6.9}$$

Here the factor ζ^2 comes due to the sum. Consequently, the sum in this case is bounded by

$$\begin{aligned}&\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k}\sum_{\zeta=1}^{2k-2} n^{\zeta}p_{n,\text{av}}^{\zeta}\zeta^2 2^{4k}(2k)^{6(2k-\zeta+2)}\\ &\le\left(\frac{1}{1-p_{n,\text{av}}}\right)^{2k}\sum_{\zeta=1}^{2k-2} 2^{4k}(4k)^2(2k)^{12}\left(\frac{(2k)^6}{np_{n,\text{av}}}\right)^{(2k-\zeta)}.\end{aligned} \tag{5.6.10}$$

Now (5.6.10) can be analyzed similarly as (5.6.7) to get that (5.6.10) goes to 0 also. As a consequence, one can simply ignore the term containing $\mathcal{W}_3$.

Analysis of the term containing $\mathcal{W}_2$: At first observe that for any $w \in \mathcal{W}_2$ the graph G_w contains exactly one cycle and removing the cycle the rest of the graph is a collection of forests. At first we prove that if the graph G_w has a non empty forest then it is a single straight line without having any branches. Further the walk induced by any such word has to start from one end of the line which is not connected to the cycle.

To begin with we at first prove if G_w has a non empty forest then the corresponding walk cannot start from a point on the cycle. This is true because if the walk would start from a point on the cycle and at some time point it goes into the forest then after entering the forest it has to come back to the cycle since the walk is closed. Now we look at the instants when the walk explores any component of the forest and comes back to the cycle. This is a closed walk on the component of the forest which is a tree. In Proposition 18 we have proved any such walk has to backtrack at some point which cannot happen since the walk induced by the word is non-backtracking.

Now we prove the other part. It is clear if a non-backtracking walk has a non empty forest then it has to start at some point on the forest. Observe that there is a unique path from the starting point of the walk to the cycle. If there are branches hanging from this trail at some point the walk has to enter the branch and return to the point from where the branch started. Again applying Proposition 18 we get to the required result. Also note that there cannot be multiple trees hanging from the cycles. In this case one needs to enter at least one tree from the cycle which is again not possible by Proposition 18.

Finally observe that the only way an word $w \in \mathcal{W}_2$ and have a trail in the following way. The word w start from the point on the trail which is not connected to the cycle and goes straight into the cycle then moves on the cycle several times in a fixed direction (clockwise or counter clockwise) and follows the trail back to the starting point. Any other exploration on the cycle will lead to backtracking of the edges. We now prove among the words in $\mathcal{W}_2$, the only words which matters are those where the cycle is traversed exactly twice.

Now we introduce some other notations. For any word $w \in \mathcal{W}_2$, the cycle in G_w will be called a bracelet and denoted by Z. Observe that the word w induces an ordering in which the vertices in the bracelet are traversed. Also note that since the word w is non-backtracking this order is preserved in every exploration on the cycle. We shall call the ordered tuple in which the cycle is explored $(\alpha_1, \ldots, \alpha_r)$. Let m be the length of the trail and t be the number of times the bracelet is explored. Then any word $w \in \mathcal{W}_2$ looks like

$$(\beta_1, \ldots, \beta_m, \underbrace{\alpha_1, \ldots, \alpha_r}_{t \text{ times}}, \alpha_1, \beta_m, \ldots, \beta_1). \tag{5.6.11}$$

where all the $\beta_1, \ldots, \beta_m$ and $\alpha_1, \ldots, \alpha_r$ are distinct. We now further divide the class $\mathcal{W}_2$ into two further sub classes. $\mathcal{W}_{2,1}$ and $\mathcal{W}_{2,2}$. Here $\mathcal{W}_{2,1}$ contains all words with $t \geq 3$ and $\mathcal{W}_{2,2}$ contains all words with $t = 2$. At first we prove

$$\operatorname{Var}\left[\sum_{w \in \mathcal{W}_2} X_w\right] \to 0 \tag{5.6.12}$$

whenever $k = o\left(\log(np_{n,\mathrm{av}})\right)$. To see this we expand the left hand side of (5.6.12):

$$\begin{aligned} &\operatorname{Var}\left[\sum_{w \in \mathcal{W}_2} X_w\right] \\ &= \sum_{w,x \in \mathcal{W}_2} \mathrm{E}\left[\left(X_w - \mathrm{E}[X_w]\right)\left(X_x - \mathrm{E}[X_x]\right)\right]. \end{aligned} \tag{5.6.13}$$

First observe that $\mathrm{E}\left[\left(X_w - \mathrm{E}[X_w]\right)\left(X_x - \mathrm{E}[X_x]\right)\right] = 0$ unless the sentence $a = [w, x]$ is a weak clt sentence. This simply follows from the following two reasons. Firstly if the graphs G_w and G_x don't share an edge, then the random variables $(X_w - \mathrm{E}[X_w])$ and $(X_x - \mathrm{E}\left[X_x\right])$ are mutually independent which implies $\mathrm{E}\left[\left(X_w - \mathrm{E}[X_w]\right)\left(X_x - \mathrm{E}[X_x]\right)\right] = 0$. Finally if there exist at least one edge in G_a (say e) which is traversed exactly once, then this edge must belong to one of G_w and G_x. Without loss of generality we assume this edge belongs to G_w. This implies $\mathrm{E}\left[X_w\right] = 0$ and the product X_w can be written as follows:

$$X_w = \frac{1}{\sqrt{np_{n,\mathrm{av}}(1 - p_{n,\mathrm{av}})}}(x_e - p_{n,\mathrm{av}})Y_w$$

where the random variable $(x_e - p_{n,\mathrm{av}})$ is independent of Y_w and X_x. As a consequence, we have $\mathrm{E}\left[X_w\right] = 0$.

Now

$$\begin{aligned}
&\mathrm{E}\left[\left(X_w - \mathrm{E}[X_w]\right)\left(X_x - \mathrm{E}[X_x]\right)\right] \\
&= \mathrm{E}\left[X_w X_x\right] - \mathrm{E}[X_w]\mathrm{E}[X_x] \leq \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k}\left(p_{n,\text{av}}^{\#E_a} + p_{n,\text{av}}^{\#E_w+\#E_x}\right) \qquad (5.6.14)\\
&\leq \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k} 2p_{n,\text{av}}^{\#E_a}.
\end{aligned}$$

By definition of $\mathcal{W}_2$, any word $w \in \mathcal{W}_2$ will have $\#E_w \leq k$ and $\#V_w = \#E_w \leq k$. Since $a = [w, x]$ is a weak clt sentence, we have $\#V_a \leq \#E_a < \#E_w + \#E_x < 2k$. Now by applying Lemma 24 we have the right hand side of (5.6.13) is bounded by

$$\left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{2k} \sum_{t}^{2k-1} 2^{4k+2} n^t p_{n,\text{av}}^t \left(C_1(4k+2)\right)^{2C_2} (4k+2)^{3(4k+2-2t)} \to 0 \qquad (5.6.15)$$

whenever $k = o\left(\log(np_{n,\text{av}})\right)$. Observe now that for any $w \in \mathcal{W}_{2,1}$, we have $wt(w) = \zeta < k$. As a consequence,

$$\begin{aligned}
\sum_{w\in\mathcal{W}_{2,1}} \mathrm{E}\left[|X_w|\right] &\leq \left(\frac{1}{np_{n,\text{av}}(1-p_{n,\text{av}})}\right)^{k} \sum_{\zeta=1}^{k-1} p_{n,\text{av}}^{\zeta} n^{\zeta} 2^{2k} (2k)^{6(k-\zeta+1)} \\
&\leq 2^{2k}(2k)^6 \left(\frac{1}{(1-p_{n,\text{av}})}\right)^{k} \sum_{\zeta=1}^{k-1} \left(\frac{(2k)^6}{np_{n,\text{av}}}\right)^{k-\zeta} \to 0.
\end{aligned} \qquad (5.6.16)$$

Finally we are left with the case when $w \in \mathcal{W}_{2,2}$. For any $w \in \mathcal{W}_{2,2}$, we have $wt(w) = \#E_w = k$ and each edge in G_w is traversed exactly twice. As a consequence X_w is positive and $\mathrm{E}\left[X_w\right] = \left(p_{n,\text{av}}(1-p_{n,\text{av}})\right)^k$. Further once the length of the bracelet is fixed, there is exactly one equivalent class for such w as described in (5.6.11) with $t = 2$. Since the length of the bracelet can vary from 3 to k and there are $n^k\left(1 + O(k^2/n)\right)$ many words corresponding to each equivalent class, there are $(k-2)n^k\left(1 + O(k^2/n)\right)$ words of this type. Hence

$$\sum_{w\in\mathcal{W}_{2,1}} \mathrm{E}\left[|X_w|\right] = (k-2)\left(1 + O(k^2/n)\right).$$

<u>Analysis of terms containing $\mathcal{W}_1$:</u> Our final task boils down to the analysis of the words in

$\mathcal{W}_1$. Firstly observe that for any word $w \in \mathcal{W}_1$, $\mathrm{E}[X_w] = 0$. So any word in $\mathcal{W}_1$ does not change the mean of $\mathsf{Tr}\left[B_u^{2k}\right]$. However the randomness in $\mathsf{Tr}\left[B_u^{2k}\right]$ comes from the words in $\mathcal{W}_1$.

