Information-theoretic Limits for Testing Community Structures in Weighted Networks

Mingao Yuan\textsuperscript{1} and Zuofeng Shang\textsuperscript{2}

\textsuperscript{1}Department of Statistics, North Dakota State University; e-mail: mingao.yuan@ndsu.edu

\textsuperscript{2}Department of Mathematical Sciences, New Jersey Institute of Technology; e-mail: zshang@njit.edu

Abstract: Community detection refers to the problem of clustering the nodes of a network into groups. Existing inferential methods for community structure mainly focus on unweighted (binary) networks. Many real-world networks are nonetheless weighted and a common practice is to dichotomize a weighted network to an unweighted one which is known to result in information loss. Literature on hypothesis testing in the latter situation is still missing. In this paper, we study the problem of testing the existence of community structure in weighted networks. Our contributions are threefold: (a). We use the (possibly infinite-dimensional) exponential family to model the weights and derive the sharp information-theoretic limit for the existence of consistent test. Within the limit, any test is inconsistent; and beyond the limit, we propose a useful consistent test. (b). Based on the information-theoretic limits, we provide the first formal way to quantify the loss of information incurred by dichotomizing weighted graphs into unweighted graphs in the context of hypothesis testing. (c). We propose several new and practically useful test statistics. Simulation study show that the proposed tests have good performance. Finally, we apply the proposed tests to an animal social network.

MSC2020 subject classifications: Primary 62G10; secondary 05C80.

Keywords and phrases: community detection, weighted network, hypothesis testing.

1. Introduction

In recent decades, network data analysis has attracted increasing attention. One of the most important research topics in network data analysis is to infer the underlying network structures. For instance, in ordinary or hypergraphic stochastic block models (SBM), community detection has been extensively studied by [24, 13, 7, 16, 38, 39, 8, 20, 11, 23, 1, 3, 5], among others. Another research direction is to study hypothesis testing problems regarding the existence of community structures which has been recently studied by [4, 31, 11, 19, 10, 17, 18, 35, 33, 37, 36]. Existing hypothesis testing literature mainly focus on unweighted networks in which an edge between two nodes is either present or absent. Real-world networks are nonetheless often weighted in the sense that the observed edges may be weighted by interaction frequency, volume or similarity, etc; see [1, 2, 5, 32]. In weighted networks, edges can be either discrete or continuous random variables characterizing the strength of connectivity (see [1, 2, 5, 32]). For example, in airport networks, an edge is weighted by the number of airlines between two airports ([14]); in brain-image networks, the
weight can represent the strength of association between two brain regions ([27, 25]). To the best of our knowledge, literature on hypothesis testing in general weighted networks are still missing, with the exception of some empirical or ad hoc studies ([30, 34]). Moreover, a common practice in analyzing weighted networks is to dichotomize the weighted edges to binary ones based on which existing unweighted network techniques can be applied. According to some experimental findings, e.g., [1, 2, 32], dichotomizing the weighted networks may result in information loss, whose impact on hypothesis testing is still largely unclear.

In this paper, we are interested in the problem of testing community structures in general undirected weighted networks in which the distributions of the weighted edges belong to exponential family. Our contributions can be summarized into threefold. (a) We use the general (possibly infinite-dimensional) exponential family to model the distribution of network weights and use contiguity theory coupled with second moment method to derive a sharp information-theoretic limit for the existence of a consistent test. Within the limit, any test is proven inconsistent; and beyond the limit, we propose a useful consistent test. (b) We derive a sharp information-theoretic limit for the existence of a consistent test under dichotomized weighted network, which dramatically differs from the one under original weighted network. Such different information-theoretic limits explicitly quantify the impact of information loss incurred by dichotomizing weighted networks in the context of hypothesis testing. (c) We propose useful consistent tests that may outperform the existing empirical or ad hoc approaches proposed in [30, 34]; the latter tests were proposed under the unrealistic assumption that the weights follow some single- or two-parameter distributions whose applications may be restricted. More specifically, the limits in (a) and (b) can be characterized by different ellipsoids, and the radius of the latter is significantly larger than the former under any dichotomizing scheme. The proposed consistent tests are based on weighted signed long mixture cycles (WSLMC) which are new in literature with satisfactory numerical performances. Since the exponential family assumed on the weights is more general than Bernoulli distributions assumed on classic unweighted networks, derivation of the limits is substantially more challenging.

This paper is organized as follows. Section 2 provides a sharp information-theoretic limit for the existence of a consistent test (Section 2.1) and a quantification of information loss when dichotomizing a weighted network (Section 2.2). In Section 2.3, the results are strengthened to obtain a more explicit description of the inconsistency of the test, compared with random guessing, under an additional differential equation assumption. Practical tests are given in Section 3. Section 4 involves numerical studies based on simulated and real data. Technical proofs are deferred to Section 5.

1.1. Problem Formulation

For a positive integer \( n \), let \( V = [n] := \{1, 2, \ldots, n\} \) denote the set of network nodes. Any node \( i \in V \) is assigned, independently and uniformly at random, a label \( \sigma_i \in \{\pm\} \). Let \( P, Q \) be probability
distributions over $\mathbb{R}$ and $A \in \mathbb{R}^{n \times n}$ be a random symmetric matrix with zero diagonal entries. We say that $A$ follows an undirected weighted stochastic block model (WSBM) with weight distributions $P, Q$, denoted $A \sim \mathcal{G}(n, P, Q)$, if $A_{ij}, 1 \leq i < j \leq n$ are independent and satisfying

$$A_{ij} \sim \begin{cases} P, & \text{given } \sigma_i = \sigma_j, \\ Q, & \text{given } \sigma_i \neq \sigma_j. \end{cases}$$

When $P, Q$ are Bernoulli distributions, WSBM degenerates to the classic unweighted SBM considered by [3]. For convenience, we call $A_{ij}$’s the network weights. Clearly, $P = Q$ implies that all network weights are equally distributed regardless of whether a pair of node labels are equal or not, hence, the network is equivalent to one without community structure. In practice, only $A$ is observable while $P, Q$ are not. We are interested in the problem of testing whether $P = Q$ given $A$, namely, the following hypothesis testing problem:

$$H_0 : P = Q \text{ versus } H_1 : P \neq Q.$$ (1)

A statistical test $T(A)$, a real-valued measurable function of $A$, is said to be consistent if it has asymptotic power approaching one as $n$ goes to infinity. Otherwise, the test is said to be inconsistent. One specific aim of this paper is to study under what circumstances there is a consistent test for (1) and how to propose an asymptotically powerful test statistic. This problem will be resolved when $P, Q$ belong to exponential family that covers a broad range of weighted network models. In the same setting, [1, 2, 32] studied community detection problems. Whereas the problem of testing the existence of community structure still remains elusive.

1.2. Notation

For a positive integer $k$, and a vector of non-negative integers $\alpha = (\alpha_1, \ldots, \alpha_k)$, define $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_k$ and $\alpha! = \alpha_1! \cdots \alpha_k!$. For $x = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, denote $x^\alpha = (x_1^{\alpha_1}, x_2^{\alpha_2}, \ldots, x_k^{\alpha_k})$. For a function $f(x)$, denote $\partial^\alpha f(x) = \frac{\partial^{\alpha_1} f(x_1, \ldots, x_k)}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_k^{\alpha_k}}$. Let $Df(x_0)$ and $D^2f(x_0)$ denote the gradient and Hessian of $f$ evaluated at $x_0$, respectively. Let $I(E)$ be the indicator function of an event $E$, $\mathbf{1} = (1, 1, \ldots, 1)^T$ and $\lambda_{\text{max}}(B)$ be the largest eigenvalue of a matrix $B$. Let $D^k_F \psi$ denote the $k$th Fréchet derivative of functional $\psi$ and $D^k_F \psi(f)g^k$ denote the value of multi-linear map $D^k_F \psi(f)$ applied to $(g, g, \ldots, g)^T$. Let $\|x\|_2$ denote the Euclidean norm of $x$.

2. Sharp Information-theoretic Limits and Quantification of Information Loss

It might be challenging to derive sharp information-theoretic limits for testing (1) without any regularity assumptions on $P, Q$. Throughout, we assume that $P, Q$ belong to exponential family, which has also been adopted by [2] in community detection. The parameter dimension of the
expansive family can be either finite or infinite, in both cases we shall derive sharp information-theoretic limits. Moreover, we provide the first formal way to quantify the loss of information incurred by dichotomizing the weighted network in the context of hypothesis testing.

2.1. Sharp Information-theoretic Limit

Definition 2.1. Let $\Theta \subset \mathbb{R}^m$ be an $m$-dimensional subset and, for $\theta \in \Theta$, denote its coordinates $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$. The family of distributions $\{P_{\theta}, \theta \in \Theta\}$ is said to be an expansive family if the probability density of $P_\theta$ has a form

$$f(x; \theta) = h(x)e^{\sum_{i=1}^m \theta_i T_i(x) - \psi(\theta)}, \quad x \in \mathbb{R},$$

where $h(x), \psi(\theta), T_i(x), i = 1, 2, \ldots, m$ are known functions.

Let $P$ and $Q$ satisfy (2) with canonical parameters $\theta_1$ and $\theta_2$, respectively. Let $\tau = (\tau_1, \tau_2, \ldots, \tau_m)^T$ and $d = (d_1, d_2, \ldots, d_m)^T$ be vectors of $m$ fixed constants independent of $n$. Alternatively, we can express $\theta_1$ and $\theta_2$ as follows:

$$\theta_1 = \tau - \frac{\tau_d}{\sqrt{n}}, \quad \theta_2 = \tau + \frac{\tau_d}{\sqrt{n}},$$

where $\tau_d = (\tau_1 d_1, \tau_2 d_2, \ldots, \tau_m d_m)^T$. Under (3), the hypotheses (1) can be rewritten as follows:

$$H_0 : \|d\|_2 = 0 \quad \text{versus} \quad H_1 : \|d\|_2 \neq 0.$$ (4)

Under $H_0$, $\theta_1 = \theta_2$, and hence, $P = Q$. Under $H_1$, $P$ and $Q$ are different since at least one component of $d$ is nonzero.

To derive the sharp limit, firstly we propose a novel test statistic for (7) as follows. For integer $k = \log \log \log n$ and distinct nodes $i_1, i_2, \ldots, i_k$, let $C(i_1, i_2, \ldots, i_k)$ be the set of all the circular permutations of $i_1, i_2, \ldots, i_k$. It is well-known that $C(i_1, i_2, \ldots, i_k)$ has $\frac{(k-1)!}{2}$ elements. Let $\mathcal{I}_n = \{(i_1, i_2, \ldots, i_k) \in C(j_1, j_2, \ldots, j_k) | 1 \leq j_1 < j_2 < \cdots < j_k \leq n\}$. Define the weighted-signed-long-mixture-cycle (WSLMC) test statistic $Z_n$ as follows:

$$Z_n = \frac{\sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^k \frac{\tau_d^T (T(A_{i_{t+1} i_t}) - D\psi(\tau))}{\sqrt{n}}}{\sqrt{\frac{1}{2\pi} \left[ \tau_d^T D^2 \psi(\tau) \tau_d \right]^k}},$$

where $T(x) = (T_1(x), T_2(x), \ldots, T_m(x))^T$ and $i_{k+1} = i_1$. Note that each circular permutation of $(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n$ can be considered as a cycle. In this sense, the numerator of $Z_n$ just counts the number of some weighted cycles. Here “signed” means $T(x)$ is centered by subtracting its mean $D\psi(\tau)$; “weighted” means $T(x) - D\psi(\tau)$ has weight $\tau_d$; “long” means the length $k$ of the circular permutation (or cycle) goes to infinity as $n$ tends to infinity; ‘mixture’ means $\tau_d^T (T(A_{i_{t+1} i_t}) - D\psi(\tau))$ is a weighted sum of $m$ terms. This test statistic is motivated by but
Theorem 2.2. Suppose $\partial^2 \psi$, for $\alpha$ with $|\alpha| = 5$, exist and are uniformly bounded for all $\theta \in \Theta$. Then the following results hold.

(I) If $\tau_d D^2 \psi(\tau)\tau_d < 1$, any test is inconsistent.
(II) If $\tau_d D^2 \psi(\tau)\tau_d > 1$, the WSLMC test is consistent.

Theorem 2.2 says that, when $\tau_d^T D^2 \psi(\tau)\tau_d < 1$, any statistical test for (7) cannot achieve asymptotic power one. When $\tau_d^T D^2 \psi(\tau)\tau_d > 1$, the WSLMC test is consistent and hence optimal in this sense. The solution set $\{d : \tau_d^T D^2 \psi(\tau)\tau_d = 1\}$ is an elliptic curve in $d$, which is the sharp boundary for the existence of a consistent test. When $m = 2$, these regions are demonstrated in Figure 1.

