Testing Community Structures for Hypergraphs

Mingao Yuan†, Ruiqi Liu §, Yang Feng* and Zuofeng Shang &†

†Department of Statistics
North Dakota State University
mingao.yuan@ndsu.edu

§Department of Mathematical Sciences
IUPUI
liuruiq@iu.edu

*Department of Statistics
Columbia University
yang.feng@columbia.edu

&Department of Mathematical Sciences
IUPUI
shangzf@iu.edu

Abstract: Many complex networks in real world can be formulated as hypergraphs where community detection has been widely used. However, the fundamental question of whether communities exist or not in an observed hypergraph still remains unresolved. The aim of the present paper is to tackle this important problem. Specifically, we study when a hypergraph with community structure can be successfully distinguished from its Erdős-Renyi counterpart, and propose concrete test statistics based on hypergraph cycles when the models are distinguishable. Our contributions are summarized as follows. For uniform hypergraphs, we show that successful testing is always impossible when average degree tends to zero, might be possible when average degree is bounded, and is possible when average degree is growing. We obtain asymptotic distributions of the proposed test statistics and analyze their power. Our results for growing degree case are further extended to nonuniform hypergraphs in which a new test involving both edge and hyperedge information is proposed. The novel aspect of our new test is that it is provably more powerful than the classic test involving only edge information. Simulation and real data analysis support our theoretical findings. The proofs rely on Janson’s contiguity theory ([32]) and a high-moments driven asymptotic normality result by Gao and Wormald ([28]).

AMS 2000 subject classifications: Primary 62G10; secondary 05C80.

Keywords and phrases: hypergraph, stochastic block model, hypothesis testing, contiguity, l-cycle.

1. Introduction

Community detection is a fundamental problem in network data analysis. For instance, in social networks ([18, 30, 53]), protein to protein interactions ([14]), image segmentation ([49]), among others, many algorithms have been developed for identifying community structure. Theoretical studies on community detection have mostly been focusing on ordinary graph setting in which each possible edge contains exactly two vertices (see [7, 3, 46, 53, 54, 27, 4]). One common assumption made in these references is the existence of communities. Recently, a number of researchers have been devoted to testing this assumption, e.g., [12, 34, 41, 10, 6, 25, 26, 51].

Real-world networks are usually more complex than ordinary graphs. Unlike ordinary graphs where data structure is typically unique, e.g., edges only contain two vertices, hypergraphs demonstrate a number of possibly overlapping data structures. For instance, in coauthorship data ([17, 44, 47, 42]), the number of coauthors varies so that one cannot consider edges consisting of two coauthors only. Instead, a new type of “edge,” called as hyperedge, must be considered which allows the connectivity of arbitrarily many coauthors. The complex structures of hypergraphs create new challenges in both theoretical and methodological study. As far as we know, existing hypergraph literature mostly focus on community detection in algorithmic aspects ([48, 13, 7, 46, 3, 22, 33, 35]). Only recently Ghoshdastidar and Dukkipati [22, 23] provided a statistical study in which a spectral algorithm based on adjacency tensor was proposed for identifying community structure and asymptotic results were developed. Nonetheless, the important problem of testing the existence of community structure in an observed hypergraph still remains untreated.

*Co-corresponding author. Supported by NSF CAREER Grant DMS-1554804.
†Corresponding author. Supported by NSF DMS-1764280 and NSF DMS-1821157.
In this paper, we aim to tackle the problem of testing community structure for hypergraphs. We first consider the relatively simpler but widely useful uniform hypergraphs in which each hyperedge consists of equal number of vertices. For instance, the (user, resource, annotation) structure in folksonomy may be represented as a uniform hypergraph where each hyperedge consists of three vertices ([29]); the (user, remote host, log-in time, log-out time) structure in the login-data can be modeled as a uniform hypergraph where each hyperedge contains four vertices ([24]); the point-set matching problem is usually formulated as identifying a strongly connected component in a uniform hypergraph ([13]). We provide various theoretical or methodological studies ranging from dense uniform hypergraphs to sparse ones and investigate the possibility of a successful test in each scenario. Our testing results in dense case are then extended to the more general nonuniform hypergraph setting, in which a new test statistic involving both edge and hyperedge is proposed. One important finding is that our new test is more powerful than the classic one involving edge information only, which is an advantage of using hyperedge information to boost the testing performance.