Now observe that any word in $\mathcal{W}_1$ has the following form

$$(\beta_1, \ldots, \beta_m, \alpha_1, \ldots, \alpha_r, \alpha_1, \beta_m, \ldots, \beta_1).$$

The argument for this structure is exactly same as the argument given in the analysis of $\mathcal{W}_2$. For any word w, we call $(\beta_1, \ldots, \beta_m)$ to be the stem of the word w and the cycle induced by $(\alpha_1, \ldots, \alpha_r, \alpha_1)$ to be the bracelet of w. We at first observe that r is always even and has to be greater than equal to 4. We further divide the class $\mathcal{W}_2$ according to their bracelet lengths and denote the corresponding class by $\mathcal{W}_{1,r}$. First of all if $w_1 \in \mathcal{W}_{1,r_1}$ and $w_2 \in \mathcal{W}_{1,r_2}$ then $\mathrm{E}\left[X_{w_1} X_{w_2}\right] = 0$. So

$$\mathrm{Var}\left[\sum_{w \in \mathcal{W}_2} X_w\right] = \sum_r \mathrm{Var}\left[\sum_{w \in \mathcal{W}_{1,r}} X_w\right].$$

Now we analyze

$$\mathrm{Var}\left[\sum_{w \in \mathcal{W}_{1,r}} X_w\right]$$

for each given r. We have

$$\begin{aligned}\mathrm{Var}\left[\sum_{w \in \mathcal{W}_{2,r}} X_w\right] &= \sum_{w,x \in \mathcal{W}_2} \mathrm{E}\left[X_w X_x\right] \\ &= \sum_{w,x\ :\ \text{they have common bracelet}} \mathrm{E}\left[X_w X_x\right]\end{aligned} \tag{5.6.17}$$

Given w and x any two words with common bracelet, we now look at the sentence $a = [w, x]$. First of all we claim that $\#E_a \geq \#V_a$ and $E_a \leq 2k$. The fact $\#V_a \leq \#E_a$ is trivial hence omitted. For the second part observe that $2m + r = 2k$ where m is the length of the stem corresponding to the word w or x. Since the bracelet is common between w and x and there might be possible common edges between the stems of w and x we have $\#E_a \le r + 2m$. Here the equality occurs only if the stems of w and x are disjoint. In any case the sentence a is a weak clt sentence. At first we consider the case when $\#V_a = \zeta < 2k$. By Lemma 24 with $m = 2$, we have the cardinality of this class of sentences are bounded by $2^{4k+2}\left(C_1(4k+2)\right)^{2C_2}(4k+2)^{3(4k+2-2t)}$. Further $\mathrm{E}\left[|X_w X_x|\right]$ in this case is bounded by $p_{n,\mathrm{av}}^{\#E_a} \le p_{n,\mathrm{av}}^{\zeta}$. As a consequence, the sum in (5.6.17) corresponding to this case is bounded by:

$$\begin{aligned} &\sum_{\zeta=1}^{2k-1}\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k} n^{\zeta}p_{n,\mathrm{av}}^{\zeta}2^{4k+2}\left(C_1(4k+2)\right)^{2C_2}(4k+2)^{3(4k+2-2\zeta)} \\ &\le \left(\frac{1}{(1-p_{n,\mathrm{av}})}\right)^{2k}\sum_{\zeta=1}^{2k-1}2^{4k+2}\left(C_1(4k+2)\right)^{2C_2}(4k+2)^{6}\left(\frac{(4k+2)^6}{np_{n,\mathrm{av}}}\right)^{2k-\zeta} \to 0. \end{aligned} \tag{5.6.18}$$

****(There should be a sum of over $p_{n,\mathrm{av}}^{\zeta}$ which I am overlooking for now. This doesn't change any of the results.) Finally we are left with the case when $\#V_a = \#E_a = 2k$. In this case the stems of w and x are mutually disjoint while their bracelets are the same. We shall do a precise enumeration in this case. At first we fix a word w and calculate the number of words x such that the aforesaid conditions hold. Suppose $w = (\beta_1,\ldots,\beta_m,\alpha_1,\ldots,\alpha_r,\alpha_1,\beta_m,\ldots,\beta_1)$ and $x = (\beta'_1,\ldots,\beta'_m,\alpha'_1,\ldots,\alpha'_r,\alpha'_1,\beta'_m,\ldots,\beta'_1)$. Firstly observe that either $(\alpha'_1,\ldots,\alpha'_r,\alpha'_1) = (\alpha_{\tau(1)},\alpha_{\tau(2)},\ldots,\alpha_{\tau(r)},\alpha_{\tau(1)})$ or $(\alpha'_1,\ldots,\alpha'_r,\alpha'_1) = (\alpha_{\tau(1)},\alpha_{\tau(r)},\alpha_{\tau(r-1)},\ldots,\alpha_{\tau(2)},\alpha_{\tau(1)})$. Here τ is a cyclic permutation of $\{1,\ldots,r\}$. So fixing w, there are exactly $2r$ choices for the sub-word which gives rise to the same bracelet as w. Further the stem $(\beta'_1,\ldots,\beta'_m)$ is disjoint from $(\beta_1,\ldots,\beta_m)$. So fixing w there are $2rn^m\left(1+O(k^2/n)\right)$ many choices for x and there are $n^{m+r}\left(1+O(k^2/n)\right)$ choices for w. As a consequence, there are total $2rn^{2m+r}\left(1+O(k^2/n)\right) = 2rn^{2k}\left(1+O(k^2/n)\right)$ many choices for a. On the other hand for any choice of w, x of this kind we have $\mathrm{E}\left[X_w X_x\right] = \left(p_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^{2k}$. So the sum in (5.6.17) corresponding to this case is given by:

$$2r\left(\frac{1}{np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})}\right)^{2k}\left(np_{n,\mathrm{av}}(1-p_{n,\mathrm{av}})\right)^{2k}\left(1+O(k^2/n)\right)=2r\left(1+O(k^2/n)\right).$$

Hence we have

$$\mathrm{Var}\left[\sum_{w\in\mathcal{W}_{2,r}}X_w\right]=2r\left(1+O(k^2/n)\right)+o(1).$$

A similar argument shows that

$$\mathrm{Cov}\left[\sum_{w\in\mathcal{W}_{2,r}}X_w,C_{n,r}\right]=2r\left(1+O(k^2/n)\right)$$

and we know that

$$\mathrm{Var}\left(C_{n,r}\right)=2r\left(1+O(k^2/n)\right).$$

As a consequence,

$$\mathrm{Var}\left[\sum_{w\in\mathcal{W}_{2,r}}X_w-C_{n,r}\right]=O(k^3/n).$$

Since for different values of r, $C_{n,r}$'s and $\sum_{w\in\mathcal{W}_{2,r}}X_w$'s are uncorrelated, we have

$$\mathrm{Var}\left[\sum_{4\le r\le 2k\,:\,r\text{ even}}\ \sum_{w\in\mathcal{W}_{1,r}}X_w-\sum_{4\le r\le 2k\,:\,r\text{ even}}C_{n,r}\right]=O(k^4/n)\to 0.$$

This proves that

$$\mathsf{Tr}\left[B_u^{2k}\right]-\sum_{4\le r\le 2k\,:\,r\text{ even}}C_{n,r}-(k-2)\xrightarrow{p}0.$$

Proof of part (i): Proof of part (i) is covered in the proof of part (ii). Here everything is the same only one can further show that the class analogous to $\mathcal{W}_{2,2}$ is actually empty. □

Proofs of Corollary 2 and Theorem 15:

Proof of Corollary 2: Proof of Corollary 2 is a direct application of Theorems 12 and 14.