We point out the WSLMC test is not directly applicable in practice, since the vector $\tau_d$ and the function $T(x) = (T_1(x), T_2(x), \ldots, T_m(x))^T$ are unknown. Our results only provide some theoretical insights and serve as a benchmark for developing practical statistical tests. Motivated by $Z_n$, more practical tests shall be proposed in Section 3.

Next, we illustrate our results by restricting $P, Q$ to the exponential distribution and the normal distribution.

Example 1 (Exponential distribution). The exponential distribution has density

$$f(x; \theta) = e^{-\theta x + \log \theta}, \quad \theta > 0,$$

(5)
The exponential distribution belongs to the single-parameter exponential family with $T(x) = -x$, $\psi(\theta) = -\log \theta$. The mean of exponential distribution is $\frac{1}{\theta}$. In this case, (3) is simplified to

$$\theta_1 = \tau - \frac{d}{\sqrt{n}}, \quad \theta_2 = \tau + \frac{d}{\sqrt{n}}. \quad (6)$$

Since $\frac{d^2 \psi(\theta)}{d\theta^2} = \frac{1}{\theta^2}$, then $\tau_d^T D^2 \psi(\theta) \tau_d = d^2$. On the region $d^2 < 1$, there is no consistent.

**Example 2 (Normal distribution).** The normal distribution $N(\mu, \sigma^2)$ has density

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$  

It is a member of the 2-parameter exponential family with $h(x) = \frac{1}{\sqrt{2\pi}}$, $T_1(x) = x$, $T_2(x) = x^2$, $\theta_1 = \frac{\mu}{\sqrt{2}}$, $\theta_2 = -\frac{1}{2\sigma^2}$ and $\psi(\theta_1, \theta_2) = -\frac{\theta_1^2}{2\theta_2} - \frac{1}{2} \log(-2\theta_2)$. Straightforward calculation yields

$$D^2 \psi(\theta) = \begin{bmatrix} -\frac{1}{2\theta_2} & \frac{\theta_1}{\theta_2} \\ \frac{\theta_1}{\theta_2} & -\frac{\theta_1^2}{2\theta_2} + \frac{1}{2\theta_2} \end{bmatrix}.$$  

Then

$$\tau_d^T D^2 \psi(\theta) \tau_d = -\frac{1}{2\theta_2} \tau_1^2 d_1^2 + 2 \frac{\tau_1}{\theta_2} \tau_1 \tau_2 d_1 d_2 + \left( -\frac{\tau_1^2}{2\theta_2^3} + \frac{1}{2\theta_2^2} \right) \tau_2^2 d_2^2.$$  

For $d$ within the ellipsoid $\tau_d^T D^2 \psi(\theta) \tau_d < 1$, there is no consistent test.

Theorem 2.2 assumes the dimension $m$ of the exponential family to be finite. Actually, Theorem 2.2 holds even for infinite-dimensional exponential family defined in [15]:

$$\mathcal{P} = \left\{ p_f(x) = e^{f(x) - \psi(f)} q_0(x), \ x \in \Omega \subset \mathbb{R} : f \in \mathcal{F} \right\},$$  

where $\mathcal{F}$ is a subspace of a reproducing kernel Hilbert space $\mathcal{H}$, $q_0(x)$ is a reference density function and

$$\mathcal{F} = \left\{ f \in \mathcal{H} : e^{\psi(f)} < \infty \right\}, \quad \psi(f) = \log \int_{\Omega} e^{f(x)} q_0(x) dx.$$  

The infinite-dimensional exponential family includes a very broad class of distributions (see [15, 28]).

Assume $P, Q \in \mathcal{P}$ are parametrized by $f_1$ and $f_2$ respectively. Let

$$f_1 = f - \frac{g}{\sqrt{n}}, \quad f_2 = f + \frac{g}{\sqrt{n}},$$  

where $g \in \mathcal{F}$ such that $f_1, f_2 \in \mathcal{F}$. Then the hypotheses (1) can be reformulated as follows:

$$H_0 : \|g\|_\mathcal{H} = 0 \quad \text{versus} \quad H_1 : \|g\|_\mathcal{H} \neq 0. \quad (7)$$

where $\|g\|_\mathcal{H}$ represents the norm of $g$ in the Hilbert space $\mathcal{H}$. In this case, the WSLMC test has the following form:

$$Z_n = \sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^k \frac{g(A_{it+k}) - \mathbb{E}_f(g(A_{it+k}))}{\sqrt{n}} \sqrt{\frac{1}{2\pi} \left[ D^2 \psi(f) g^2 \right]^k},$$

where $\mathbb{E}_f$ represents expectation with respect to density $p_f(x) \in \mathcal{P}$.
Theorem 2.3. Suppose $D^5_F \psi(f)$ exist and are uniformly bounded for all $f \in \mathcal{F}$. Then the following results hold.

(I) If $D^2_F \psi(f) g^2 < 1$, any test is inconsistent.
(II) If $D^2_F \psi(f) g^2 > 1$, the WSLMC test is consistent.

To prove Theorem 2.3, one needs to replace the partial derivatives by Fréchet derivatives, and follows the line of proof of Theorem 2.2. In this sense, our result is pretty general.

2.2. Information Loss of Dichotomizing Weighted Networks

Existing community detection algorithms and statistical tests for community structure are mainly developed for binary edges ([3, 4, 6, 8, 10, 11, 17, 19, 20, 23, 31, 39, 8]). When a network is weighted, a common way is to convert the weighted network to a binary one. It has been empirically verified that there is information loss in the dichotomizing process ([1, 2, 29]). However, it is unclear how much information is lost quantitatively. In this subsection, we provide the first formal quantification of information loss in the context of hypothesis testing via statistical limits.

Given a fixed real number $t_0$, the weights $A_{ij}, i < j$ can be naturally dichotomized to binary ones $\tilde{A}_{ij}$ as follows

$$\tilde{A}_{ij} = I[A_{ij} > t_0], \quad 1 \leq i < j \leq n.$$ 

That is, all the weights smaller than $t_0$ are discarded and weights larger than $t_0$ are converted to ones ([1, 2, 29]). In this way, the weighted work $A$ is converted to an unweighted network $\tilde{A}$. The network $\tilde{A}$ inherits the community structure of $A$. Given $\sigma$ and $\theta(\sigma_i, \sigma_j) = \tau - \frac{D_\sigma}{\sqrt{n}} \sigma_i \sigma_j$, the probability of the presence of an edge in $\tilde{A}$ is

$$p_{ij}(\sigma) = \mathbb{P}(\tilde{A}_{ij} = 1|\sigma) = \mathbb{P}(A_{ij} > t_0|\sigma) = \int_{t_0}^{\infty} h(x) e^{\sum_{i=1}^{n} \theta(\sigma_i, \sigma_j) T(x) - \psi(\sigma_i, \sigma_j)} dx.$$ 

Hence, the hypotheses are still the same as (7). Under $H_0$, there is no community structure and the edge presence probability is

$$p_0 = \int_{t_0}^{\infty} h(x) e^{\tau T(x) - \psi(\tau)} dx.$$ 

To get the sharp testing limit, we propose the signed-long-cycle (SLC) test statistic $R_n$ as

$$R_n = \sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^{k} \left( \tilde{A}_{i_{t+1}} - p_0 \right) \sqrt{\frac{(k-1)! (\binom{n}{k})}{2} \frac{\left[p_0 (1-p_0)\right]^k}{\left[p_0 (1-p_0)\right]^k}}, \quad k = \log \log \log n.$$ 

The SLC test rejects $H_0$ if $|R_n| > C$ for some constant $C$ dependent on the type I error.

Theorem 2.4. Suppose $\partial^\alpha \psi$, for $\alpha$ with $|\alpha| = 5$, exist and are uniformly bounded for all $\theta \in \Theta$. Let

$$a(t_0) = \int_{t_0}^{\infty} h(x) e^{\tau T(x) - \psi(\tau)} (-T(x) + D\psi(\tau)) dx.$$
Then the following results hold.

(I) If \( \frac{(a(t_0)^T \tau_d)^2}{p_0(1-p_0)} < 1 \), any test is inconsistent.

(II) If \( \frac{(a(t_0)^T \tau_d)^2}{p_0(1-p_0)} > 1 \), the SLC test is consistent.

For the dichotomized network, the sharp limit for existence of consistent test is given by the quantity \( \frac{(a(t_0)^T \tau_d)^2}{p_0(1-p_0)} \). In general, \( \frac{(a(t_0)^T \tau_d)^2}{p_0(1-p_0)} \) differs from \( \tau_d D^2 \psi(\tau) \tau_d \) in Theorem 2.2, since it depends on \( t_0 \). The difference between them can be considered as a measure of information loss incurred by dichotomizing a weighted network in the context of hypothesis testing. In this sense, our result provides the first theoretical characterization of loss of information.

In the general case, it is not immediately clearly what is the difference between \( \frac{(a(t_0)^T \tau_d)^2}{p_0(1-p_0)} \) and \( \tau_d D^2 \psi(\tau) \tau_d \). For better illustration, we restrict \( P, Q \) to be exponential distributions. In this case, \( p_0 = e^{-\tau t_0} \), \( a(t_0) = t_0 p_0 \) and \( \frac{(a(t_0)^T \tau_d)^2}{p_0(1-p_0)} = \frac{\tau^2 t_0^2 p_0^2}{e^{\tau t_0} - 1} \). Then the following corollary follows.

**Corollary 2.5.** Suppose \( P, Q \) are exponential distributions given by (5), (6) and \( t_0 > 0 \) is a fixed constant. Then the following results hold.

(I) If \( d^2 < \frac{e^{\tau t_0} - 1}{\tau^2 t_0} \), any test is inconsistent.

(II) If \( d^2 > \frac{e^{\tau t_0} - 1}{\tau^2 t_0} \), the SLC test is consistent.

For graphs with weights following the exponential distribution, the region without consistent test is \( d^2 < 1 \) by Example 1. For the dichotomized network, the region without consistent test is \( d^2 < \frac{e^{\tau t_0} - 1}{\tau^2 t_0} \) by Corollary 2.5. Note that \( \min_{x>0} \frac{e^x - 1}{x^2} = 1.544 \) as shown in Figure 2. Hence, \( \frac{e^{\tau t_0} - 1}{\tau^2 t_0^2} \geq 1.544 > 1 \) for any \( \tau t_0 > 0 \) and dichotomizing weighted exponential network to binary network always enlarges the region where no consistent test exists. This reflects the loss of information. For fixed \( \tau, t_0 = \frac{1.544}{\tau} \) leads to the least loss of information. This interesting finding theoretically confirms the intuition that the dichotomy threshold should not be too small or too large.

### 2.3. A Stronger Result

The result (I) in Theorem 2.2 only guarantees the inconsistency of any statistical test, without providing further description of their performances. This can actually be made stronger under an additional ODE condition. The following theorem says that, under such a condition, all tests perform not better than random guess.

**Theorem 2.6.** Suppose \( \partial^\alpha \psi \), for \( \alpha \) with \( |\alpha| = 5 \), exist and are uniformly bounded for \( \theta \in \Theta \). Moreover, \( \psi \) satisfies the following differential equation

\[
\left( \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \tau_d \right)^\alpha \right)^2 + 3 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \tau_d \right)^\alpha = 0. \tag{8}
\]

Then any test is not better than random guessing if \( \tau_d D^2 \psi(\tau) \tau_d < 1 \).
Figure 2. Comparison of the testing radius under exponential distributions. Blue dashed: testing radius (≡ 1) under weighted network; Green solid: testing radius (y) under dichotomized weighted network. The latter is displayed as a function of threshold $t_0$ which achieves minimum at $t_0 = \frac{1.594}{\tau}$.

The following example provides distributions that satisfy condition (8).

**Example 3.** For single-parameter exponential family, (8) is reduced to

$$\left( \frac{d^2 \psi}{d\eta^2}(\tau) \right)^2 + \frac{1}{2} \frac{d^4 \psi}{d\eta^4}(\tau) = 0.$$ 

Consider the Gamma distribution with density given by

$$f(x; \lambda, \eta) = \frac{\eta^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\eta x}. $$

When $\lambda = 3$, the Gamma distribution belong to the single-parameter exponential family with $\psi(\eta) = -3 \log \eta$. Then direct calculation yields

$$\frac{d^2 \psi}{d\eta^2}(\tau) = \frac{3}{\tau^2}, \quad \frac{d^4 \psi}{d\eta^4}(\tau) = -\frac{18}{\tau^4}, $$

which satisfies (8). Similarly, the inverse Gamma distribution with scale parameter 2 satisfies (8).