1.1. Review of Hypergraph Model And Relevant Literature.

In this section, we review some basic notion in hypergraphs and recent progress in literature. Let us first review the notion of uniform hypergraph. An \( m \)-uniform hypergraph \( H_m = (V, E) \) consists of a vertex set \( V \) and a hyperedge set \( E \), where each hyperedge in \( E \) is a subset of \( V \) consisting of exactly \( m \) vertices. Two hyperedges are the same if they are equal as vertex sets. An \( l \)-cycle in \( H_m \) is a cyclic ordering \( \{v_1, v_2, \ldots, v_l\} \) of the vertex set with hyperedges like \( \{v_i, v_{i+1}, \ldots, v_{i+m-1}\} \) and any two adjacent hyperedges have exactly \( l \) common vertices. An \( l \)-cycle is loose if \( l = 1 \) and tight if \( l = m - 1 \). To better illustrate the notion, consider a 3-uniform hypergraph \( H_3 = (V, E) \), where \( V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\} \), \( E = \{(v_i, v_j, v_l) | 1 \leq i < j < l \leq 7\} \). Then \( \{(v_1, v_2, v_3, v_4, v_5, v_6), (v_1, v_2, v_3), (v_3, v_4, v_5), (v_5, v_6, v_1)\} \) is a loose cycle and \( \{(v_1, v_2, v_3, v_4), (v_1, v_2, v_3), (v_2, v_3, v_4), (v_3, v_4, v_1), (v_4, v_1, v_2)\} \) is a tight cycle (see Figure 1).

**Fig 1:** Left: a loose cycle of three edges \( E_1, E_2, E_3 \). Right: a tight cycle of four edges \( E_1, E_2, E_3, E_4 \). Both cycles are subgraphs of the 3-uniform hypergraph \( H_3(V, E) \).

Next, let us review uniform hypergraphs with a planted partitioning structure, also known as stochastic block model (SBM). For positive integers \( n, m, k \) with \( m, k \geq 2 \) and positive constants \( a, b, \alpha \) with \( a > b \), let \( \mathcal{H}_m^k(n, \frac{a}{m}, \frac{b}{m}) \) denote an \( m \)-uniform hypergraph of \( n \) vertices and \( k \) balanced communities, in which \( \frac{a}{m} \) (\( \frac{b}{m} \)) represents the hyperedge probability within (between) communities. More explicitly, any vertex \( i \in [n] \equiv \{1, 2, \ldots, n\} \) is assigned, independently and uniformly at random, a label \( \sigma_i \in [k] \), and then each possible hyperedge \( (i_1, i_2, \ldots, i_m) \) is included with probability \( \frac{a}{m} \) if \( \sigma_{i_1} = \sigma_{i_2} = \cdots = \sigma_{i_m} \) and with probability \( \frac{b}{m} \) otherwise. In particular, \( \mathcal{H}_2^k(n, \frac{a}{m}, \frac{b}{m}) \) (with \( m = k = 2 \)) becomes the ordinary bisection stochastic block models considered by [39, 51]. Let \( A \in \{0, 1\}^{\frac{n \times n \times \cdots \times n}{m}} \) denote the symmetric adjacency
tensor of order $m$ associated with $\mathcal{H}^k_m(n, \frac{a}{n^m}, \frac{b}{n^m})$. By symmetry we mean that $A_{i_1i_2...i_m} = A_{\psi(i_1)\psi(i_2)\psi(i_m)}$ for any permutation $\psi$ of $(i_1, i_2, ..., i_m)$. For convenience, assume $A_{i_1i_2...i_m} = 0$ if $i_s = i_t$ for some distinct $s, t \in \{1, 2, ..., m\}$, i.e., the hypergraph has no self-loops. Conditional on $\sigma_1, ..., \sigma_n$, the $A_{i_1i_2...i_m}$’s, with $i_1, ..., i_m$ pairwise distinct, are assumed to be independent following the distribution below:

$$\mathbb{P}(A_{i_1i_2...i_m} = 1|\sigma) = p_{i_1i_2...i_m}(\sigma), \mathbb{P}(A_{i_1i_2...i_m} = 0|\sigma) = q_{i_1i_2...i_m}(\sigma),$$