Using Theorem 12, we know when $t < 1$ the sequence of measures $\mathbb{P}_{1,n}$ and $\mathbb{P}_{0,n}$ are mutually contiguous. As a consequence, for any sequence of sets A_n, $\mathbb{P}_{0,n}(A_n) \to 0 \Leftrightarrow \mathbb{P}_{1,n}(A_n) \to 0$. Now from Theorem 14 we know for any $k = o\left(\log(np_{n,\mathrm{av}})\right)$,

$$\mathbb{P}_{0,n}\left[\left|C_{n,2k+1} - \left[\mathsf{Tr}\left(B_u^{2k+1}\right) - \mathsf{Tr}\left(B_u^{2k-1}\right)\right]\right| > \varepsilon\right] \to 0 \tag{5.6.19}$$

and

$$\mathbb{P}_{0,n}\left[\left|C_{n,2k} - \left[\mathsf{Tr}\left(B_u^{2k}\right) - \mathsf{Tr}\left(B_u^{2k-2}\right) - 1\right]\right| > \varepsilon\right] \to 0. \tag{5.6.20}$$

So (5.6.19) and (5.6.20) hold for the sequence of measures $\mathbb{P}_{1,n}$ also.

Proof of Theorem 15: Consider

$$\tilde{L}_n := \sum_{k=3}^{m_n} \frac{2t^k C_{n,k} - t^{2k}}{4k}.$$

Using Theorem 12 and the fact that $\mathrm{E}_{\mathbb{P}_{0,n}}\left[C_{n,k}\right] = 0$, $\mathrm{Cov}_{\mathbb{P}_{0,n}}\left[C_{n,k_1}, C_{n,k_2}\right] = 2k_1 \mathbb{I}_{k_1=k_2}\left(1 + \frac{k_1^2}{n}\right)$, $\mathrm{E}_{\mathbb{P}_{1,n}}\left[C_{n,k}\right] = t^k\left(1 + \frac{k^2}{n}\right)$ and $\mathrm{Cov}_{\mathbb{P}_{1,n}}\left[C_{n,k_1}, C_{n,k_2}\right] = 2k_1 \mathbb{I}_{k_1=k_2}\left(1 + \frac{k_1^2}{n}\right) + O\left(\frac{k^k}{n}\right)$(See the proof of Proposition 4.1 in Banerjee (2018) for details) one gets that $\tilde{L}_n$ has the appropriate limiting distribution under $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$. ******(Another way is to use Lecam's Third Lemma and the decomposition result in our paper. In this case one can take $m_n = o\log(np_{n,\mathrm{av}})$)******** However the variance estimates obtained in the proof of Theorem 14 can be used to prove under $\mathbb{P}_{0,n}$, $L_n - \tilde{L}_n \xrightarrow{p} 0$. As a consequence, by applying contiguity we have under $\mathbb{P}_{1,n}$ also $L_n - \tilde{L}_n \xrightarrow{p} 0$. Hence L_n has the appropriate distribution under $\mathbb{P}_{1,n}$. This concludes the proof. □

CHAPTER 6 : A Bernstein type inequality for sums of choices in three dimensional arrays

6.1. Overview

We consider the three dimensional array $\mathcal{A} = \{a_{i,j,k}\}_{1\leq i,j,k\leq n}$, with $a_{i,j,k} \in [0,1]$, and the statistics $T_2 := \sum_{i=1}^{n} a_{i,\sigma(i),\pi(i)}$, where σ and π are chosen independently from the set of permutations of $\{1,2,\ldots,n\}$. These can be viewed as natural three dimensional generalizations of the statistic $T_1 = \sum_{i=1}^{n} a_{i,\sigma(i)}$, considered by Hoeffding Hoeffding et al. (1951). Here we give Bernstein type concentration inequalities T_2 by extending the argument for concentration of T_1 by Chatterjee Chatterjee (2005).

6.2. Arrays and Concentration Inequalities

Let $\mathcal{A} = \{a_{i,j,k}\}_{1\leq i,j,k\leq n}$ be a three dimensional array with $a_{i,j,k} \in [0,1]$, and consider the following two statistics

$$T_2 := \sum_{i=1}^{n} a_{i,\sigma(i),\pi(i)} \tag{6.2.1}$$

where σ and π are chosen independently and uniformly from the set S_n of permutations of $[n] = \{1,2,\ldots,n\}$. Our goal is to obtain Bernstein type tail bounds for the statistics T_2. Statistics of these type have already been considered in literature; for example, when the dimension is two, the statistic

$$T_1 := \sum_{i=1}^{n} a_{i,\sigma(i)}$$

where σ is drawn uniformly from S_n was studied by Hoeffding Hoeffding et al. (1951), who proved that, under certain conditions, it has an asymptotic normal distribution as n goes to infinity. In fact, the special case when $a_{i,j} = c_i \cdot d_j$ dates back to the works of Wald and Wolfwitz Wald and Wolfowitz (1944) and Noether Noether (1949). Another example of the statistic T_1 is the Spearman's footrule, useful in non-parametric statistics, where $a_{i,j} = |i-j|$. Statistic T_2 can be viewed as natural generalizations of statistic T_1 in three dimensions. However, in this paper we are concerned about concentration inequalities for

T_2, and not on their asymptotic distribution. The concentration of T_1 was considered by Chatterjee Chatterjee (2005) (page 52); specifically he obtained an elegant tail bound of Bernstein type.

Theorem 16. *Let* $\{a_{i,j}\}_{1\leq i,j,\leq n} \in [0,1]$ *and* T_1 *be as above. Then for any* $t \geq 0$,

$$\mathbb{P}\left(|T_1 - E(T_1)| \geq t\right) \leq 2\exp\left\{-\frac{t^2}{4\mathrm{E}[T_1] + 2t}\right\}$$

Chatterjee obtains this bound by the method of exchangeable pairs, and here we extend his method to obtain Bernstein type concentration inequalities for T_2.

Theorem 17. *If* T_2 *is as defined in (6.2.1) and* $\{a_{i,j,k}\}_{1\leq i,j,k\leq n} \in [0,1]$, *then*

$$\mathbb{P}\left(|T_2 - \mathrm{E}[T_2]| \geq t\right) \leq 2\exp\left\{-\frac{(t-3+O(1/n))^2}{12\mathrm{E}(T_2) + 18 + 6(1+O(1/n))(t-3)}\right\}. \qquad (6.2.2)$$

Concentrations of functions of random permutations have also been studied by Talagrand (Theorem 5.1) Talagrand (1995), Murray Maurey (1979) and McDiarmid McDiarmid (2002). However, as mentioned in Chatterjee Chatterjee (2005), apart from Talagrand's Theorem 5.1 none of these results are able to give Bernstein type concentration inequalities as above.

6.3. On the method of exchangeable pair

We first need to recall some notions on the theory of exchangeable pairs as used by Chatterjee Chatterjee (2005).

Definition 34. Suppose X is a random variable on the measure space $(\Omega, \mathcal{F}, \mathbb{P})$ and X' is another random variable defined on the same measure space. The pair (X, X') is called an exchangeable pair if $(X, X') \stackrel{d}{=} (X', X)$.

The method of exchangeable pairs exploits three useful functions:

- A function $F : \mathbb{R}^2 \to \mathbb{R}$, measurable and almost surely anti-symmetric, i.e. such that $F(X, X') = -F(X', X)$ almost surely.

- The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(X) := \mathrm{E}\left[F(X, X')\,|X\right]$. This is a fundamental quantity in the the concentration inequality.

- The function $v(X)$, that serves as a stochastic bound size of $f(X)$, and which is defined by

$$v(X) := \frac{1}{2}\mathrm{E}\left[|(f(X) - f(X')) \cdot F(X, X')| \;\middle|\; X\right]. \tag{6.3.1}$$

The following lemma from Chatterjee Chatterjee (2005) tells us how the concentration of $f(X)$ is governed by a bound on $v(X)$.

Lemma 27. *(Theorem 3.9 in Chatterjee (2005)) Suppose (X, X') is an exchangeable pair and $F(X, X')$, $f(X)$ and $v(X)$ are defined as before, with $v(X) \leq C + Bf(X)$ almost surely for some known fixed constants B and C. Then*

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

The fundamental idea of the method of exchangeable pairs is to construct $F(X, X')$, $f(X)$ and $v(X)$ so that Lemma 27 yields concentration for $f(X)$. One example to keep in mind of $F(X, X')$ is $c(X - X')$, where c is a nonrandom constant.