3. Practical Tests

The WSLMC proposed in Section 2 is not practically feasible since it involves unknown quantities. Specifically, it is unclear which exponential distribution the weights follow, and the vector $T(x)$ is unknown as well. However, the WSLMC test still provides useful insights about how to construct a powerful test. Note that for exponential distribution, $T(x) = -x$; and for the normal distribution, $T(x) = (x, x^2)^T$. This hints us that we should consider higher moments of weights when proposing
novel test statistics. Based on this observation, we propose more practical tests that work for a broad class of distributions $P$ and $Q$.

To be convenient, we reformulate the hypotheses (1) as follows. Given a fixed positive integer $m$ and $\Theta \subset \mathbb{R}^m$, let $f(x; \mu)$ be a probability density with parameter $\mu = (\mu_1, \mu_2, \ldots, \mu_m)^T \in \Theta$, where $\mu_t$ $(t \in \{1, 2, \ldots, m\})$ is the $t$th moment. Define the $m$-parameter distribution family as

$$F_m = \{f(x; \mu) | \mu = (\mu_1, \mu_2, \ldots, \mu_m)^T \in \Theta \subset \mathbb{R}^m\}.$$  

Suppose $P$ and $Q$ belong to $F_m$ and has parameter $\mu_1$ and $\mu_2$ respectively. Let

$$\mu_1 = \mu + \frac{\mu_d}{\sqrt{n}}, \quad \mu_2 = \mu - \frac{\mu_d}{\sqrt{n}},$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_m)^T$, $d = (d_1, d_2, \ldots, d_m)^T$ and $\mu_d = (\mu_1 d_1, \mu_2 d_2, \ldots, \mu_m d_m)^T$. Then (1) can be rewritten as

$$H_0 : \|d\|_2 = 0 \quad \text{versus} \quad H_1 : \|d\|_2 \neq 0.$$  

Let $M(x) = (x, x^2, \ldots, x^m)^T$ and $k \geq 3$ be a positive integer. Define

$$\overline{A}_t = \frac{1}{n} \sum_{1 \leq i < j \leq n} A_{ij}, \quad t = 1, 2, \ldots, m,$$

and $\overline{M}(A) = (\overline{A}_1, \overline{A}_2, \ldots, \overline{A}_m)^T$. The sample covariance of $M(A_{12})$ is

$$S^2 = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} (M(A_{ij}) - \overline{M}(A))(M(A_{ij}) - \overline{M}(A))^T.$$  

We propose several tests based on the number of cycles. The first test is based on the signed-long-mixture-cycle (SLMC) test statistic defined as

$$\mathcal{T}_n = \frac{\sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^{k} (M(A_{it+1}) - \overline{M}(A))}{\sqrt{\left(\frac{(k-1)!}{2}\binom{n}{2}\right)(1^T S^2 1)^k}}.$$  

**Theorem 3.1.** Suppose $3 \leq k = O(\log \log \log n)$ and all the moments of $P$ and $Q$ exist. Then under $H_0$, $\mathcal{T}_n$ converges in distribution to the standard normal distribution as $n$ goes to infinity.

Based on Theorem 3.1, the SLMC test rejects $H_0$ if $|\mathcal{T}_n| > Z_{\frac{\alpha}{2}}$, where $Z_{\frac{\alpha}{2}}$ is the $100\frac{\alpha}{2}\%$ quantile of the standard normal distribution. Since the number of cycles with length $k$ can be expressed as a function of the trace of $A^k$ and the number of $k$ walks, the computation complexity of $\mathcal{T}_n$ is at most $O(n^3k)$. Hence, $\mathcal{T}_n$ is a practical test statistic. We point out that the condition all the moments of $P, Q$ exist can be relaxed to that the $4m$-th moments exist. The current proof of Theorem 3.1 employs the method of moment which requires all the moments are finite. An alternative proof is to use the Martingale central limit theorem, which only requires finite $4m$-th moment.

**Theorem 3.2.** Suppose $3 \leq k = O(\log \log \log n)$, $2m$-th moments of $P$ and $Q$ exist and $\max_{1 \leq t \leq m} d_t = o(\sqrt{n})$. Then under $H_1$, $\mathcal{T}_n = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{\sum_i 1}}\right)^k (1 + o_p(1))$.  

Based on Theorem 3.2, the power of the SLMC test approaches one as \( n \to \infty \) if \( \frac{1}{\sqrt{2k}} \left( \frac{1}{\sqrt{1 + \Sigma_k}} \right)^k \to \infty \). In this case, the SLMC test with larger \( k \) may achieve higher power. If \( \max_{1 \leq t \leq m} d_t \to \infty \), the power can tend to one for finite \( k \). Most importantly, even when \( \max_{1 \leq t \leq m} d_t \) is bounded, our test can still have asymptotic power one whenever \( k \to \infty \) and \( k = O(\log \log \log n) \). In this sense, the SLMC test is almost optimal. The condition \( \max_{1 \leq t \leq m} d_t = o(\sqrt{n}) \) is just to simplify the order of \( T_n \) under \( H_1 \). When \( \max_{1 \leq t \leq m} d_t = c\sqrt{n} \) for some constant \( c > 0 \), the order of \( T_n \) has a tedious expression but the power still converges to one. Theorem 3.2 only requires finite 2\( m \)-th moments of \( P \) and \( Q \), since we did not pursue the asymptotic distribution of \( T_n \). It suffices to get the order of \( T_n \) under \( H_1 \) for power analysis.

The SLMC test statistic employs all the \( m \) moments. Alternatively, we can also use a single moment to construct a test statistic. Let \( l \in \{1, 2, \ldots, m\} \) and

\[
S_l^2 = \frac{1}{n(n-2)} \sum_{i < j} (A_{ij}^l - \bar{A}^l)^2.
\]

Define the signed long-cycle (SLC) test statistic as

\[
T_{n,l} = \sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^{k} \left( A_{i_t i_{t+1}}^l - \bar{A}_t \right) \sqrt{\frac{(k-1)!}{2} \left( \frac{n}{k} \right) S_l^2}.
\]

The following results follow by a similar proof of Theorem 3.1 and Theorem 3.2.

**Corollary 3.3.** Suppose \( l \in \{1, 2, \ldots, m\} \) is a fixed integer, \( 3 \leq k = O(\log \log \log n) \) and all the moments of \( P \) and \( Q \) exist. Then under \( H_0 \), \( T_{n,l} \) converges in distribution to the standard normal distribution as \( n \) goes to infinity. Under \( H_1 \), if \( d_t = o(\sqrt{n}) \), then \( T_{n,l} = \frac{1}{\sqrt{2k}} \left( \frac{\mu_d}{\sigma_d} \right)^k (1 + o_p(1)) \). Here \( \sigma_l^2 = \text{Var}(A_{12}^l) \) under \( H_0 \).

The SLC test rejects \( H_0 \) if \( |T_{n,l}| > Z_{0.975} \), where \( Z_{0.975} \) is the 100\% quantile of the standard normal distribution. The power approaches one as \( n \to \infty \) if \( \frac{1}{\sqrt{2k}} \left( \frac{\mu_d}{\sigma_d} \right)^k \to \infty \). When \( \mu_d \) only has a single nonzero component, the SCL test may have higher power than the SLMC test.

### 4. Simulation and Application

#### 4.1. Simulation

In this subsection, we illustrate the performance of the proposed tests and compare them with the spectral test ([30]) in various simulations. The nominal type I error is set to be 0.05. The empirical type I errors and powers are calculated based on 500 repetitions.

For each \( l \in \{1, 2, \ldots, m\} \), the spectral test statistics are defined as

\[
\Lambda_{n,l} = n^{\frac{1}{2}} (\lambda_{l,\text{max}} - 2), \quad \Lambda_n = n^{\frac{1}{2}} (\lambda_n - 2),
\]

where \( \lambda_{l,\text{max}} \) is the maximum of the \( l \)-th eigenvalue.
Figure 3. Type I errors of the spectral tests and cycle-based tests in three models.

where
\[ \lambda_{l,\text{max}} = \lambda_{\text{max}} \left( A_l - \overline{A}_l \right) / \sqrt{nS_l} , \quad \lambda_n = \lambda_{\text{max}} \left( M(A) - \overline{M}(A) \right) / \sqrt{n^{\frac{1}{2}} S^2} . \]

Note that \( \Lambda_{n,1} \) is given in [30]. The rejection threshold is calculated based on the Tracy-Widom law with degree of freedom one.

We assess the performance of the cycle-based tests \( T_n, T_{n,1}, T_{n,2} \) and the spectral tests \( \Lambda_{n,1}, \Lambda_{n,2} \). The results for \( \Lambda_n \) are not reported due to large type I errors. The random labels \( \sigma_i \) (1 ≤ \( i \) ≤ \( n \)) are generated from Bernoulli distribution with success probability 0.5. Given \( \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n) \), let \( \lambda = (\lambda_1, \lambda_2) \) and \( \epsilon = (\epsilon_1, \epsilon_2) \), and

\[ \mu_1 = \lambda_1 + \epsilon_1 \sigma_i \sigma_j, \quad \mu_2 = \lambda_2 + \epsilon_2 \sigma_i \sigma_j, \quad (10) \]

In the first simulation, we generate the weights from the normal distribution with mean \( \mu_1 \) and second moment \( \mu_2 \) defined in (10). Figure 3 summarizes the empirical type I errors and Figure 4 visualizes the powers.

In the second simulation, the weights are assumed to follow the Gamma distribution with mean and second moment given in (10). The density of Gamma distribution is given by

\[ f(x; \lambda, \theta) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-\frac{x}{\theta}}, \]

where \( \lambda, \theta \) are functions of \( \mu_1, \mu_2 \):

\[ \lambda = \frac{\mu_2^2}{\mu_2 - \mu_1^2}, \quad \theta = \frac{\mu_2 - \mu_1^2}{\mu_1}. \]

Figure 3 presents the empirical type I errors and Figure 5 plots the powers.

In the last simulation, we generate weights from the mixture of two exponential distributions with density given by

\[ f(x; \lambda_1, \lambda_2) = 0.5 \lambda_1 e^{-\lambda_1 x} + 0.5 \lambda_2 e^{-\lambda_2 x}, \]
Figure 4. Powers of the spectral tests and cycle-based tests when weights follow Normal distribution with $\lambda = (0, 1)$.
Figure 5. Powers of the spectral tests and cycle-based tests when weights follow Gamma distribution with $\lambda = (4, 28)$.
where
\[ \lambda_1 = \frac{2}{2\mu_1 + \sqrt{2\mu_2 - 4\mu_1^2}}, \quad \lambda_2 = \frac{2}{2\mu_1 - \sqrt{2\mu_2 - 4\mu_1^2}}. \]

Note that this distribution does not belong to exponential family. The empirical type I errors are plotted in Figure 3 and Figure 6 visualizes the powers.

This simulation study has the following indications: (I) The cycle-based tests with \( k = 3 \) converge much faster than the spectral tests for non-Gaussian weights, since the cycle-based tests have empirical type I errors close to the nominal for smaller \( n \). (II) The empirical powers increase as the differences (\( \epsilon_1 \) or \( \epsilon_2 \)) of parameters get larger. (III) No single test can dominate the others in all cases. In summary, this simulation highlights the necessity to incorporate higher order moments of weights in test statistic and the proposed tests have significant advantage over spectral tests for smaller \( n \).

### 4.2. Animal Social Network Data

In this subsection, we apply the proposed tests to an animal social network “aves-wildbird-network-5” publicly available at [26]. In this network, each bird is a node and edges are weighted by interaction between two birds. This network has 145 nodes and 2512 weighted edges. We apply test statistics \( T_n, T_{n,1}, T_{n,2} \) with \( k = 3 \) to this network and the calculated test statistics are listed in Table 1. With type I error 0.05, all these three tests reject the null hypothesis, which indicates the network contains community structure.

| \( T_{n,1} \) | \( T_{n,2} \) | \( T_n \) |
|---|---|---|
| 225.7204 | 47.9405 | 174.9434 |

### 5. Proof of main result

#### 5.1. Proof of Theorem 2.2

For the proof of part (I) of Theorem 2.2, we will use the second moment methods. Specifically, we show that the second moment of the likelihood ratio under \( H_0 \) is bounded if \( 2\sum_{|\alpha|=2} \frac{\partial^\alpha \log \psi}{\alpha!} (\tau_d)^\alpha < 1 \). For the proof of part (II) of Theorem 2.2, we prove the WSLMC test has asymptotic power one if \( 2\sum_{|\alpha|=2} \frac{\partial^\alpha \log \psi}{\alpha!} (\tau_d)^\alpha > 1 \). For convenience, we will let \( m = 2 \). The proof for general \( m \) is exactly the same as \( m = 2 \).