(1)

where $\sigma = (\sigma_1, ..., \sigma_n)$,

$$p_{i_1i_2...i_m}(\sigma) = \begin{cases} \frac{a}{n^m}, & \sigma_1 = \cdots = \sigma_m, \\ \frac{b}{n^m}, & \text{otherwise} \end{cases}, q_{i_1i_2...i_m}(\sigma) = 1 - p_{i_1i_2...i_m}(\sigma).$$

In other words, each possible hyperedge $(i_1, ..., i_m)$ is included with probability $\frac{a}{n^m}$ if the vertices $i_1, ..., i_m$ belong to the same community, and with probability $\frac{b}{n^m}$ if they belong to different communities. Let $\mathcal{H}_m(n, \frac{a+(k^{m-1}-1)b}{k^{m-1}n^m})$ denote the $m$-uniform hypergraph without community structure, i.e., an Erdős-Rényi model in which each possible hyperedge is included with common probability $\frac{a+(k^{m-1}-1)b}{k^{m-1}n^m}$. We consider such a special choice of hyperedge probability in order to make the model have the same average degree as $\mathcal{H}_m(n, \frac{a}{n^m}, \frac{b}{n^m})$. In particular, $\mathcal{H}_2(n, \frac{a+(k-1)b}{kn})$ with $m = 2$ becomes the traditional Erdős-Rényi model that has been well studied in ordinary graph literature; see [8, 9, 20, 16, 50]. Nonuniform hypergraphs can be simply viewed as a superposition of uniform ones; see Section 2.4.

Given an observed adjacency tensor $A$, does $A$ represent a hypergraph that exhibits community structure?

In the present setting, this problem can be formulated as testing the following hypothesis:

$$H_0 : A \sim \mathcal{H}_m(n, \frac{a+(k^{m-1}-1)b}{k^{m-1}n^m}) \text{ vs. } H_1 : A \sim \mathcal{H}^k_m(n, \frac{a}{n^m}, \frac{b}{n^m}),$$

(2)

where $a, b$ are predetermined positive constants. When $m = k = 2$, problem (2) has been well studied in the literature. Specifically, for extremely sparse scenario ($\alpha > 1$), [39] show that $H_0$ and $H_1$ are always indistinguishable; for bounded degree case ($\alpha = 1$), the two models are distinguishable if and only if the signal-to-noise ratio (SNR) is greater than 1 ([39, 40, 51]); for dense scenario ($\alpha < 1$), $H_0$ and $H_1$ are always distinguishable and a number of algorithms have been developed (see [34, 25, 26, 10, 2, 12]). When $m = 2$ and $k \geq 3$, the above statements remain true for $\alpha > 1$ and $\alpha < 1$; but for $\alpha = 1$, SNR $> 1$ is only a sufficient condition for successfully distinguishing $H_0$ from $H_1$ while a necessary condition remains an open problem (see [2, 11, 52]). Abbe ([1]) provides a comprehensive review about the recent development in this field. As far as we know, there is a lack of literature dealing with the testing problem (2) for general $m$. The literature on hypergraph analysis merely focus on community detection; see [5, 22, 23, 48, 13, 21, 33, 35, 38].

1.2. Our Contributions.

The aim of this paper is to provide a study on hypergraph testing under a spectrum of hyperedge density scenarios. Our results consist of four major parts. Section 2.1 deals with the extremely sparse scenario ($\alpha > m-1$), in which we show that $H_0$ and $H_1$ are always indistinguishable in the sense of contiguity. Section 2.2 deals with bounded degree case ($\alpha = m-1$), in which we show that $H_1$ and $H_0$ are distinguishable if the SNR of uniform hypergraph is greater than one, but indistinguishable if the SNR is below certain threshold. We also construct a powerful test statistic in the former case based on counting the “long loose cycles.” Section 2.3 deals with dense scenario. Specifically, for $\alpha \in (m-\frac{k+3}{k}, m-1)$ with $1 \leq l \leq \frac{m}{2}$, we propose a test based on counting the $l$-cycles, and show that the power of the proposed test approaches one as the number of vertices goes to infinity. In Section 2.4, we extend some of the previous results to nonuniform hypergraph testing. We propose a new test involving both edge and hyperedge information and show that it is generally more powerful than the classic test using edge information only (see Remark 2.2). The results of the present paper can be viewed as nontrivial extensions of the ordinary graph testing results such as [39, 40, 25]. Section 3 provides empirical studies to support our theory.
2. Main Results