Remark 12. Chatterjee Chatterjee (2007) further proved that for the lower tail one can get an actual Gaussian decay. In particular, one can show that

$$\mathbb{P}\left[f(X) < -t\right] \leq \exp\left\{-\frac{t^2}{2C}\right\}.$$

6.4. Strategy of the Proofs

For proving the concentration inequalities for T_2, we use the following general strategy. At first we construct the statistics T_2' by applying "small" changes to T_2, such that two properties hold. We require (T_2, T_2') to form an exchangeable pair and $\mathrm{E}\left[T_2 - T_2'\,|T_2\right]$ to be somewhat close to $c\left(T_2 - \mathrm{E}[T_2]\right)/n$. We then define the quantity $v(T_2)$ as in the previous section and bound it in terms of $T_2 - \mathrm{E}[T_2]$. Finally, we derive the concentration inequality for $|T_2 - \mathrm{E}[T_2]|$ by applying Lemma 27.

The construction of T_2' is not as simple as the procedure used in Chatterjee (2005). The main reason is that $\mathrm{E}[T_2]$ is a sum over three independent directions i, j, k, while, fixing σ and π, T_2 is a sum over only one single direction. As a consequence, one might check that it is not possible to get $\mathrm{E}\left[T_2 - T_2' \,|T_2\right]$ close to $c\left(T_2 - \mathrm{E}[T_2]\right)/n$ by simply moving two indexes. Instead, one needs to move three indexes in a systematic way. We then choose (I_1, I_2, I_3) extracted uniformly *without replacement* from $[n]$ and define the functions $\tau_{1,2} : [n]^3 \to \mathcal{S}_n$ such that $\tau_1(I_1, I_2, I_3) = (I_1, I_2, I_3)$ and $\tau_2(I_1, I_2, I_3) = (I_1, I_3, I_2)$. These are the only cyclic permutations which are not the identity. The permutations σ', π' are defined as follows:

$$(\sigma', \pi') = \begin{cases} \sigma \circ \tau_1(I_1, I_2, I_3), \pi \circ \tau_2(I_1, I_2, I_3) & \text{with probability } \frac{1}{2} \\ \sigma \circ \tau_2(I_1, I_2, I_3), \pi \circ \tau_1(I_1, I_2, I_3) & \text{with probability } \frac{1}{2}. \end{cases}$$

Note that this σ' is different from the one defined in the construction of T_1', but it will always be clear which one of the two we are considering. Finally we define

$$T_2' := \sum_i a_{i,\sigma'(i),\pi'(i)}.$$

For σ' and π' to be valid permutations one needs all the indexes (I_1, I_2, I_3) to be distinct or for all three to be the same. We only consider the case when (I_1, I_2, I_3) are all distinct for convenience, since the case when they are all the same does not affect the exchangeability of T_2 and T_2' and it just gives a slight change in the result which is negligible as n grows to infinity. It is important to note that one needs σ' and π' to be valid permutations in order for T_2' to have the same distribution as T_2, necessary condition to have exchangeability.

6.5. Proofs of the results

Now to prove (6.2.2), we need a more delicate argument. For $k \in \{1, 2, 3\}$ we define C_k to be the set of all ordered tuples $(I_1, I_2, I_3) \in [n]^3$ such that $\#\{I_1, I_2, I_3\} = k$. Observe that in order σ' and π' to be valid permutations, one needs $(I_1, I_2, I_3) \in C_k$ where $k = 1$ or 3. When $k = 1$, one has $\sigma' = \sigma$ and $\pi' = \pi$. In this paper we only consider the case when I_1, I_2, I_3 are all distinct. It is easy to see that in that case we have $\tau_1(I_1, I_2, I_3)^{-1} = \tau_2(I_1, I_2, I_3)$. On the contrary, when $(I_1, I_2, I_3) \in C_2$, the permutations σ' and π' are not well-defined. With the choice of $(I_1, I_2, I_3) \in C_3$, and (σ', π') defined as before, we have

$$\begin{aligned} T_2' &:= \sum_i a_{i,\sigma'(i),\pi'(i)} \\ &= T_2 - \sum_{j=1}^{3} a_{I_j,\sigma(I_j),\pi(I_j)} + \begin{cases} \sum_{j=1}^{3} a_{I_j,\sigma\circ\tau_1(I_1,I_2,I_3)(I_j),\pi\circ\tau_2(I_1,I_2,I_3)(I_j)} & \text{with prob. } \frac{1}{2} \\ \sum_{j=1}^{3} a_{I_j,\sigma\circ\tau_2(I_1,I_2,I_3)(I_j),\pi\circ\tau_1(I_1,I_2,I_3)(I_j)} & \text{with prob. } \frac{1}{2}. \end{cases} \end{aligned}$$

Proposition 19. (T_2, T_2') *forms an exchangeable pair.*

Proof. At first observe that

$$\mathbb{P}\left[T_2 = x, T_2' = x'\right] = \mathrm{E}\left[\mathbb{P}\left[T_2 = x, T_2' = x' \,\middle|\, \sigma, \pi, \sigma', \pi'\right]\right]$$

Since (T_2, T_2') is a function of $(\sigma, \pi, \sigma', \pi')$, we have

$$\mathbb{P}\left[T_2 = x, T_2' = x' \,\middle|\, \sigma, \pi, \sigma', \pi'\right] = \mathbb{I}\left(T_2(\sigma, \pi) = x, T_2(\sigma', \pi') = x'\right).$$

We at first fix a value of (I_1, I_2, I_3) We set $\gamma_3 = \gamma_1 \circ \tau_1(I_1, I_2, I_3),\ \gamma_4 = \gamma_2 \circ \tau_2(I_1, I_2, I_3),\ \gamma_5 =$

$\gamma_1 \circ \tau_2(I_1, I_2, I_3)$ and $\gamma_6 = \gamma_2 \circ \tau_1(I_1, I_2, I_3)$, to get

$$
\begin{aligned}
&\mathbb{P}\left[T_2 = x, T_2' = x'\right] \\
&= \mathrm{E}\left[\mathbb{P}\left[T_2 = x, T_2' = x' \,|\sigma, \pi, \sigma', \pi'\right]\right] \\
&= \frac{1}{2n!^2}\frac{1}{n(n-1)(n-2)} \\
&\sum_{i_1,i_2,i_3 \in C_3}\sum_{\gamma_1,\gamma_2}\mathbf{1}\left[\sum_i a_{i,\gamma_1(i),\gamma_2(i)} = x, \sum_i a_{i,\gamma_1\circ\tau_1(i_1,i_2,i_3)(i),\gamma_2\circ\tau_2(i_1,i_2,i_3)(i)} = x'\right] \\
&+ \frac{1}{2n!^2}\frac{1}{n(n-1)(n-2)} \\
&\sum_{i_1,i_2,i_3 \in C_3}\sum_{\gamma_1,\gamma_2}\mathbf{1}\left[\sum_i a_{i,\gamma_1(i),\gamma_2(i)} = x, \sum_i a_{i,\gamma_1\circ\tau_2(i_1,i_2,i_3)(i),\gamma_2\circ\tau_1(i_1,i_2,i_3)(i)} = x'\right] \qquad (6.5.1)\\
&= \frac{1}{2n!^2}\frac{1}{n(n-1)(n-2)} \\
&\sum_{i_1,i_2,i_3 \in C_3}\sum_{\gamma_1,\gamma_2}\mathbf{1}\left[\sum_i a_{i,\gamma_1\circ\tau_1(i_1,i_2,i_3)(i),\gamma_2\circ\tau_2(i_1,i_2,i_3)(i)} = x, \sum_i a_{i,\gamma_1(i),\gamma_2(i)} = x'\right] \\
&+ \frac{1}{2n!^2}\frac{1}{n(n-1)(n-2)} \\
&\sum_{i_1,i_2,i_3 \in C_3}\sum_{\gamma_1,\gamma_2}\mathbf{1}\left[\sum_i a_{i,\gamma_1\circ\tau_2(i_1,i_2,i_3)(i),\gamma_2\circ\tau_1(i_1,i_2,i_3)(i)} = x, \sum_i a_{i,\gamma_1(i),\gamma_2(i)} = x'\right]
\end{aligned}
$$

□

We now give an expression for $\mathrm{E}\left[T_2 - T_2' \,|T_2\right]$. First observe that $\mathrm{E}\left[T_2' \,|T_2\right] = \mathrm{E}\left[\mathrm{E}\left[T_2' \,|\sigma, \pi\right] |T_2\right]$. We at first find an expression for $\mathrm{E}[T_2' \,|\sigma, \pi]$.