**Proof of Theorem 2.2 (I):** The proof strategy is to show the second moment of the likelihood ratio under \( H_0 \) is bounded. Given random label vector \( \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n) \), the parameters of the distribution of \( A_{ij} \) can be concisely written as

\[
\theta_{t,ij}(\sigma) = \tau_t - \tau_1 \frac{d_t}{\sqrt{n}} \sigma_i \sigma_j, \quad t = 1, 2.
\]  

(11)
Figure 6. Powers of the spectral tests and cycle-based tests when weights follow mixture exponential distribution with $\lambda = (3.6, 36)$. 
Let $\theta_{ij}(\sigma) = (\theta_{1,ij}(\sigma), \theta_{2,ij}(\sigma))$ and $\tau = (\tau_1, \tau_2)$. Then the likelihood ratio $L_n$ is equal to

$$L_n = \frac{1}{2^n} \sum_{\sigma, \eta \in \{\pm\}^n} \left[ \prod_{1 \leq i < j \leq n} \frac{h(A_{ij}) \exp\{\theta_{1,ij}(\sigma)T_1(A_{ij}) + \theta_{2,ij}(\sigma)T_2(A_{ij}) - \psi(\theta_{ij}(\sigma))\}}{h(A_{ij}) \exp\{\tau_1 T_1(A_{ij}) + \tau_2 T_2(A_{ij})\} - \psi(\tau)} \right].$$

Let $\eta$ be an independent copy of $\sigma$. Then under $H_0$, the second moment of $L_n$ is equal to

$$\mathbb{E}[L_n^2] = \frac{1}{4^n} \sum_{\sigma, \eta \in \{\pm\}^n} \mathbb{E} \left[ \prod_{1 \leq i < j \leq n} \frac{h(A_{ij}) \exp\{\theta_{1,ij}(\sigma)T_1(A_{ij}) + \theta_{2,ij}(\sigma)T_2(A_{ij}) - \psi(\theta_{ij}(\sigma))\}}{h(A_{ij}) \exp\{\tau_1 T_1(A_{ij}) + \tau_2 T_2(A_{ij})\} - \psi(\tau)} \right] \times \prod_{1 \leq i < j \leq n} \int h(A_{ij}) \exp\{\theta_{1,ij}(\sigma)T_1(A_{ij}) + \theta_{2,ij}(\sigma)T_2(A_{ij}) - \psi(\theta_{ij}(\sigma))\} \frac{d\theta_{ij}(\sigma)}{\exp\{\tau_1 T_1(A_{ij}) + \tau_2 T_2(A_{ij})\} - \psi(\tau)} \right] \times \exp\{\theta_{1,ij}(\sigma)T_1(A_{ij}) + \theta_{2,ij}(\sigma)T_2(A_{ij}) - \psi(\theta_{ij}(\sigma))\} dA_{ij}$$

$$= \mathbb{E}_{\sigma, \eta} \left[ \prod_{1 \leq i < j \leq n} \int h(A_{ij}) \exp\{[\theta_{1,ij}(\sigma) + \theta_{1,ij}(\eta) - \tau_1]T_1(A_{ij}) + [\theta_{2,ij}(\sigma) + \theta_{2,ij}(\eta) - \tau_2]T_2(A_{ij}) - \psi(\theta_{ij}(\sigma)) - \psi(\theta_{ij}(\eta)) + \psi(\tau)\} dA_{ij} \right]$$

$$= \mathbb{E}_{\sigma, \eta} \left[ \prod_{1 \leq i < j \leq n} \exp\{\psi(\theta_{ij}(\sigma) + \theta_{ij}(\eta) - \tau) - \psi(\theta_{ij}(\sigma)) - \psi(\theta_{ij}(\eta)) + \psi(\tau)\} \right]$$

(12)

For $\sigma_i, \sigma_j = 1$ and $\eta_i, \eta_j = 1$, we have

$$\psi(\theta_{ij}(\sigma) + \theta_{ij}(\eta) - \tau) - \psi(\theta_{ij}(\sigma)) - \psi(\theta_{ij}(\eta)) + \psi(\tau)$$

$$= \psi\left(\tau_1 - \frac{2\tau_1 d_1}{\sqrt{n}}, \tau_2 - \frac{2\tau_2 d_2}{\sqrt{n}}\right) - \psi\left(\tau_1 - \frac{\tau_1 d_1}{\sqrt{n}}, \tau_2 - \frac{\tau_2 d_2}{\sqrt{n}}\right) + \psi(\tau_1, \tau_2).$$

(13)

By Taylor expansion, we have

$$\psi\left(\tau_1 - \frac{2\tau_1 d_1}{\sqrt{n}}, \tau_2 - \frac{2\tau_2 d_2}{\sqrt{n}}\right)$$

$$= \psi(\tau_1, \tau_2) - \frac{2\tau_1 d_1}{\sqrt{n}} D_\tau(\psi) + \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left(-\frac{2\tau_1 d_1}{\sqrt{n}}\right)^\alpha + \sum_{|\alpha|=3} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left(-\frac{2\tau_1 d_1}{\sqrt{n}}\right)^\alpha$$

$$+ \sum_{|\alpha|\geq4} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left(-\frac{2\tau_1 d_1}{\sqrt{n}}\right)^\alpha + O\left(\frac{1}{n^2 \sqrt{n}}\right).$$

(14)
\[
\psi \left( \tau_1 - \frac{\tau_1 d_1}{\sqrt{n}}, \tau_2 - \frac{\tau_2 d_2}{\sqrt{n}} \right) = \psi(\tau_1, \tau_2) - \frac{\tau_d}{\sqrt{n}} D\psi(\tau) + \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( -\frac{\tau_d}{\sqrt{n}} \right)^\alpha + \sum_{|\alpha|=3} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( -\frac{\tau_d}{\sqrt{n}} \right)^\alpha + \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( -\frac{\tau_d}{\sqrt{n}} \right)^\alpha + O \left( \frac{1}{n^{2\sqrt{n}}} \right). 
\]

(15)

Hence, by (13), (14),(15), we get

\[
\psi \left( \theta_{ij}(\sigma) + \theta_{ij}(\eta) - \tau \right) - \psi(\theta_{ij}(\sigma)) - \psi(\theta_{ij}(\eta)) + \psi(\tau) = 2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + 14 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha - 6 \sum_{|\alpha|=3} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + O \left( \frac{1}{n^{2\sqrt{n}}} \right). 
\]

(16)

For \( \sigma_i\sigma_j = 1 \) and \( \eta_i\eta_j = -1 \) or \( \sigma_i\sigma_j = -1 \) and \( \eta_i\eta_j = 1 \), one has

\[
\psi \left( \theta_{ij}(\sigma) + \theta_{ij}(\eta) - \tau \right) - \psi(\theta_{ij}(\sigma)) - \psi(\theta_{ij}(\eta)) + \psi(\tau) = \psi(\tau_1, \tau_2) - \psi \left( \tau_1 + \frac{\tau_1 d_1}{\sqrt{n}}, \tau_2 + \frac{\tau_2 d_2}{\sqrt{n}} \right) - \psi \left( \tau_1 - \frac{\tau_1 d_1}{\sqrt{n}}, \tau_2 - \frac{\tau_2 d_2}{\sqrt{n}} \right) + \psi(\tau_1, \tau_2) = -2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha - 2 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + O \left( \frac{1}{n^{2\sqrt{n}}} \right). 
\]

(17)

For \( \sigma_i\sigma_j = -1 \) and \( \eta_i\eta_j = -1 \), the following equations are true.

\[
\psi \left( \theta_{ij}(\sigma) + \theta_{ij}(\eta) - \tau \right) - \psi(\theta_{ij}(\sigma)) - \psi(\theta_{ij}(\eta)) + \psi(\tau) = \psi \left( \tau_1 + \frac{2\tau_1 d_1}{\sqrt{n}}, \tau_2 + \frac{2\tau_2 d_2}{\sqrt{n}} \right) - 2\psi \left( \tau_1 + \frac{\tau_1 d_1}{\sqrt{n}}, \tau_2 + \frac{\tau_2 d_2}{\sqrt{n}} \right) + \psi(\tau_1, \tau_2) = 2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + 14 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + 6 \sum_{|\alpha|=3} \frac{\partial^\alpha \psi}{\alpha!}(\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + O \left( \frac{1}{n^{2\sqrt{n}}} \right). 
\]

(18)

Let

\[
s_+ = \# \{(i, j) : i < j, \sigma_i\sigma_j\eta_i\eta_j = +1 \}, \quad s_- = \# \{(i, j) : i < j, \sigma_i\sigma_j\eta_i\eta_j = -1 \},
\]
and \( \rho = \frac{1}{n} \sum_{i=1}^{n} \sigma_i \eta_i \). Then \( s_+ = \frac{n^2}{4} (1 + \rho^2) - \frac{n}{2} \) and \( s_- = \frac{n^2}{4} (1 - \rho^2) \). By (12) and (16)-(18), we get

\[
\mathbb{E}[L_n^2] = (1 + o(1)) \mathbb{E}_{\sigma, \eta} \left[ \exp \left\{ - s_- \left( 2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + 2 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha \right) \right\} \right.
\]

\[
\times \exp \left\{ s_+ \left( 2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + 14 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha \right) \right\} \right. 
\]

\[
\times \exp \left\{ 6 \sum_{|\alpha|=3} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha \sum_{i<j} \left( \frac{1 - \sigma_i \sigma_j}{2} \right) \left( 1 - \eta_i \eta_j \right) - \frac{1}{2} \left( \frac{1 + \sigma_i \sigma_j}{2} \right) \left( 1 + \eta_i \eta_j \right) \right\} 
\]

\[
= (1 + o(1)) \mathbb{E}_{\sigma, \eta} \left[ Y_n \right] \exp \left\{ - \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) (\tau_d)^\alpha + 3 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\}, \tag{19} \]

where

\[
Y_n = \exp \left\{ n \rho^2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} \exp \left\{ 4 \rho^2 \sum_{|\alpha|=4} \frac{\partial^\alpha \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} \right. 
\]

\[
\times \exp \left\{ -3 \sum_{|\alpha|=3} \frac{\partial^\alpha \psi}{\alpha!} (\tau) (\tau_d)^\alpha \left[ \frac{1}{n \sqrt{n}} \left( \sum_{i=1}^{n} \sigma_i \right)^2 + \frac{1}{n \sqrt{n}} \left( \sum_{i=1}^{n} \eta_i \right)^2 \right] \right\}. \tag{20} \]

Next we find the limit of \( \mathbb{E}_{\sigma, \eta} [Y_n] \). Note that \( \exp \left\{ n \rho^2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} \) is uniformly integrable if \( 2 \sum_{|\alpha|=2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) (\tau_d)^\alpha = \tau_d^T D^2 \psi(\tau) \tau_d < 1 \). Besides, \( n \rho^2 \) converges in law to chi-square distribution with degree of freedom one. Hence

\[
\mathbb{E} \left( \exp \left\{ \frac{n \rho^2}{2} \tau_d^T D^2 \psi(\tau) \tau_d \right\} \right) = (1 + o(1)) \left( 1 - \tau_d^T D^2 \psi(\tau) \tau_d \right)^{-\frac{1}{2}}. \tag{21} \]

Let \( M = 32 \sum_{|\alpha|=3} \frac{|\partial^\alpha \psi|}{\alpha!} (\tau) (\tau_d)^\alpha \). Decompose \( Y_n \) as follows.

\[
Y_n = Y_{n,1} + Y_{n,2} + Y_{n,3} + Y_{n,4},
\]
where

\[
Y_{n,1} = Y_n I \left[ \frac{\sum_i \sigma_i^2}{n \sqrt{n}} \leq M, \frac{\sum_i \eta_i^2}{n \sqrt{n}} \leq M \right],
\]

\[
Y_{n,2} = Y_n I \left[ \frac{\sum_i \sigma_i^2}{n \sqrt{n}} > M, \frac{\sum_i \eta_i^2}{n \sqrt{n}} \leq M \right],
\]

\[
Y_{n,3} = Y_n I \left[ \frac{\sum_i \sigma_i^2}{n \sqrt{n}} > M, \frac{\sum_i \eta_i^2}{n \sqrt{n}} > M \right],
\]

\[
Y_{n,4} = Y_n I \left[ \frac{\sum_i \sigma_i^2}{n \sqrt{n}} \leq M, \frac{\sum_i \eta_i^2}{n \sqrt{n}} > M \right].
\]

Let \( p = 1 + \frac{1 - \tau_d^T D^2 \psi(\tau) \tau_d}{2 \tau_d^T D^2 \psi(\tau) \tau_d} \). Then \( p > 1 \) and \( p \tau_d^T D^2 \psi(\tau) \tau_d < 1 \). In this case, by (21),

\[
E_{\sigma, \eta} \left[ Y_{n1}^p \right] = O(1) \exp \left( \frac{n \rho^2}{2} p \tau_d^T D^2 \psi(\tau) \tau_d \right) = O(1),
\]

which implies \( Y_{n1} \) is uniformly integrable. Note that \( Y_{n1} \) converges in distribution to \( \exp \left\{ \frac{x^2}{2} \tau_d^T D^2 \psi(\tau) \tau_d \right\} \).