2.1. A Contiguity Theory for Sparse Scenario ($\alpha > m - 1$).

In this section, we consider the testing problem (2) with $\alpha > m - 1$, i.e., the hyperedge density of the hypergraph is extremely low. We will show that no test can successfully distinguish $H_0$ from $H_1$ in such situation. The proof proceeds by showing that the probability measures associated with $H_0$ and $H_1$ are contiguous (see Theorem 2.1). We remark that contiguity has also been used to prove indistinguishability for ordinary graphs (see [39, 40]).

Let $P_n$ and $Q_n$ be sequences of probability measures on a common probability space $(\Omega_n, \mathcal{F}_n)$. We say that $P_n$ and $Q_n$ are mutually contiguous if for every sequence of measurable sets $A_n \subset \Omega_n$, $P_n(A_n) \to 0$ if and only if $Q_n(A_n) \to 0$ as $n \to \infty$. They are said to be orthogonal if there exists a sequence of measurable sets $A_n$ such that $P_n(A_n) \to 0$ and $Q_n(A_n) \to 1$ as $n \to \infty$. According to [39], two probability models are indistinguishable if their associated probability measures are mutually contiguous, and two probability models are distinguishable if their associated probability measures are orthogonal. The following Theorem 2.1 shows that $H_0$ and $H_1$ are indistinguishable for any $\alpha > m - 1$.

**Theorem 2.1.** If $\alpha > m - 1$, then for any fixed $a > b > 0$, the probability measures associated with $H_0$ and $H_1$ are mutually contiguous.

The proof of Theorem 2.1 proceeds by showing that the ratio of the likelihood function of $H_1$ over $H_0$ converges in distribution to 1 under $H_0$, which implies the contiguity of $H_1$ and $H_0$ ([32]). Theorem 2.1 says that when $\alpha > m - 1$, the hypergraphs in $H_0$ and $H_1$ are indistinguishable, hence, no test can successfully separate the two hypotheses. Indeed, when $\alpha > m - 1$, the average degree of both hypergraph models converges to zero. To see this, the average degree is

$$\left( \frac{n}{m-1} \right) \frac{a + (k^{m-1} - 1)b}{k^{m-1}n^\alpha},$$

which goes to zero as $n \to \infty$ if $\alpha > m - 1$. Therefore, the signals in both models are not strong enough to support a successful test. It is easy to see that the average degree becomes bounded when $\alpha = m - 1$ which will be investigated in next section.

2.2. Bounded Degree Case ($\alpha = m - 1$).

In this section, we consider $\alpha = m - 1$ which leads to bounded average degrees for the models in $H_0$ and $H_1$; see (3). Define

$$\kappa = \frac{(a-b)^2}{k^{m-1}(m-2)!(a + (k^{m-1} - 1)b)}.$$  \hspace{1cm} (4)

When $m = k = 2$, it is easy to check that $\kappa = \frac{(a-b)^2}{2(a+b)}$ which becomes the classic SNR of ordinary stochastic block models considered by [39]. Hence, it is reasonable to view $\kappa$ defined in (4) as a generalization of the classic SNR to the hypergraph model $H_{\alpha}(n, \frac{m}{n}, \frac{b}{m^\alpha})$. Like the classic SNR, the value of $\kappa$ characterizes the separability between communities. Intuitively, when $\kappa$ is large which means that the communities are very different, the testing problem (2) becomes simpler. The following result says that when $\kappa > 1$ successful testing becomes possible.

**Theorem 2.2.** Suppose that $a > b > 0$ are fixed constants, $m, k \geq 2$ and $\alpha = m - 1$. If $\kappa > 1$, then the probability measures associated with $H_0$ and $H_1$ are orthogonal.