$$
\begin{aligned}
&\mathrm{E}\left[T_2' \,|\sigma, \pi\right] \\
&= \mathrm{E}\left[T_2 - \sum_{j=1}^{3} a_{I_j,\sigma(I_j),\pi(I_j)} + \frac{1}{2}\left[\sum_{j=1}^{3} a_{I_j,\sigma\circ\tau_1(I_1,I_2,I_3)(I_j),\pi\circ\tau_2(I_1,I_2,I_3)(I_j)}\right]\right. \qquad (6.5.2)\\
&\qquad \left. + \frac{1}{2}\left[\sum_{j=1}^{3} a_{I_j,\sigma\circ\tau_2(I_1,I_2,I_3)(I_j),\pi\circ\tau_1(I_1,I_2,I_3)(I_j)}\right] |\sigma, \pi\right]
\end{aligned}
$$

We deal separately with the terms in the last expression. First observe that for any $j \in$

$\{1, 2, 3\}$, we have

$$\mathrm{E}\left[\sum_{j=1}^{3} a_{I_j,\sigma(I_j),\pi(I_j)}\,|\sigma,\pi\right] = \frac{3}{n}T_2.$$

As a consequence,

$$\mathrm{E}\left[\sum_{j=1}^{3} a_{I_j,\sigma(I_j),\pi(I_j)}\,|T_2\right] = \frac{3}{n}T_2.$$

Now, for any $j \in \{1, 2, 3\}$,

$$\begin{aligned}&\mathrm{E}\left[a_{I_j,\sigma\circ\tau_1(I_1,I_2,I_3)(I_j),\pi\circ\tau_2(I_j)}\,|\sigma,\pi\right]\\ &= \frac{1}{n(n-1)(n-2)}\sum_{(i_1,i_2,i_3)\in C_3} a_{i_1,\sigma(i_2),\pi(i_3)}.\end{aligned} \tag{6.5.3}$$

The same thing happens when τ_1 and τ_2 are interchanged.

$$\begin{aligned}&\sum_{j=1}^{3}\mathrm{E}\left[\frac{1}{2}\left[a_{I_j,\sigma\circ\tau_1(I_1,I_2,I_3)(I_j),\pi\circ\tau_2(I_j)} + \left[a_{I_j,\sigma\circ\tau_2(I_1,I_2,I_3)(I_j),\pi\circ\tau_1(I_j)}\right]\,|\sigma,\pi\right]\right.\\ &= 3\frac{1}{n(n-1)(n-2)}\sum_{(i_1,i_2,i_3)\in C_3} a_{i_1,\sigma(i_2),\pi(i_3)}\\ &= 3\frac{n^2}{n(n-1)(n-2)}\mathrm{E}[T_2]-\\ &3\frac{1}{n(n-1)(n-2)}\sum_{(i_1,i_2,i_3)\in C_2} a_{i_1,\sigma(i_2),\pi(i_3)} - 3\frac{1}{n(n-1)(n-2)}\sum_i a_{i,\sigma(i),\pi(i)}\end{aligned} \tag{6.5.4}$$

Since $0 \leq a_{i,j,k} \leq 1$, one has

$$0 \leq \sum_{(i_1,i_2,i_3)\in C_2} a_{i_1,\sigma(i_2),\pi(i_3)} \leq 3n(n-1).$$

As a consequence, we have

$$\begin{aligned}&T_2 - \frac{3}{n}T_2 + 3\frac{n}{(n-1)(n-2)}E[T_2] - 3\frac{1}{n(n-1)(n-2)}T_2\\ &\geq \mathrm{E}\left[T_2'\,|T_2\right]\\ &\geq T_2 - \frac{3}{n}T_2 + 3\frac{n}{(n-1)(n-2)}E[T_2] - 9\frac{1}{n-2} - 3\frac{1}{n(n-1)(n-2)}T_2\end{aligned} \tag{6.5.5}$$

Hence

$$\begin{aligned}
&\frac{3}{n}T_2 - 3\frac{n}{(n-1)(n-2)}\mathrm{E}[T_2] + 3\frac{1}{n(n-1)(n-2)}T_2 \\
&\leq \mathrm{E}\left[T_2 - T_2' \,|T_2\right] \\
&\leq \frac{3}{n}T_2 - 3\frac{n}{(n-1)(n-2)}E[T_2] + 9\frac{1}{n-2} + 3\frac{1}{n(n-1)(n-2)}T_2
\end{aligned} \tag{6.5.6}$$

As a consequence,

$$\begin{aligned}
&\left(\frac{3}{n} + \frac{3}{n(n-1)(n-2)}\right)T_2 - 3\frac{n}{(n-1)(n-2)}\mathrm{E}[T_2] + 3\frac{1}{n(n-1)(n-2)}T_2 \\
&\leq \mathrm{E}\left[T_2 - T_2' \,|T_2\right] \\
&\leq \left(\frac{3}{n} + \frac{3}{n(n-1)(n-2)}\right)T_2 - 3\frac{n}{(n-1)(n-2)}\mathrm{E}[T_2] + 9\frac{1}{n-2} + 3\frac{1}{n(n-1)(n-2)}T_2
\end{aligned} \tag{6.5.7}$$

Writing

$$\begin{aligned}
&\left(\frac{3}{n} + \frac{3}{n(n-1)(n-2)}\right)T_2 - 3\frac{n}{(n-1)(n-2)}\mathrm{E}[T_2] \\
&= 3\left(\frac{n^2-3n+3}{n(n-1)(n-2)}\right)(T_2 - \mathrm{E}[T_2]) - \frac{9}{n(n-2)}\mathrm{E}[T_2],
\end{aligned} \tag{6.5.8}$$

we have

$$\begin{aligned}
&3\left(\frac{n^2-3n+3}{n(n-1)(n-2)}\right)(T_2 - \mathrm{E}[T_2]) - \frac{9}{n(n-2)}\mathrm{E}[T_2] \\
&\leq \mathrm{E}\left[T_2 - T_2' \,|T_2\right] \\
&\leq 3\left(\frac{n^2-3n+3}{n(n-1)(n-2)}\right)(T_2 - \mathrm{E}[T_2]) - \frac{9}{n(n-2)}\mathrm{E}[T_2] + \frac{9}{n-2}.
\end{aligned} \tag{6.5.9}$$

We now have

$$F(T_2, T_2') = \frac{n(n-1)(n-2)}{3(n^2-3n+3)}(T_2 - T_2'). \tag{6.5.10}$$

Hence

$$
\begin{aligned}
f(T_2) &= \mathrm{E}\left[F(T_2,T_2')\,|T_2\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[F(T_2,T_2')\,|\sigma,\pi\right]|T_2\right] \\
&= \frac{n(n-1)(n-2)}{3(n^2-3n+3)}\mathrm{E}\left[\mathrm{E}\left[\left(T_2-T_2'\right)|\sigma,\pi\right]|T_2\right] \\
&= \frac{n(n-1)(n-2)}{3(n^2-3n+3)}\mathrm{E}\left[\left(3\left(\frac{n^2-3n+3}{n(n-1)(n-2)}\right)(T_2-\mathrm{E}(T_2))-\frac{9}{n(n-2)}\mathrm{E}[T_2]\right.\right. \\
&\qquad\left.\left.+\frac{3}{n(n-1)(n-2)}\sum_{(i_1,i_2,i_3)\in C_3}a_{i_1,\sigma(i_2),\pi(i_3)}\right)|\sigma,\pi\;|T_2\right]
\end{aligned}
\tag{6.5.11}
$$

So

$$
\begin{aligned}
f(T_2)-f(T_2') &\leq (T_2-T_2')+\frac{1}{n^2-3n+3}\mathrm{E}\left[\sum_{(i_1,i_2,i_3)\in C_3}a_{i_1,\sigma(i_2),\pi(i_3)}\,|T_2\right] \\
&\leq (T_2-T_2')+3+O\left(\frac{1}{n}\right).
\end{aligned}
\tag{6.5.12}
$$

Also

$$
(T_2-\mathrm{E}(T_2))-3+O\left(\frac{1}{n}\right)\leq f(T_2)\leq (T_2-\mathrm{E}(T_2))+3+O\left(\frac{1}{n}\right).
\tag{6.5.13}
$$

So

$$
\begin{aligned}
v(T_2) &= \frac{1}{2}\left[\left|\left(f(T_2)-f(T_2')\right)F(T_2,T_2')\right|\,|T_2\right] \\
&\leq \frac{n(n-1)(n-2)}{6(n^2-3n+3)}\mathrm{E}\left[(T_2-T_2')^2\,|T_2\right]+\frac{n(n-1)(n-2)}{2(n^2-3n+3)}\left(1+O\left(\frac{1}{n}\right)\right)\mathrm{E}\left[\left|T_2-T_2'\right|\,|T_2\right]
\end{aligned}
\tag{6.5.14}
$$