Hence

\[
E(Y_{n1}) = (1 + o(1)) \left( 1 - \tau_d^T D^2 \psi(\tau) \tau_d \right)^{-\frac{1}{2}}. \tag{22}
\]

Next we show \( E_{\sigma, \eta} \left[ Y_{nt} \right] = o(1) \) for \( t = 2, 3, 4 \). By Bernstein inequality, for \( n > M^2 \), we have

\[
P \left( \left( \sum_i \sigma_i^2 \right) \leq \frac{M}{\sqrt{n}} \right) \leq 2 \exp \left( - \frac{nM^2}{4} \right). \tag{23}
\]

Since \( \sigma_i \eta_i \) and \( \sigma_i \) are independent, then \( \rho \) and \( \sum_{i=1}^n \sigma_i \) are independent. By the fact that \( \frac{\sum_i \sigma_i^2}{n \sqrt{n}} \leq \sqrt{n}, \frac{\sum_i \eta_i^2}{n \sqrt{n}} \leq \sqrt{n}, \) (21) and (23), one has

\[
E_{\sigma, \eta} \left[ Y_{n2} \right] \leq O(1) \exp \left\{ 6 \sqrt{n} \sum_{|\alpha| = 3} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} E_{\sigma, \eta} \left[ \exp \left\{ n \rho^2 \sum_{|\alpha| = 2} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} I \left[ \frac{\sum_i \sigma_i^2}{n \sqrt{n}} > M \right] \right\] \leq O(1) \exp \left\{ 6 \sqrt{n} \sum_{|\alpha| = 3} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} E_{\sigma, \eta} \left[ \exp \left\{ n \rho^2 \sum_{|\alpha| = 2} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} \right]\]

\[
\leq 2O(1) \exp \left\{ 6 \sqrt{n} \sum_{|\alpha| = 3} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha - \frac{nM}{4} \right\} \exp \left\{ n \rho^2 \sum_{|\alpha| = 2} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} \right\] \leq o(1). \tag{24}
\]

Similarly, \( E_{\sigma, \eta} \left[ Y_{nt} \right] = o(1) \) for \( t = 3, 4 \). Then by (19) and (24),

\[
E(L_n^2) = (1 + o(1)) \exp \left\{ \sum_{|\alpha| = 2} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha + 3 \sum_{|\alpha| = 4} \frac{\partial^3 \psi}{\alpha!} (\tau) (\tau_d)^\alpha \right\} (1 - \tau_d^T D^2 \psi(\tau) \tau_d)^{-\frac{1}{2}}. \tag{25}
\]
Hence, $E(L_n^2) = O(1)$ under $H_0$, which implies any test is inconsistent.

(II). The proof strategy is to show the WSLMC test has asymptotic power one. To this end, we prove that $Z_n = O_P(1)$ under $H_0$ and $Z_n = (-1)^k \frac{\sum_{i=1}^k \left( \tau_d^T D^2 \psi(\tau) \tau_d \right)^{1/2}}{\sqrt{2k}} (1 + o(1)) + o_P(1)$ under $H_1$.

By the property of exponential family, the mean and covariance of $T(X)$ are equal to

$$E(T(X)) = D\psi, \quad Cov(T(X)) = D^2\psi.$$ 

Under $H_0$, $A_{ij} (1 \leq i < j \leq n)$ are independent. Hence, $E(Z_n) = 0$. Note that

$$\sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \mathbb{E} \left[ \prod_{t=1}^k \frac{\tau_d^T (T(A_{i_t}) - D\psi(\tau))}{\sqrt{n}} \right]^2$$

$$= \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} \prod_{t=1}^k \tau_d^T \mathbb{E} \left[ (T(A_{i_t}) - D\psi(\tau))(T(A_{i_t+1}) - D\psi(\tau))^T \right] \tau_d$$

$$= \binom{n}{k} \left[ \frac{\tau_d^T D^2 \psi(\tau) \tau_d}{n^k} \right]^k. \quad (26)$$

For $(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n$ and $(j_1, j_2, \ldots, j_k) \in \mathcal{I}_n$, if $(i_1, i_2, \ldots, i_k) \neq (j_1, j_2, \ldots, j_k)$, then

$$\mathbb{E} \left[ \prod_{t=1}^k \tau_d^T (T(A_{i_t}) - D\psi(\tau)) \prod_{t=1}^k \tau_d^T (T(A_{j_t}) - D\psi(\tau)) \right] = 0.$$

For $k = \log \log n$, it is easy to verify $\binom{n}{k} = \frac{n^k}{k!} (1 + o(1))$. Hence, by (26), we have

$$\mathbb{E} Z_n^2 = \frac{\binom{n-1}{k} \binom{n}{k}}{\frac{1}{2k} \left[ \frac{\tau_d^T D^2 \psi(\tau) \tau_d}{n^k} \right]^k} = 1 + o(1),$$

which implies $Z_n = O_P(1)$ under $H_0$.

Next, we show $Z_n = \frac{1}{\sqrt{2k}} \left[ \frac{\tau_d^T D^2 \psi(\tau) \tau_d}{n^k} \right]^{1/2} + o_P(1)$ under $H_1$. Given $\sigma$, the mean and covariance of $T(A_{ij})$ are equal to

$$E(T(A_{ij})) = D\psi(\theta_{ij}(\sigma)), \quad Cov(T(A_{ij})) = D^2\psi(\theta_{ij}(\sigma)),$$

where $\theta_{ij}(\sigma)$ is defined in (11). Let $\lambda_i \in \{0, 1\}, (1 \leq i \leq k)$ and $\lambda = (\lambda_1, \ldots, \lambda_k)$. Denote $|\lambda| = \lambda_1 + \lambda_2 + \cdots + \lambda_k$. Then

$$Z_n = \sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^k \frac{\tau_d^T (T(A_{i_t}) - D\psi(\theta_{i_t+1}(\sigma)) + D\psi(\theta_{i_t+1}(\sigma)) - D\psi(\tau)}{\sqrt{\frac{1}{2k} \binom{n}{k} \left[ \frac{\tau_d^T D^2 \psi(\tau) \tau_d}{n^k} \right]^k}}$$

$$= \sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^k \frac{\tau_d^T (D\psi(\theta_{i_t+1}(\sigma)) - D\psi(\tau))}{\sqrt{\frac{1}{2k} \binom{n}{k} \left[ \frac{\tau_d^T D^2 \psi(\tau) \tau_d}{n^k} \right]^k}}$$

$$+ \sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \sum_{|\lambda| > 0} \prod_{t=1}^k \frac{\tau_d^T (T(A_{i_t+1}) - D\psi(\theta_{i_t+1}(\sigma)))}{\sqrt{\frac{1}{2k} \binom{n}{k} \left[ \frac{\tau_d^T D^2 \psi(\tau) \tau_d}{n^k} \right]^k}}$$

$$= R_1 + R_2. \quad (27)$$
Next we show the first term $R_1$ in (27) is the leading term.

By Taylor expansion, we get
\[
\frac{\partial \psi}{\partial \theta_1}(\theta_{ij}(\sigma)) = \frac{\partial \psi}{\partial \theta_1}(\tau) + \frac{\partial^2 \psi}{\partial \theta_1^2}(\tau) \frac{\tau_1 d_1}{\sqrt{n}} \sigma_i \sigma_j + \frac{\partial^2 \psi}{\partial \theta_1 \partial \theta_2}(\tau) \frac{\tau_2 d_2}{\sqrt{n}} \sigma_i \sigma_j + O\left(\frac{1}{n}\right).
\]
\[
\frac{\partial \psi}{\partial \theta_2}(\theta_{ij}(\sigma)) = \frac{\partial \psi}{\partial \theta_2}(\tau) + \frac{\partial^2 \psi}{\partial \theta_1^2}(\tau) \frac{\tau_1 d_1}{\sqrt{n}} \sigma_i \sigma_j + \frac{\partial^2 \psi}{\partial \theta_2^2}(\tau) \frac{\tau_2 d_2}{\sqrt{n}} \sigma_i \sigma_j + O\left(\frac{1}{n}\right).
\]
Hence $\tau_d^T (D\psi(\theta_{ij}(\sigma)) - D\psi(\tau)) = -\frac{\tau_d^T D^2 \psi(\tau) \tau_d}{\sqrt{n}} \sigma_i \sigma_j + O\left(\frac{1}{n}\right)$ and
\[
\prod_{t=1}^k \tau_d^T (D\psi(\theta_{i_{t+1}}(\sigma)) - D\psi(\tau)) = \prod_{t=1}^k \left(-\frac{\tau_d^T D^2 \psi(\tau) \tau_d}{\sqrt{n}} + O\left(\frac{1}{n}\right)\right)^{\|\sigma_i \sigma_{i+1}\|^2} = \frac{\tau_d^T D^2 \psi(\tau) \tau_d}{\sqrt{n}} (1 + o(1)).
\]
(28)

Then (28) implies $R_1 = (1 + o(1)) (-1)^k \sqrt{\frac{n^k}{2k}} \tau_d^T D^2 \psi(\tau) \tau_d^k$.

Next, we prove $R_2 = o_p(1)$. Since $\tau_d$ is a vector of constants, then for a large constant $C > 0$,
\[
\mathbb{E} \left[ \tau_d^T \left( T(A_{i_{i+1}}(\sigma)) - D\psi(\theta_{i_{i+1}}(\sigma)) \right) \right]^2 \leq C,
\]
\[
\mathbb{E} \left| \tau_d^T \left( D\psi(\theta_{ij}(\sigma)) - D\psi(\tau) \right) \right| \leq \frac{C}{\sqrt{n}}.
\]
Recall that $A_{ij} (1 \leq i < j \leq n)$ are independent conditional on $\sigma$. Then fixing a $\lambda$ with $|\lambda| = l > 0$, we have
\[
\mathbb{E} \left[ \sum_{i_1 < i_2 < \ldots < i_k} \prod_{t=1}^k \tau_d^T \left( T(A_{i_{i+1}}(\sigma)) - D\psi(\theta_{i_{i+1}}(\sigma)) \right) \right]^{\lambda_t} \left[ \tau_d^T \left( D\psi(\theta_{i_{i+1}}(\sigma)) - D\psi(\tau) \right) \right]^{1-\lambda_t} \leq \frac{1}{2k n^k \tau_d^T D^2 \psi(\tau) \tau_d} \sum_{i_1 < i_2 < \ldots < i_k} \prod_{t=1}^k \mathbb{E} \left[ \tau_d^T \left( T(A_{i_{i+1}}(\sigma)) - D\psi(\theta_{i_{i+1}}(\sigma)) \right) \right]^{2\lambda_t}
\]
\[
\times \left[ \tau_d^T \left( D\psi(\theta_{i_{i+1}}(\sigma)) - D\psi(\tau) \right) \right]^{1-\lambda_t} \left[ \tau_d^T \left( D\psi(\theta_{j_{j+1}}(\sigma)) - D\psi(\tau) \right) \right]^{1-\lambda_t}
\]
\[
\leq \frac{2k}{n^k \tau_d^T D^2 \psi(\tau) \tau_d} \frac{C^{2k-2l}}{\sqrt{n}^{2k-2l}} \left( \frac{2k}{2kC^{2k-l}} \right)^{2k-l} \leq \frac{n^{s-l} \tau_d^T D^2 \psi(\tau) \tau_d^k}{\tau_d^T D^2 \psi(\tau) \tau_d^k} \leq \frac{2k}{n^{s-l} \tau_d^T D^2 \psi(\tau) \tau_d^k} = o(1).
\]
Here, $s$ is the number of distinct nodes that any $l$ edges on the cycle $i_1, i_2, \ldots, i_k$ have and hence $s \geq l + 1$. Note that there are $2^k - 1$ possible choices of $\lambda$ such that $|\lambda| = l > 0$. Since $k! \leq k^k$ and $k = \log \log \log n$, then
\[
\mathbb{E}(R_2^2) \leq \frac{(k-1)!^2}{4} \frac{2kC^{2k}}{n \tau_d^T D^2 \psi(\tau) \tau_d} = o(1).
\]
Then the proof is complete. \qed
5.2. Proof of Theorem 2.4