Now

$$\begin{aligned}
&\mathrm{E}\left[\left|T_2 - T_2'\right| |T_2\right] \\
&= \mathrm{E}\left[\left|\sum_{j=1}^{3} a_{I_j,\sigma(I_j),\pi(I_j)} - \sum_{j=1}^{3} a_{I_j,\sigma'(I_j),\pi'(I_j)}\right| |T_2\right] \\
&\leq \mathrm{E}\left[\sum_{j=1}^{3} a_{I_j,\sigma(I_j),\pi(I_j)} + \sum_{j=1}^{3} a_{I_j,\sigma'(I_j),\pi'(I_j)} \,|T_2\right] \\
&\leq \frac{3}{n}T_2 + \frac{3n}{(n-1)(n-2)}\mathrm{E}[T_2]
\end{aligned} \tag{6.5.15}$$

and

$$\begin{aligned}
&\mathrm{E}\left[\left(T_2 - T_2'\right)^2 |T_2\right] \\
&\leq \frac{9}{n}T_2 + \frac{9n}{(n-1)(n-2)}\mathrm{E}[T_2].
\end{aligned} \tag{6.5.16}$$

Here we have used the fact for $0 \leq a, b \leq 3$, $(a-b)^2 \leq 3(a+b)$. Hence we have

$$\begin{aligned}
v(T_2) &\leq \frac{3n(n-1)(n-2)}{(n^2-3n+3)}\left(\frac{1}{n}T_2 + \frac{n}{(n-1)(n-2)}\mathrm{E}[T_2]\right) \\
&= \left(3 + O\left(\frac{1}{n}\right)\right)(T_2 + \mathrm{E}[T_2]) \\
&= \left(3 + O\left(\frac{1}{n}\right)\right)\left(f(T_2) + 2\mathrm{E}[T_2] + 3 + O\left(\frac{1}{n}\right)\right)
\end{aligned} \tag{6.5.17}$$

Now by Lemma 27 we have for $t > 0$,

$$\begin{aligned}
\mathbb{P}\left[f(T_2) > t\right] &\leq \exp\left\{-\frac{t^2}{\left(6 + O\left(\frac{1}{n}\right)\right)t + 12\mathrm{E}[T_2] + 18 + O\left(\frac{1}{n}\right)}\right\} \\
\Rightarrow \mathbb{P}\left[T_2 - \mathrm{E}[T_2] > t\right] &\leq \mathbb{P}\left[f(T_2) + 3 + O\left(\frac{1}{n}\right) > t\right] \\
\Rightarrow \mathbb{P}\left[T_2 - \mathrm{E}[T_2] > t\right] &\leq \exp\left\{-\frac{(t - 3 + O(1/n))^2}{\left(6 + O\left(\frac{1}{n}\right)\right)(t-3) + 12\mathrm{E}[T_2] + 18 + O(1/n)}\right\}
\end{aligned} \tag{6.5.18}$$

Similarly

$$\mathbb{P}\left[T_2 - \mathrm{E}[T_2] < -t\right] \leq \exp\left\{-\frac{(t-3+O(1/n))^2}{\left(6+O\left(\frac{1}{n}\right)\right)(t-3)+12\mathrm{E}[T_2]+18+O(1/n)}\right\}. \quad (6.5.19)$$

Hence the proof is completed by union bound.

CHAPTER 7 : Future Scopes

We discuss some preliminary ideas about local weak convergence of dense Wigner matrices. Suppose we have a symmetric Wigner matrix of dimension $n \times n$ $A = \frac{1}{\sqrt{n}} \left(X_{i,j}\right)_{1 \leq i,j \leq n}$ where $\left(X_{i,j}\right)_{1 \leq i < j \leq n}$ are i.i.d. standard Gaussian. Often times one comes with weighted graph isomorphism random variables. For a fixed multi graph H we define the weighted graph isomprphism of H as follows:

$$\langle A, H \rangle = \sum_{i_0, \ldots, i_{\#V(H)} \text{ distinct}} \prod_{e \in E(H)} \left(\frac{X_{i_{e_1}, i_{e_2}}}{\sqrt{n}} \right)^{\# N_e} \tag{7.0.1}$$

where for any edge $e \in E(H)$ e_1 and e_2 are two endpoints of the edge e and $N(e)$ denotes the number of times e is traversed. We fix a vertex i and we are interested in understanding the process rooted i. For an exceptional vertex i we consider a rooted multi graph $H(\rho)$ rooted at ρ. We consider the rooted graph isomorphism of $H(\rho)$ as follows:

$$\langle A(i), H(\rho) \rangle = \sum_{i_\rho = i, \ldots, i_{\#V(H)} \text{ distinct}} \prod_{e \in E(H(\rho))} \left(\frac{X_{i_{e_1}, i_{e_2}}}{\sqrt{n}} \right)^{\# N_e}. \tag{7.0.2}$$

We want to understand the process rooted at i indexed by $H(\rho)$ at the limit. The basic intuition is to introduce a sparse random graph as a proxy and use exchangeability. We believe $\mathcal{G}(i)$ can be recursively expressed as a function of $\mathcal{G}(j)$ and the edge weights incident to i.

Proposition 20. *(Janson's second moment method): Let* $\mathbb{P}_n$ *and* $\mathbb{Q}_n$ *be two sequences of probability measures such that for each* n*, both are defined on the common* σ*-algebra* $(\Omega_n, \mathcal{F}_n)$*. Suppose that for each* $i \geq 1$, $W_{n,i}$ *are random variables defined on* $(\Omega_n, \mathcal{F}_n)$*. Then the probability measures* $\mathbb{P}_n$ *and* $\mathbb{Q}_n$ *are asymptotically mutually contiguous if the following conditions hold simultaneously:*

(i) $\mathbb{Q}_n$ *is absolutely continuous with respect to* $\mathbb{P}_n$ *for each* n*;*

(ii) For any fixed $k \geq 1$*, one has* $(W_{n,1}, \ldots, W_{n,k})\,|\mathbb{P}_n \xrightarrow{d} (Z_1, \ldots, Z_k)$ *and* $(W_{n,1}, \ldots, W_{n,k})\,|\mathbb{Q}_n \xrightarrow{d} (Z'_1, \ldots, Z'_k)$*.*

(iii) $Z_i \sim N(0, \sigma_i^2)$ *and* $Z'_i \sim N(\mu_i, \sigma_i^2)$ *are sequences of independent random variables.*

(iv) The likelihood ratio statistic $Y_n = \frac{\mathrm{d}\mathbb{Q}_n}{\mathrm{d}\mathbb{P}_n}$ *satisfies*

$$\limsup_{n\to\infty} \mathrm{E}_{\mathbb{P}_n}\left[Y_n^2\right] \leq \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\} < \infty. \tag{.0.3}$$

In addition, we have that under $\mathbb{P}_n$*,*

$$Y_n \xrightarrow{d} \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i Z_i - \frac{1}{2}\mu_i^2}{\sigma_i^2}\right\}. \tag{.0.4}$$

Furthermore, given any $\epsilon, \delta > 0$ *there exists a natural number* $K = K(\delta, \epsilon)$ *such that for any sequence* n_l *there is a further subsequence* n_{l_m} *such that*

$$\limsup_{m\to\infty} \mathbb{P}_{n_{l_m}}\left(\left|\log(Y_{n_{l_m}}) - \sum_{k=1}^{K} \frac{2\mu_k W_{n_{l_m},k} - \mu_k^2}{2\sigma_k^2}\right| \geq \epsilon\right) \leq \delta. \tag{.0.5}$$

Proof. We only give the proof of (.0.5) since this is the only part where the extra condition was used. **proof of (.0.5)** First of all observe that the limiting random variable

$$\sum_{k=1}^{\infty} \frac{\mu_k Z_k - \frac{1}{2}\mu_k^2}{\sigma_k^2}$$

is well defined from the fact that L^2 spaces are complete. In particular observe that

$$\sup_k \mathrm{E}\left[\left(\sum_{k=1}^{K} \frac{\mu_k Z_k - \frac{1}{2}\mu_k^2}{\sigma_k^2}\right)^2\right] < C.$$

Hence there exists a unique M invariant over K such that given any $\delta > 0$

$$\begin{aligned} \mathbb{P}\left[-M < \sum_{k=1}^{K} \frac{\mu_k Z_k - \frac{1}{2}\mu_k^2}{\sigma_k^2} < M\right] &\leq \frac{\delta}{100} \\ \mathbb{P}\left[-M < \log(L) < M\right] &\leq \frac{\delta}{100}. \end{aligned} \tag{.0.6}$$

On this interval $(\exp\{-M\}, \exp\{M\})$ choose $\tilde{\epsilon}$ such that

$$\begin{aligned} |x-y| \leq \tilde{\epsilon} &\Rightarrow |\log(x) - \log(y)| \leq \epsilon \\ \Leftrightarrow |x-y| > \tilde{\epsilon} &\Leftarrow |\log(x) - \log(y)| > \epsilon. \end{aligned} \tag{.0.7}$$

Now choose K so large that

$$\exp\left\{\sum_{k=1}^{\infty} \frac{\mu_k^2}{\sigma_k^2}\right\} - \exp\left\{\sum_{k=1}^{K} \frac{\mu_k^2}{\sigma_k^2}\right\} < \frac{\delta\tilde{\epsilon}^2}{100}.$$

We have seen that given any sequence $(Y_{n_l}, W_{n_l,1}, \ldots, W_{n_l,K})$ is tight for any given K. We know that there is a further sub-sequence n_{l_m} such that $(Y_{n_{l_m}}, W_{n_{l_m},1}, \ldots, W_{n_{l_m},K})$ converges jointly in distribution to

$$(Y_{n_{l_m}}, W_{n_{l_m},1}, \ldots, W_{n_{l_m},K}) \xrightarrow{d} (H_1, H_2, \ldots, H_{K+1}) \in (\Omega\{n_{l_m}\}, \mathcal{F}\{n_{l_m}\}, \mathbb{P}\{n_{l_m}\}).$$

Let $\mathcal{F}\{n_{l_m}, 1\} \subset \mathcal{F}\{n_{l_m}\}$ be the sigma algebra generated by $(H_2, \ldots, H_{K+1})$. Here $H_1 \stackrel{d}{=} L$ and $(H_2, \ldots, H_{K+1}) \stackrel{d}{=} (Z_1, \ldots, Z_K)$. Using the arguments same as the previous proof we see that

$$\mathrm{E}\left[H_1 \,|\mathcal{F}_{n_1,1}\right] = \exp\left\{\sum_{k=1}^{K} \frac{2\mu_k H_{k+1} - \mu_k^2}{2\sigma_k^2}\right\}.$$

As a consequence, we have

$$0 \le \mathrm{E}\left(H_1 - \exp\left\{\sum_{k=1}^{K}\frac{2\mu_k H_{k+1} - \mu_k^2}{2\sigma_k^2}\right\}\right)^2 \le \exp\left\{\sum_{k=1}^{\infty}\frac{\mu_k^2}{\sigma_k^2}\right\} - \exp\left\{\sum_{k=1}^{K}\frac{\mu_k^2}{\sigma_k^2}\right\}.$$

Now by Chebyshev's inequality

$$\mathbb{P}\left[\left|H_1 - \exp\left\{\sum_{k=1}^{K}\frac{2\mu_k H_{k+1} - \mu_k^2}{2\sigma_k^2}\right\}\right| \ge \frac{\tilde{\epsilon}}{2}\right] \le \frac{\delta\tilde{\epsilon}^2}{25\tilde{\epsilon}^2} = \frac{\delta}{100}.$$

Now observe that

$$\limsup_{n_{l_m}} \mathbb{P}_{n_{l_m}}\left[Y_{n_{l_m}} \notin (e^{-M}, e^{M})\right] \le \mathbb{P}\left[H_1 \notin (e^{-M}, e^{M})\right] \le \frac{\delta}{100}$$

and

$$\limsup_{n_{l_m}} \mathbb{P}_{n_{l_m}}\left[\exp\left\{\sum_{k=1}^{K}\frac{2\mu_k W_{n_{l_m}} - \mu_k^2}{2\sigma_k^2}\right\} \notin (e^{-M}, e^{M})\right] \le \frac{\delta}{100}$$

Since

$$\left(Y_{n_{l_m}}, W_{n_{l_m},1}, \ldots, W_{n_{l_m},k}\right) \xrightarrow{d} (H_1, H_2, \ldots, H_{k+1})$$

by continuous mapping theorem for in distributional convergence, we have

$$Y_{n_{l_m}} - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2}\right\} \xrightarrow{d} H_1 - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2}\right\}.$$

Since the set $[\frac{\tilde{\epsilon}}{2}, \infty)$ is closed, we have by Portmanteau theorem,

$$\begin{aligned}
&\limsup_{n_{l_m}} \mathbb{P}_{n_{l_m}}\left[\left|Y_{n_{l_m}} - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2}\right\}\right| > \tilde{\epsilon}\right] \\
&\le \limsup_{n_{l_m}} \mathbb{P}_{n_{l_m}}\left[\left|Y_{n_{l_m}} - \exp\left\{\sum_{i=1}^{k}\frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2}\right\}\right| \ge \frac{\tilde{\epsilon}}{2}\right] \\
&\le \frac{\delta}{25}.
\end{aligned} \tag{.0.8}$$

□

As a consequence,

$$
\begin{aligned}
\frac{\delta}{25} &\geq \limsup \mathbb{P}_{n_{l_m}} \left[\left| Y_{n_{l_m}} - \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i W_{n_{l_m},i} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \tilde{\epsilon} \right] \\
&\geq \limsup \mathbb{P}_{n_{l_m}} \left[Y_{n_{l_m}} \in [e^{-M}, e^{M}] \cap \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \in [e^{-M}, e^{M}] \right. \\
&\qquad \left. \cap \left| Y_{n_{l_m}} - \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \tilde{\epsilon} \right] \\
&\geq \limsup \mathbb{P}_{n_{l_m}} \left[Y_{n_{l_m}} \in [e^{-M}, e^{M}] \cap \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \in [e^{-M}, e^{M}] \right. \\
&\qquad \left. \cap \left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \epsilon \right] \\
&\geq \limsup 1 - \mathbb{P}_{n_{l_m}} \left[\left(Y_{n_{l_m}} \in [e^{-M}, e^{M}] \cap \exp\left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \in [e^{-M}, e^{M}] \right)^c \right] \\
&\qquad\qquad - \mathbb{P}_{n_{l_m}} \left[\left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| \leq \epsilon \right] \\
&\geq \limsup \left(\mathbb{P}_{n_{l_m}} \left[\left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \epsilon \right] - \frac{\delta}{100} \right) \\
&\Rightarrow \limsup \mathbb{P}_{n_{l_m}} \left[\left| \log\left(Y_{n_{l_m}}\right) - \left\{ \sum_{i=1}^{k} \frac{2\mu_i H_{i+1} - \mu_i^2}{2\sigma_i^2} \right\} \right| > \epsilon \right] \leq \frac{\delta}{25} + \frac{\delta}{100} < \delta.
\end{aligned}
\tag{.0.9}
$$

BIBLIOGRAPHY

E. Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, To appear, 2017.

E. Abbe and C. Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic BP, and the information-computation gap. *ArXiv e-prints*, Dec. 2015. URL `https://arxiv.org/abs/1512.09080`.

E. Abbe and C. Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 670–688. IEEE, 2015.

E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *CoRR*, abs/1405.3267, 2014. URL `http://arxiv.org/abs/1405.3267`.

M. Aizenman, J. L. Lebowitz, and D. Ruelle. Some rigorous results on the sherrington-kirkpatrick spin glass model. *Communications in mathematical physics*, 112(1):3–20, 1987.

G. W. Anderson and O. Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006.

G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

Z. Bai, X. Wang, and W. Zhou. CLT for linear spectral statistics of Wigner matrices. *Electron. J. Probab*, 14(83):2391–2417, 2009.

Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605, 2004.

J. Baik and J. O. Lee. Fluctuations of the free energy of the spherical sherrington–kirkpatrick model. *Journal of Statistical Physics*, 165(2):185–224, 2016.

J. Baik and J. O. Lee. Fluctuations of the free energy of the spherical sherrington–kirkpatrick model with ferromagnetic interaction. *Annales Henri Poincaré*, 18(6):1867–1917, 2017.

J. Baik, G. B. Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

J. Baik, J. O. Lee, and H. Wu. Ferromagnetic to paramagnetic transition in spherical spin glass. *Journal of Statistical Physics*, 173(5):1484–1522, 2018.

D. Banerjee. Contiguity and non-reconstruction results for planted partition models: the dense case. *Electronic Journal of Probability*, 23, 2018.