The proof strategy is similar to that of Theorem 2.2. The likelihood ratio $L_n$ is equal to

$$L_n = \frac{1}{2^n} \sum_{\sigma \in \{\pm\}^n} \left[ \prod_{1 \leq i < j \leq n} \frac{p_{ij}(\sigma) \tilde{A}_{ij} (1 - p_{ij}(\sigma))^{1 - \tilde{A}_{ij}}}{p_0^{\tilde{A}_{ij}} (1 - p_0)^{1 - \tilde{A}_{ij}}} \right].$$

Let $\eta$ be an independent copy of $\sigma$. Then the second moment under $H_0$ is

$$\mathbb{E}(L_n^2) = \mathbb{E}_{\sigma, \eta} \left[ \prod_{1 \leq i < j \leq n} \left( \frac{p_{ij}(\sigma)p_{ij}(\eta) + (1 - p_{ij}(\sigma))(1 - p_{ij}(\eta))}{p_0} \right) \right]$$

$$= \mathbb{E}_{\sigma, \eta} \left[ \prod_{1 \leq i < j \leq n} \frac{p_0 - p_0[p_{ij}(\sigma) + p_{ij}(\eta)] + p_{ij}(\sigma)p_{ij}(\eta)}{p_0(1 - p_0)} \right]. \quad (29)$$

By Taylor expansion, we have

$$\psi \left( \tau - \frac{\tau_d}{\sqrt{n}} \sigma_i \sigma_j \right) = \psi(\tau) - \frac{\tau_d}{\sqrt{n}} D\psi(\tau) \sigma_i \sigma_j + \sum_{|\alpha| = 2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha + O \left( \frac{1}{n^{3/2}} \right) \sigma_i \sigma_j.$$

Hence,

$$p_{ij}(\sigma) - p_0$$

$$= \int_{t_0}^\infty h(x) e^{\theta(\sigma_i \sigma_j)^T T(x) - \psi(\theta(\sigma_i \sigma_j))} - h(x)e^{\tau^T T(x) - \psi(\tau)} dx$$

$$= \int_{t_0}^\infty h(x) e^{\tau^T T(x) - \psi(\tau)} \left[ \exp \left\{ \frac{\tau_d}{\sqrt{n}} (T(x) + D\psi(\tau)) \sigma_i \sigma_j \right. \right.$$

$$\left. - \sum_{|\alpha| = 2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha \right] - 1 \right] dx$$

$$= \int_{t_0}^\infty h(x) e^{\tau^T T(x) - \psi(\tau)} \left[ \left( \frac{\tau_d}{\sqrt{n}} (T(x) + D\psi(\tau)) \sigma_i \sigma_j \right. \right.\right.$$

$$\left. - \sum_{|\alpha| = 2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha \right] + O \left( \frac{1}{n^{3/2}} \right) \sigma_i \sigma_j \right] dx + O \left( \frac{1}{n^2} \right)$$

$$= \sigma_i \sigma_j \frac{\tau_d a(t_0)}{\sqrt{n}} + \frac{b(t_0)}{n} + O \left( \frac{1}{n^{3/2}} \right) \sigma_i \sigma_j,$$

where

$$b(t_0) = \int_{t_0}^\infty h(x) e^{\tau^T T(x) - \psi(\tau)} \left( \frac{\tau_d}{\sqrt{n}} (T(x) + D\psi(\tau)) \right) \left[ \left( \frac{\tau_d}{\sqrt{n}} (T(x) + D\psi(\tau)) \right) - \sum_{|\alpha| = 2} \frac{\partial^\alpha \psi}{\alpha!} (\tau) \left( \frac{\tau_d}{\sqrt{n}} \right)^\alpha \right] dx.$$
Then we get

$$p_{ij}(\sigma) = p_0 + \frac{\sigma_i \sigma_j \tau_d^T a(t_0)}{\sqrt{n}} + \frac{b(t_0)}{n} + O\left(\frac{1}{n^{3/2}}\right) \sigma_j + O\left(\frac{1}{n^2}\right),$$

and then

$$p_0 - p_0[p_{ij}(\sigma) + p_{ij}(\eta)] + p_{ij}(\sigma)p_{ij}(\eta)$$

$$= p_0 - p_0\left(2p_0 + \frac{2b(t_0)}{n} + \frac{(\sigma_i \sigma_j + \eta_i \eta_j) \tau_d^T a(t_0)}{\sqrt{n}} + O\left(\frac{1}{n^{3/2}}\right) \sigma_j + \eta_j \eta_j + O\left(\frac{1}{n^2}\right)\right)$$

$$+ p_0 \frac{2p_0 b(t_0)}{n} + \frac{(\sigma_i \sigma_j + \eta_i \eta_j) p_0 \tau_d^T a(t_0)}{\sqrt{n}} + \frac{(\tau_d^T a(t_0))^2}{n} \sigma_i \sigma_j \eta_j \eta_j + O\left(\frac{1}{n^{3/2}}\right) \sigma_j + \eta_j \eta_j$$

$$= p_0(1 - p_0) + \frac{(\tau_d^T a(t_0))^2}{n} \sigma_i \sigma_j \eta_j \eta_j + O\left(\frac{1}{n^{3/2}}\right) \sigma_j + \eta_j \eta_j).$$

Plugging (30) into (29) yields

$$\mathbb{E}(L^2_n) = \mathbb{E}_{\sigma, \eta}\left[\prod_{1 \leq i < j \leq n} p_0 - p_0[p_{ij}(\sigma) + p_{ij}(\eta)] + p_{ij}(\sigma)p_{ij}(\eta)\right]/p_0(1 - p_0)$$

$$= \mathbb{E}_{\sigma, \eta}\left[\prod_{1 \leq i < j \leq n} \left(1 + \frac{(\tau_d^T a(t_0))^2}{n p_0(1 - p_0)} \sigma_j \eta_j + O\left(\frac{1}{n^{3/2}}\right) \sigma_j + \eta_j \eta_j\right)\right]$$

$$= O(1)\mathbb{E}_{\sigma, \eta}\left[\exp\left(\frac{(\tau_d^T a(t_0))^2}{2p_0(1 - p_0)} \sum_{i<j}^n \sigma_i \sigma_j \eta_i \eta_j\right) + O\left(1\right) \sum_{i<j}^n \sigma_i \sigma_j + \eta_j \eta_j\right]\right]$$

$$= O(1)\mathbb{E}_{\sigma, \eta}\left[\exp\left(\frac{(\tau_d^T a(t_0))^2}{2p_0(1 - p_0)} \sum_{i}^n \sigma_i \eta_i\right)^2 - n + O\left(1\right) \sum_{i}^n \sigma_i \sigma_i + \sum_{i}^n \eta_i \eta_i - 2n\right]\right]$$

If \(\frac{(\tau_d^T a(t_0))^2}{p_0(1 - p_0)} < 1\), then \(\exp\left(\frac{(\tau_d^T a(t_0))^2}{2p_0(1 - p_0)} \sum_{i}^n \sigma_i \eta_i\right)^2\) is uniformly integrable and

$$\mathbb{E}_{\sigma, \eta}\left[\exp\exp\left(\frac{(\tau_d^T a(t_0))^2}{2p_0(1 - p_0)} \sum_{i}^n \sigma_i \eta_i\right)^2\right] = (1 + o(1)) \left(1 - \frac{(\tau_d^T a(t_0))^2}{p_0(1 - p_0)}\right)^{-\frac{1}{2}}.$$

By a similar truncation technique in the proof of Theorem 2.2, we conclude \(\mathbb{E}(L^2_n) = O(1)\) if \(\frac{(\tau_d^T a(t_0))^2}{p_0(1 - p_0)} < 1\).

(II). By a similar proof of (II) of Theorem 2.2, \(Z_n = O_P(1)\) under \(H_0\) and \(Z_n = \frac{(-1)^k}{\sqrt{2k}} \left(\frac{(\tau_d^T a(t_0))^2}{p_0(1 - p_0)}\right)^{\frac{1}{2}} \left(1 + o(1)\right) + O_P(1)\) under \(H_1\). Then the proof is complete.

\[\square\]

5.3. Proof of Theorem 2.6

Firstly we recall the contiguity theorem ([9]) and several useful lemmas.
Proposition 5.1. Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures and $X_{ni}$ be random variables on the same sample space. Then $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if the following conditions hold.

i). $\mathbb{P}_n \ll \mathbb{Q}_n$ and $\mathbb{Q}_n \ll \mathbb{P}_n$.

ii). For any fixed $m \geq 3$, $X_{ni}(3 \leq i \leq m)$ jointly converges in distribution to $Z_i$ with $Z_i \sim N(0, 2i)$ under $\mathbb{P}_n$ and $Y_i$ with $Y_i \sim N(t^2, 2i)$ ($|t| < 1$) under $\mathbb{Q}_n$ respectively.

iii). $Z_i$ and $Y_i$ $(3 \leq i \leq m)$ are independent.

iv). \[
\mathbb{E}_{\mathbb{P}_n}\left[\frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right]^2 = (1 + o(1)) \exp\left(-\frac{t^2}{2} - \frac{t^4}{4}\right) \frac{1}{\sqrt{1-t}}.
\]

The following two lemmas are well-known.

Lemma 5.2. Let $Y_{n,1}, \ldots, Y_{n,m}$ be $m$ random variables. Then $Y_{n,1}, \ldots, Y_{n,m}$ jointly converges in distribution to $Z_1, \ldots, Z_m$ if the following conditions hold.

i). For any fixed $k$ and $\lambda_1 + \cdots + \lambda_m = k$ with integers $\lambda_t \geq 0$, $\mathbb{E}[Y_{n,1}^{\lambda_1}, \ldots, Y_{n,m}^{\lambda_m}] = (1+o(1))\mathbb{E}[Z_1^{\lambda_1}, \ldots, Z_m^{\lambda_m}]$.

ii). \[
\sum_{l=1}^{\infty} \left(\lim_{n \to \infty} \mathbb{E}(Y_{n,l}^{2l})\right)^{-\frac{3}{4}} < 1.
\]

Lemma 5.3. Let $Y_1, \ldots, Y_m$ follow a $m$-variate distribution $F$ with mean 0 and covariance $\Sigma$. Then $F$ is Gaussian distribution if and only if for even $l$,

\[
\mathbb{E}[X_1X_2 \ldots X_l] = \sum_{\eta} \prod_{i=1}^{\frac{l}{2}} \mathbb{E}[X_{\eta(i,1)}X_{\eta(i,2)}]
\]

and $\mathbb{E}[X_1X_2 \ldots X_l] = 0$ for odd $l$. Here $X_i \in \{Y_1, Y_2, \ldots, Y_m\}$, $\eta$ is a partition of $\{1, 2, \ldots, l\}$ into $\frac{l}{2}$ equal-size subsets and $\eta(i,t)$ is the $t$th element of ith subset.

Given integer $k \geq 3$, let $J_{n,k} = \{(i_1, i_2, \ldots, i_k) | i_1, \ldots, i_k \text{ distinct}\}$ and define

\[
U_{n,k} = \frac{\sum_{(i_1, i_2, \ldots, i_k) \in J_{n,k}} \prod_{t=1}^{k} \tau_d (T(A_{ii_{t+1}}) - D\psi(\tau))}{\sqrt{n^k \left(\tau_d D^2 \psi(\tau) \tau_d\right)^k}}.
\]

Proposition 5.4. The following results hold.

(a). For fixed integers $3 \leq k_1 < k_2 < \cdots < k_m$, $\frac{U_{n,k_t}^2}{\sqrt{2k_t}} (1 \leq t \leq m)$ converges jointly in distribution to the standard $m$-variate normal distribution under $H_0$.

(b). For fixed integers $3 \leq k_1 < k_2 < \cdots < k_m$, $\frac{U_{n,k_t}^2}{\sqrt{2k_t}} D^2 \psi(\tau) \tau_d (1 \leq t \leq m)$ converges jointly in distribution to the standard $m$-variate normal distribution under $H_1$.

Proof of Proposition 5.4. By the proof of Theorem 2.2, it is easy to get that $Var(U_{n,k}) = 2k(1+o(1))$ for each fixed integer $k \geq 3$ under $H_0$ or $H_1$. By Lemma 5.2 and Lemma 5.3, to prove (a) or (b), it suffices to prove that

\[
\mathbb{E}[U_{n,k_t}^2] = o(1), \quad k_1 \neq k_2,
\]

(31)
and
\[ \mathbb{E}[\mathcal{V}_{n,1} \ldots \mathcal{V}_{n,l}] = \begin{cases} \sum_{\eta} \prod_{t=1}^{l} \mathbb{E}[Z_{\eta(i_1)}Z_{\eta(i_2)}], & \text{even } l, \\ 0, & \text{odd } l, \end{cases} \tag{32} \]

where \( \mathcal{V}_{n,i} \in \{ \mathcal{U}_{n,k_1}, \ldots, \mathcal{U}_{n,k_m} \} \).

(a) Under \( H_0 \), \( A_{ij} \) \((1 \leq i < j \leq n)\) are independent. Hence it is easy to verify that (31) holds. Next we prove (32). Let \( V_{n,i} = (i^{(1)}_t, \ldots, i^{(\lambda_i)}_t) \in \mathcal{J}_{n,\lambda_i} \) and \( V_t = \prod_{t=1}^{\lambda_i} T_d \left( T(A_{i_t(i_t+1)}) - D_\psi(\tau) \right) \). Then
\[ \mathbb{E}[\mathcal{V}_{n,1} \ldots \mathcal{V}_{n,l}] = \frac{\sum_{i_1, \ldots, i_l} \mathbb{E}(V_{i_1} \ldots V_{i_l})}{(n \tau_d^T D^2 \psi(\tau) \tau_d)^{\sum_{i=1}^{\lambda_i} \lambda_i/2}}. \tag{33} \]

Each edge in \( V_t \) must be traversed at least twice, otherwise \( \mathbb{E}(V_{i_1} \ldots V_{i_l}) = 0 \). Hence any node pair \((i^{(1)}_t, i^{(\lambda_i)}_t) \in \mathcal{J}_{n,\lambda_i} \) must be equal to at least one other pair \((i^{(1)}_{t_1}, i^{(\lambda_i)}_{t_1}) \in \mathcal{J}_{n,\lambda_i} \) with \( t_1 \neq t_2 \). Then \( i_t \) \((1 \leq t \leq l)\) are partitioned into \( s \) disjoint groups with each group containing at least 2 elements. Clearly \( s \leq l/2 \). If \( s < l/2 \), then \( i_t \) \((1 \leq t \leq l)\) has at most \( \sum_{i=1}^{\lambda_i} - 1 \) distinct nodes. Hence by (33), we have
\[ \mathbb{E}[\mathcal{V}_{n,1} \ldots \mathcal{V}_{n,l}] = O \left( \frac{n^{\sum_{i=1}^{\lambda_i} - 1}}{n^{\sum_{i=1}^{\lambda_i}/2}} \right) = o(1). \]

Note that for odd \( l \), \( s < l/2 \) always holds. When \( l \) is even and \( s = l/2 \), it is easy to check that (32) holds.

(b) By the proof of Part (II) of Theorem 2.2 and part (a) above, the proof is straightforward. Hence we omit it.

**Proof of Theorem 2.6:** By (25), if (8) holds, then iv) of Proposition 5.1 holds with \( t = \tau_d D^2 \psi(\tau) \tau_d \). The proof is straightforward based on Proposition 5.1 and Proposition 5.4.

\[ \square \]

### 5.4. Proof of Theorem 3.1 and Theorem 3.2

**Proof of Theorem 3.1:** Suppose \( H_0 \) holds. Let \( \Sigma = \text{Cov}(M(A_{12})) \). Then for \( k = \log \log \log n \), \((1^T S^2)^k\) converges to \((1^T \Sigma 1)^k\) in probability. Let
\[ \mathcal{Y}_n = \frac{\sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{I}_n} \prod_{t=1}^{k} 1^T (M(A_{i_t i_{t+1}}) - M(A))}{\sqrt{\frac{(k-1)!}{2}} (1^T \Sigma 1)^k}. \]
We only need to prove $\mathcal{Y}_n$ converges to the standard normal distribution. Let $\lambda_t \in \{0, 1\}$, $\lambda = (\lambda_1, \ldots, \lambda_k)$ and $|\lambda| = \lambda_1 + \lambda_2 + \cdots + \lambda_k$. Note that
\[
\sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} 1^T (M(A_{i_t i_{t+1}}) - \overline{M}(A))
= \sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} 1^T (M(A_{i_t i_{t+1}}) - \mu + \mu - \overline{M}(A))
= \sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} 1^T (M(A_{i_t i_{t+1}}) - \mu) + \sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} 1^T (M(A_{i_t i_{t+1}}) - \mu) \lambda_t [1^T (\mu - \overline{M}(A))]^{1-\lambda_t}
= R_1 + R_2.
\] (34)

Next we show $\mathbb{E}(R_1^2) = \binom{n}{k} (1^T \Sigma 1)^k$ and $R_2 = o_p \left( \sqrt{\binom{n}{k} (1^T \Sigma 1)^k} \right)$.

Under $H_0$, $A_{ij} (1 \leq i < j \leq n)$ are independent and $\mathbb{E}(M(A_{ij})) = \mu$. Hence
\[
\mathbb{E}\left[ \sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} 1^T (M(A_{i_t i_{t+1}}) - \mu) \right]^2
= \sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} \mathbb{E}[1^T (M(A_{i_t i_{t+1}}) - \mu)]^2
= \binom{n}{k} (1^T \Sigma 1)^k.
\] (35)

For a given $\lambda$ with $|\lambda| < k$, we have
\[
\sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} [1^T (M(A_{i_t i_{t+1}}) - \mu)]^{\lambda_t} [1^T (\mu - \overline{M}(A))]^{1-\lambda_t}
= \prod_{t=1}^{k} [1^T (\mu - \overline{M}(A))]^{1-\lambda_t} \sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} [1^T (M(A_{i_t i_{t+1}}) - \mu)]^{\lambda_t}.
\]

Note that
\[
\mathbb{E}\left[ \sum_{i_1 < i_2 < \cdots < i_{k+1}} \prod_{t=1}^{k} [1^T (M(A_{i_t i_{t+1}}) - \mu)]^{\lambda_t} \right]^2
= \prod_{t=1}^{k} \mathbb{E}[1^T (M(A_{i_t i_{t+1}}) - \mu)]^{2\lambda_t}
\leq n^{2k-s} (1^T \Sigma 1)^{|\lambda|},
\]
where $k > s \geq |\lambda| + 1$. Since $\mu - \overline{M}(A) = O_p \left( \frac{1}{n} \right)$, then
\[
R_2 = o_p \left( 2^k n^{|\lambda| - \frac{s}{2}} (1^T \Sigma 1)^{\frac{|\lambda|}{2}} \right) = o_p \left( \sqrt{\binom{n}{k} (1^T \Sigma 1)^k} \right).
\] (36)

By (34), (35) and (36), we conclude that
\[
\mathcal{Y}_n = \mathcal{X}_n + o_p(1),
\]
where
\[ \chi_n = \sum_{(i_1,i_2,\ldots,i_k) \in T_n} \prod_{t=1}^k 1^T \left( M(A_{i_t,i_{t+1}}) - \mu \right) \frac{1}{\sqrt{\frac{(k-1)!}{2}} \binom{n}{k} (1^T \Sigma 1)^k}. \]

Next, we use the method of moment to prove \( \chi_n \) converges in distribution to the standard normal distribution. To this end, we will show \( \mathbb{E}(\chi_n^r) = o(1) \) for odd \( r \) and \( \mathbb{E}(\chi_n^r) = (r - 1)!! + o(1) \) for even \( r \).

Clearly, \( \mathbb{E}(\chi_n) = 0 \). The second moment of \( \chi_n \) is
\[ \mathbb{E}(\chi_n^2) = \sum_{(i_1,i_2,\ldots,i_k) \in T_n} \prod_{t=1}^k 1^T \left( M(A_{i_t,i_{t+1}}) - \mu \right) \frac{1}{\sqrt{\frac{(k-1)!}{2}} \binom{n}{k} (1^T \Sigma 1)^k} = \sum_{(i_1,i_2,\ldots,i_k) \in T_n} \prod_{t=1}^k 1^T \left( M(A_{i_t,i_{t+1}}) - \mu \right)^2 \frac{1}{\sqrt{\frac{(k-1)!}{2}} \binom{n}{k} (1^T \Sigma 1)^k} = 1. \]

Fix a positive integer \( r \geq 3 \). For convenience, let \( i_k^{(t)} = (i_1^{(t)},i_2^{(t)},\ldots,i_k^{(t)}) \) denote a circular permutation of \( k \) distinct nodes for each \( t \in \{1,2,\ldots,r\} \). Then
\[ \mathbb{E}(\chi_n^{2r}) = \sum_{i_k^{(t)},i_k^{(v)} \in T_n} \mathbb{E}\left[ \prod_{t=1}^r \prod_{v=1}^k 1^T \left( M(A_{i_t^{(v)},i_{t+1}^{(v)}}) - \mu \right) \frac{1}{\sqrt{\frac{(k-1)!}{2}} \binom{n}{k} (1^T \Sigma 1)^k} \right]. \]

If there are two indexes \( i_k^{(v)},i_k^{(v)} \) such that \( (i_k^{(v)},i_k^{(v)}) \) is different from any other pairs, then
\[ \mathbb{E}\left[ \prod_{t=1}^r \prod_{v=1}^k 1^T \left( M(A_{i_t^{(v)},i_{t+1}^{(v)}}) - \mu \right) \right] = 0. \]

Hence, any \( i_k^{(v)} \in T_n \) has to be equal to at least one \( i_k^{(v_2)} \in T_n \) for \( v_2 \neq v_1 \). Then there exist \( s(s \leq \frac{r}{2}) \) integers \( \lambda_i, (\lambda_i \geq 2, i = 1,2,\ldots,s) \) such that \( \lambda_1 + \lambda_2 + \cdots + \lambda_s = r \) and
\[ \mathbb{E}(\chi_n^r) = C_r \sum_{i_k^{(v)} \in T_n} \mathbb{E}\left[ \prod_{t=1}^r \prod_{v=1}^k 1^T \left( M(A_{i_t^{(v)},i_{t+1}^{(v)}}) - \mu \right)^{\lambda_v} \right], \]
where \( C_r \) is a constant dependent on \( r \). Note that all the moments of \( A_{12} \) are finite. By repeatedly using Cauchy-Schwarz inequality, we have
\[ \mathbb{E}\left[ \prod_{t=1}^r \prod_{v=1}^k 1^T \left( M(A_{i_t^{(v)},i_{t+1}^{(v)}}) - \mu \right)^{\lambda_v} \right] \leq C^{kr}, \]
for a large constant \( C \). If there exists \( \lambda_v \geq 3 \), that is, \( s < \frac{r}{2} \), then
\[ \mathbb{E}(\chi_n^r) \leq C_r \frac{n^{kr} C^{kr}}{\frac{(k-1)!}{2} \binom{n}{k} (1^T \Sigma 1)^k} = \frac{C_r}{(1^T \Sigma 1)^{kr} n^{k\frac{r}{2} - s}} = o(1), \]
noting that \( k = \log \log n \). If \( r \) is an odd number, \( s < \frac{r}{2} \) holds and hence \( \mathbb{E}(\mathcal{A}_n^r) = o(1) \).

Next we assume \( r \) is even and \( s = \frac{r}{2} \). Note that there are \((r-1)!!\) ways to partition \( r \) distinct numbers into \( \frac{r}{2} \) pairs. Then \( C_r = (r-1)!! \) and

\[
\mathbb{E}(\mathcal{A}_n^r) = (r-1)!! \sum_{i_k^{(1)}, \ldots, i_k^{(s)} \in \mathcal{I}_n} \mathbb{E} \left[ \prod_{v=1}^{s} \prod_{t=1}^{k} \frac{1}{T} \left( M(A_{i_t (v)} - \mu) \right)^2 \right] \cdot \left[ \frac{(k-1)!/n^k \left( 1^T \Sigma 1 \right)^k}{(n-k^2/2)^{k^2/2}} \right]^2.
\]

If there are two \( i_k^{(v_1)} \) and \( i_k^{(v_2)} \) \((v_1 \neq v_2)\) have at least a common vertex, then \( \mathbb{E}(\mathcal{A}_n^r) = o(1) \). Hence

\[
\mathbb{E}(\mathcal{A}_n^r) = (r-1)!! \left( (r-1)!! \right) \sum_{i_k^{(1)}, \ldots, i_k^{(s)} \in \mathcal{I}_n} \prod_{v=1}^{s} \prod_{t=1}^{k} \frac{1}{T} \left( M(A_{i_t (v)} - \mu) \right)^2 \cdot \left[ \frac{(k-1)!/n^k \left( 1^T \Sigma 1 \right)^k}{(n-k^2/2)^{k^2/2}} \right]^2 = (r-1)!!.
\]

Then the proof is complete.