D. Banerjee and A. Bose. Largest eigenvalue of large random block matrices: a combinatorial approach. *Tech. Report R1/2016 Stat-Math Unit, Indian Statistical Institute, Kolkata*, 2016. URL http://www.isical.ac.in/~statmath/report/11601-blockmatrixfinaltechrepr12016.pdf.

D. Banerjee and Z. Ma. Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv preprint arXiv:1705.05305*, 2017a.

D. Banerjee and Z. Ma. Supplement to "Optimal hypothesis testing for stochastic block models with growing degrees". 2017b.

J. Banks, C. Moore, J. Neeman, and P. Netrapalli. Information-theoretic thresholds for community detection in sparse networks. *ArXiv e-prints*, July 2016. URL https://arxiv.org/abs/1607.01760.

P. J. Bickel and A. Chen. A nonparametric view of network models and newmangirvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

P. J. Bickel and P. Sarkar. Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78 (1):253–273, 2016.

R. B. Boppana. Eigenvalues and graph bisection: An average-case analysis. *In 28th Annual Symposium on Foundations of Computer Science*, pages 280–285, 1987.

C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. *ArXiv e-prints*, Jan. 2015. URL http://arxiv.org/pdf/1501.06087v2.pdf.

S. Bubeck, J. Ding, R. Eldan, and M. Rácz. Testing for high-dimensional geometry in random graphs. *ArXiv e-prints*, Nov. 2014. URL http://arxiv.org/abs/1411.5713.

T. N. Bui, S. Chaudhuri, F. T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, 1987.

T. Cai, Z. Ma, and Y. Wu. Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability theory and related fields*, 161(3-4):781–815, 2015.

T. T. Cai and Z. Ma. Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388, 2013.

M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large wigner matrices: convergence and nonuniversality of the fluctuations. *The Annals of Probability*, pages 1–47, 2009.

T. Carleman. Les fonctions quasi analytiques(in French). Leçons professées au Collège de France. 1926.

S. Chatterjee. *Concentration inequalities with exchangeable pairs.* PhD thesis, Citeseer, 2005.

S. Chatterjee. Steins method for concentration inequalities. *Probability theory and related fields*, 138(1):305–321, 2007.

S. Chatterjee. A general method for lower bounds on fluctuations of random variables. *arXiv preprint arXiv:1706.04290*, 2017.

K. Chen and J. Lei. Network cross-validation for determining the number of communities in network data. *arXiv preprint arXiv:1411.1715*, 2014.

A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Combinatorics, Probability & Computing*, 19(2):227–284, 2010.

F. Comets and J. Neveu. The sherrington-kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. *Communications in Mathematical Physics*, 166(3): 549–564, 1995.

A. Condon and R. M. Karp. *Algorithms for Graph Partitioning on the Planted Partition Model*, pages 221–232. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

B. Dabbs and B. Junker. Comparison of cross-validation methods for stochastic block models. *arXiv preprint arXiv:1605.03000*, 2016.

J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Statistics and computing*, 18(2):173–183, 2008.

A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physics Review E*, 84(6):066106, Dec. 2011. URL `https://arxiv.org/abs/1109.3041`.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

P. Dharmawansa, I. M. Johnstone, and A. Onatski. Local Asymptotic Normality of the spectrum of high-dimensional spiked F-ratios. *arXiv preprint arXiv:1411.3875*, 2014.

E. Dobriban. Sharp detection in pca under correlations: all eigenvalues matter. *The Annals of Statistics*, 45(4):1810–1833, 2017.

M. E. Dyer and A. M. Frieze. The solution of some random np-hard problems in polynomial expected time. *J. Algorithms*, 10(4):451–489, Dec. 1989.

L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Communications in Mathematical Physics*, 314(3):587–640, 2012.

L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős–Rényi graphs I: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.

Z. Fan and A. Montanari. How well do local algorithms solve semidefinite programs? In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 604–614. ACM, 2017.

D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2007.

Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

C. Gao and J. Lafferty. Testing network structure using relations between small subgraph probabilities. *arXiv preprint arXiv:1704.06742*, 2017.

C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block model. *arXiv preprint arXiv:1505.03772*, 2015.

C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Community detection in degree-corrected block models. *arXiv preprint arXiv:1607.06993*, 2016.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016a.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Transactions on Information Theory*, 62(10):5918–5937, 2016b.

W. Hoeffding et al. A combinatorial central limit theorem. *The Annals of Mathematical Statistics*, 22(4):558–566, 1951.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

D. Hsu, S. Kakade, T. Zhang, et al. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17, 2012.

L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

S. Janson. Random regular graphs: asymptotic distributions and contiguity. *Combin. Probab. Comput.*, 4(4):369–405, 1995.

V. Jog and P.-L. Loh. Information-theoretic bounds for exact recovery in weighted stochastic block models using the Rényi divergence. *arXiv preprint arXiv:1509.06418*, 2015.

S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

I. M. Johnstone and A. Onatski. Testing in high-dimensional spiked models. *arXiv preprint arXiv:1509.07269*, 2015.

W. Lang. On polynomials related to powers of the generating function of catalan's numbers. *Fibonacci Quarterly*, 38(5):408–419, 2000.

P. Latouche, E. Birmele, and C. Ambroise. Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1):93–115, 2012.

C. M. Le and E. Levina. Estimating the number of communities in networks by spectral methods. *arXiv preprint arXiv:1507.00827*, 2015.

L. Le Cam. *Asymptotic methods in statistical decision theory*. Springer Science & Business Media, 2012.

L. Le Cam and G. L. Yang. *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media, 2012.

J. O. Lee and K. Schnelli. Local law and tracy–widom limit for sparse random matrices. *Probability Theory and Related Fields*, pages 1–74, 2016.

J. Lei. A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, 44(1): 401–424, 2016.

C. L. Mallows. A note on asymptotic joint normality. *Ann. Math. Statist.*, 43(2):508–515, 1972.

L. Massoulié. Community detection thresholds and the weak ramanujan property. *CoRR*, abs/1311.3085, 2013. URL `http://arxiv.org/abs/1311.3085`.

B. Maurey. Construction de suites symétriques. *CR Acad. Sci. Paris Sér. AB*, 288(14): A679–A681, 1979.

C. McDiarmid. Concentration for independent permutations. *Combinatorics, Probability and Computing*, 11(2):163–178, 2002.

F. McSherry. Spectral partitioning of random graphs. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529–537, Oct 2001.

E. Mossel, J. Neeman, and A. Sly. A Proof Of The Block Model Threshold Conjecture. *ArXiv e-prints*, Nov. 2013. URL `https://arxiv.org/abs/1311.4115`.

E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for the planted bisection model. *Electron. J. Probab.*, 21:1–24, 2016.

M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2566–2572, 2002.

G. E. Noether. On a theorem by wald and wolfowitz. *The Annals of Mathematical Statistics*, 20(3):455–458, 1949.

A. Onatski, M. J. Moreira, and M. Hallin. Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics*, 41(3):1204–1231, 2013.

A. Onatski, M. J. Moreira, and M. Hallin. Signal detection in high dimension: The multi-spiked case. *The Annals of Statistics*, 42(1):225–254, 2014.

D. Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013.

T. P. Peixoto. Parsimonious module inference in large networks. *Physical review letters*, 110(14):148701, 2013.

A. Pizzo, D. Renfrew, and A. Soshnikov. On finite rank deformations of Wigner matrices. In *Annales de L'Institut Henri Poincare, Probabilites Et Statistiques*, volume 49, pages 64–94. Institut Henri Poincaré, 2013.

J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.*, 39(4):1878–1915, 08 2011.

D. Saldana, Y. Yu, and Y. Feng. How many communities are there? *Journal of Computational and Graphical Statistics*, 26:171–181, 2017.

J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, Aug. 2000.

M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.

M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.

M. Talagrand. The parisi formula. *Annals of mathematics*, pages 221–263, 2006.

C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996.

V. H. Vu. Spectral norm of random matrices. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 423–430. ACM, 2005.

A. Wald and J. Wolfowitz. Statistical tests based on permutations of the observations. *The Annals of Mathematical Statistics*, 15(4):358–372, 1944.

Y. R. Wang, P. J. Bickel, et al. Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2):500–528, 2017.

G. C. Wick. The evaluation of the collision matrix. *Phys. Rev.*, 80:268–272, Oct 1950.

N. C. Wormald. Models of random regular graphs. In J. D. Lamb and D. A. Preece, editors, *Surveys in Combinatorics, 1999*, pages 239–298. Cambridge University Press, 1999.

S.-Y. Yun and A. Proutiere. Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems*, pages 965–973, 2016.

A. Y. Zhang, H. H. Zhou, et al. Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280, 2016.