\( \square \)

**Proof of Theorem 3.2:** We find the order of each term in (34). Given \( \sigma \), the mean of \( M(A_{ij}) \) is equal to \( \mu_{ij} = \mu + \frac{\mu_d}{\sqrt{n}} \sigma_i \sigma_j \). Hence,

\[
\sum_{i_1 < i_2 < \cdots < i_k} \prod_{t=1}^{k} \frac{1}{T} (M(A_{i_t i_{t+1}}) - \mu)
\]

\[
= \sum_{i_1 < i_2 < \cdots < i_k} \prod_{t=1}^{k} \frac{1}{T} (M(A_{i_t i_{t+1}}) - \mu_{i_t i_{t+1}} + \mu_{i_t i_{t+1}} - \mu)
\]

\[
= \sum_{i_1 < i_2 < \cdots < i_k} \prod_{t=1}^{k} \frac{1}{T} (\mu_{i_t i_{t+1}} - \mu) + \sum_{i_1 < i_2 < \cdots < i_k} \prod_{t=1}^{k} \frac{1}{T} (M(A_{i_t i_{t+1}}) - \mu_{i_t i_{t+1}})
\]

\[
+ \sum_{i_1 < i_2 < \cdots < i_k} \sum_{0 < |\lambda| < k} \prod_{t=1}^{k} \frac{1}{T} (M(A_{i_t i_{t+1}}) - \mu_{i_t i_{t+1}})^\lambda_t [1T(M(A_{i_t i_{t+1}}) - \mu)]^{1-\lambda_t}
\]

\[
= \left( \begin{array}{c} n \\ k \end{array} \right) \frac{1}{(\sqrt{n})^k} \mu_d^k + (a) + (b).
\]

(37)

Note that the conditional second moment \( \Sigma_{ij} \) of \( M(A_{ij}) \) exists and is a function of \( \mu_{ij} = \mu + \frac{\mu_d}{\sqrt{n}} \sigma_i \sigma_j \). By Taylor expansion, it follows that \( \frac{1}{T} \Sigma_{ij} 1 = \frac{1}{T} \Sigma 1 + O \left( \frac{\max_t |d_t|}{\sqrt{n}} \right) \). Hence,

\[
\mathbb{E}((a)^2) = \sum_{i_1 < i_2 < \cdots < i_k} \mathbb{E} \left[ \prod_{t=1}^{k} \mathbb{E} \left[ \left( \frac{1}{T} (M(A_{i_t i_{t+1}}) - \mu_{i_t i_{t+1}}) \right)^2 \big| \sigma \right] \right]
\]

\[
= \sum_{i_1 < i_2 < \cdots < i_k} \mathbb{E} \left[ \prod_{t=1}^{k} \left( \frac{1}{T} \Sigma 1 + O \left( \frac{\max_t |d_t|}{\sqrt{n}} \right) \right) \right]
\]

\[
= \left( \begin{array}{c} n \\ k \end{array} \right) \left( \frac{1}{T} \Sigma 1 \right)^k (1 + o(1)).
\]
For $s$ with $1 + |\lambda| \leq s \leq 2|\lambda|$, the order of (b) is bounded by

$$2^k \left( \frac{1^T \mu_d}{\sqrt{n}} \right)^{k-|\lambda|} \sqrt{n^{2k-s} \left( 1^T \Sigma 1 + O \left( \frac{\max_i |d_i|}{\sqrt{n}} \right) \right)^{|\lambda|}} = o \left( \left( \frac{n}{k} \right)^{1^T \mu_d} \right)^k. \quad (39)$$

Hence by (37), (38) and (39), we have

$$\sum_{(i_1,i_2,\ldots,i_k) \in \mathcal{I}_n} \prod_{t=1}^k \left[ 1^T (M(A_{i_{i_t+1}}) - \mu) \right] = \frac{1}{\sqrt{2^k k! (1^T \Sigma 1)^k}} \left( 2 + o_p(1) \right).$$

Given $\sigma$, we have

$$1^T (\mu - \mu(A)) = \frac{1}{2} \sum_{1 \leq i < j \leq n} 1^T (\mu - \mu_{ij} + \mu_{ij} - M(A_{ij}))$$

$$= \frac{1}{2} \sum_{1 \leq i < j \leq n} -\frac{1^T \mu_d}{\sqrt{n}} \sigma_{ij} + \frac{1}{2} \sum_{1 \leq i < j \leq n} 1^T (\mu_{ij} - M(A_{ij}))$$

$$= -\frac{1^T \mu_d}{\sqrt{n}} \left( \sum_{i=1}^n \sigma_i \right)^2 - n + O_p \left( \sqrt{\sum_{1 \leq i < j \leq n} 1^T \Sigma 1} \right)$$

$$= O_p \left( \frac{1^T \mu_d}{\sqrt{n}} + \frac{1}{n} \right) = O_p \left( \frac{1}{n} \right), \quad (40)$$

and

$$1^T (M(A_{i_{i_t+1}}) - \mu) = 1^T (M(A_{i_{i_t+1}}) - \mu_{ij}) + \frac{1^T \mu_d}{\sqrt{n}} \sigma_{ij} = 1^T (M(A_{i_{i_t+1}}) - \mu_{ij}) + o_p(1).$$

For a given $\lambda$,

$$E \left[ \sum_{i_1 < i_2 < \ldots < i_k} \prod_{t=1}^k \left[ 1^T (M(A_{i_{i_t+1}}) - \mu_{ij}) \right]^{\lambda_t} \right]^2$$

$$= \sum_{i_1 < i_2 < \ldots < i_k} \prod_{t=1}^k \left[ 1^T (M(A_{i_{i_t+1}}) - \mu_{ij}) \right]^{2\lambda_t} = O \left( n^{2k-s} \right),$$

where $s \geq |\lambda| + 1$. Then

$$\sum_{i_1 < i_2 < \ldots < i_k} \prod_{t=1}^k \left[ 1^T (M(A_{i_{i_t+1}}) - \mu) \right]^{\lambda_t} \left[ 1^T (\mu - \mu(A)) \right]^{1-\lambda_t}$$

$$= (1 + o_p(1)) \prod_{t=1}^k \left[ 1^T (\mu - \mu(A)) \right]^{1-\lambda_t} \sum_{i_1 < i_2 < \ldots < i_k} \sum_{|\lambda| < k} \prod_{t=1}^k \left[ 1^T (M(A_{i_{i_t+1}}) - \mu_{ij}) \right]^{\lambda_t}$$

$$= (1 + o_p(1)) O_p \left( \frac{1}{n^{k-|\lambda|}} \right) O_p \left( 2^k \sqrt{n^{2k-s}} \right)$$

$$= (1 + o_p(1)) O_p \left( 2^k C_k n^{\frac{|\lambda| - s}{2}} \right)$$

$$= (1 + o_p(1)) O_p \left( 2^k C_k n^{\frac{k_s}{2}} \right)$$

$$= o_p \left( n^{\frac{s}{2}} \left( 1^T \mu_d \right)^k \right). \quad (41)$$
Then the proof is complete by (34), (37), (38), (39) and (41).

References

[1] Aicher, C. (2014). The Weighted Stochastic Block Model. *Applied Mathematics Graduate Theses & Dissertations*, 50.

[2] Aicher, C., Jacob, A. and Clauset, A. (2015). Learning Latent Block Structure in Weighted Networks. *Journal of Complex Networks*, 3, 221-248.

[3] Abbe, E. (2017). Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18, 1-86.

[4] Arias-Castro, E. and N. Verzelen. 2014. Community detection in dense random networks. *Annals of Statistics*, 42, 3: 940-969.

[5] Ahn, K., Lee, K. and Suh, C. (2018). Hypergraph spectral clustering in the weighted stochastic block model. *IEEE Journal of Selected Topics in Signal Processing*, 12, 959-974.

[6] Abbe, E. and Sandon, C. (2017). Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7), 1334-1406.

[7] Abbe, E., Banderi, A. and Hall, G. (2016). Exact Recovery in the Stochastic Block Model. *IEEE transactions on information theory*, 62(1) 471-487.

[8] Amini, A., Chen, A. and Bickel, P. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Annals of Statistics*, 41(4), 2097-2122.

[9] Banerjee, D. (2018). Contiguity and non-reconstruction results for planted partition models: the dense case. *Electronic Journal of Probability*, 23, 1-28.

[10] Banerjee, D. and Ma, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. https://arxiv.org/pdf/1705.05305.pdf.

[11] Bickel, P. J. and Sarkar, P. (2016). Hypothesis testing for automated community detection in networks. *Journal of Royal Statistical Society, Series B*, 78, 253-273.

[12] Bubeck,S., Ding, J., Eldan,R. and Rácz, M. (2016) Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3),503-532.

[13] Chen, J. and Yuan,B. (2006). Detecting functional modules in the yeast protein-protein interaction network. *Bioinformatics*, 22(18), 2283-2290.

[14] Colizza, V., Pastor-Satorras, R. and Vespignani, A.(2007). Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Phys*, 3, 276–282.

[15] Canu, S. and Smola, A. J.(2005). Kernel methods and the exponential family. Neurocomputing, 69: 714–720.

[16] Fortunato,S. (2010). Community detection in graphs. *Physics Reports, 486 (3-5)*, 75-174.

[17] Gao, C. and Lafferty, J. (2017a). Testing for global network structure using small subgraph statistics. https://arxiv.org/pdf/1710.00862.pdf
[18] Jin, J., Ke, Z. and Luo S. (2018). Network global testing by counting graphlets. *International conference on machine learning*, 2333-2341.

[19] Lei, J. (2016). A goodness-of-fit test for stochastic block models. *Annals of Statistics, 44*, 401-424.

[20] Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics, 43(1)*, 215-237.

[21] Lu, Z., Wen, Y. and Cao, G. (2015). Community detection in weighted networks: algorithms and applications. *IEEE Transactions on Parallel and Distributed Systems*, 26(11): 2916-2926.

[22] Mossel, E., Neeman, J. and Sly, A. (2015). Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields, 162*, 431-461.

[23] Montanari, A. and Sen, S. (2016). Semidefinite programs on sparse random graphs and their application to community detection. *STOC ’16 Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 814-827.

[24] Newman, M. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64*, 016-131.

[25] C. Nicolini, C. Bordier, and A. Bifone. Community detection in weighted brain connectivity networks beyond the resolution limit. *Phys. Rev. E*, 146:28–39, 2017.

[26] R. Rossi and N. Ahmed. (2015). The Network Data Repository with Interactive Graph Analytics and Visualization, https://networkrepository.com.

[27] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage, 52*:1059–1069, 2010.

[28] Sriperumbudur, B., Fukumizu, K., Gretton, A., Hyvarinen, A. adn Kumar, R. (2017). Density estimation in infinite dimensional exponential families, *Journal of Machine Learning Research*, 18,1-59.

[29] Thomas, A. C. and Blitzstein, J. K. (2011). Valued ties tell fewer lies: Why not to dichotomize network edges with thresholds. arXiv:1101.0788

[30] Tokuda, T. (2018). Statistical test for detecting community structure in real-valued edgeweighted graphs. *PLoS ONE* 13(3): e0194079.

[31] Verzelen, N., and E. Arias-Castro. 2015. Community detection in sparse random networks. *Ann. Appl. Probab.* 25.6:3465–3510.

[32] Xu, M., Jog, V. and Loh, P. (2020). Optimal rates for community estimation in the weighted stochastic block model. *Annals of Statistics, 48*(1):183-204.

[33] Yuan, M., Feng, Y. and Shang, Z. (2022). A likelihood-ratio type test for stochastic block models with bounded degrees. *Journal of Statistical Planning and Inference*, 219:98-119.

[34] Yuan, M., Yang, F. and Shang, Z. (2022). Hypothesis testing in sparse weighted stochastic block model. *Statistical Papers*, accepted.

[35] Yuan, M., Liu, R., Feng, Y. and Shang, Z. (2022). Testing community structures for hyper-
graphs. *Annals of Statistics*, 50(1): 147-169.

[36] Yuan, M. and Nan, Y. (2020). Test dense subgraphs in sparse uniform hypergraph. *Communications in Statistics - Theory and Methods*, to appear.

[37] Yuan, M. and Shang, Z. (2021). Sharp Detection Boundaries on Testing Dense Subhypergraph. *Bernoulli*, to appear.

[38] Zhao, Y., Levina, E. and Zhu., J. (2011). Community extraction for social networks. *Proc. Natn. Acad. Sci. USA*, 108, 7321-7326.

[39] Zhao, Y., Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Annals of Statistics*, 40, 2266-2292.