We prove Theorem 2.2 by constructing a sequence of events dependent on the number of long loose cycles and showing that the probabilities of the events converge to 1 (or 0) under $H_0$ (or $H_1$), based on the high
propose the following test statistic \( \kappa \) for \( \kappa > 1 \). Abbe and Sandon [2] obtained relevant results for ordinary graph setting, i.e., \( m = 2 \) and \( k \geq 2 \) in our case; see Corollary 2.8 therein which states that community detection in polynomial time becomes possible if SNR > 1. Whereas Theorem 2.2 holds for arbitrary \( m, k \geq 2 \). Hence our result can be viewed as an extension of [2] to hypergraph setting.

Let us now propose a test statistic based on “long loose cycles” that can successfully distinguish \( H_0 \) and \( H_1 \) when \( \kappa > 1 \). Let \( k_n \) be a sequence possibly diverging along with \( n \). Let \( X_{k_n} \) be the number of loose cycles each consisting of exactly \( k_n \) edges. Define

\[
\mu_{n0} = \frac{X_{k_n}}{2k_n^2}, \quad \mu_{n1} = \mu_{n0} + \frac{k - 1}{2k_n} \left[ \frac{a - b}{k^{m-1}(m-2)!} \right]^{k_n},
\]

where \( \lambda_m = \frac{a+(k^m-1)b}{k^{m-2}(m-2)!} \) for any \( m \geq 2 \). Let \( P_{H_1} \) denote the probability measure induced by \( A \) under \( H_1 \).

We have the following theorem about the asymptotic property of \( X_{k_n} \).

**Theorem 2.3.** Suppose \( \kappa > 1 \) and \( 1 \ll k_n \leq \delta_0 \log \lambda_m \log \gamma n \), where \( \gamma > 1 \) and \( 0 < \delta_0 < 2 \) are constants. Then, under \( H_1 \) for \( l = 0, 1 \),

\[
\frac{X_{k_n} - \mu_{n1}}{\sqrt{\mu_{n1}}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.
\]

Furthermore, for any constant \( C > 0 \),

\[
P_{H_1} \left( \left| \frac{X_{k_n} - \mu_{n0}}{\sqrt{\mu_{n0}}} \right| > C \right) \to 1 \text{ as } n \to \infty.
\]

The proof is based on the asymptotic normality theory developed by [28]. According to Theorem 2.3, we propose the following test statistic

\[
T_{k_n} = \frac{X_{k_n} - \mu_{n0}}{\sqrt{\mu_{n0}}},
\]

By Theorem 2.2, \( T_{k_n} \xrightarrow{d} N(0,1) \) under \( H_0 \). Hence, we construct the following testing rule at significance level \( \alpha \in (0,1) \):

\[
\text{reject } H_0 \text{ if and only if } |T_{k_n}| > z_{\alpha/2},
\]

where \( z_{\alpha/2} \) is the \((1-\alpha/2)\)-percentile of \( N(0,1) \). It follows by Theorem 2.2 that \( P_{H_1}(|T_{k_n}| > z_{\alpha/2}) \to 1 \), i.e., the power of \( T_{k_n} \) approaches one.

Theorem 2.3 requires \( k_n \) to grow slower than an iterative logarithmic order. This is due to the use of [28] which requires \( k_n \lambda_m^k = o(\log n) \). In practice, we suggest to choose \( k_n = \lceil \delta_0 \log \lambda_m \log \gamma n \rceil \) with \( \gamma \) close to 1 and \( \delta_0 \) close to 2. Such \( \gamma \) and \( \delta_0 \) will make \( k_n \) suitably large so that the test statistic \( T_{k_n} \) becomes valid. For instance, Table 1 demonstrates the values of \( k_n \) along with \( n \) with \( \delta_0 = 1.99, \gamma = 1.01, \lambda_m = 10 \). We can see that, for a moderate range of \( n \), the values of \( k_n \) are sufficiently large to make the test valid.

| Desirable \( k_n \) | 3 | 4 | 5 | 6 |
|-------------------|---|---|---|---|
| Minimal \( n \)    | 2 | 3 | 25 | 29786 |

**Table 1.** Minimal \( n \) to achieve a desirable value of \( k_n \).

Another interesting question is to investigate for what values of \( \kappa \) a successful test becomes impossible. When \( m = k = 2 \), [39] showed that no test can successfully distinguish \( H_0 \) from \( H_1 \) provided \( \kappa < 1 \), which essentially shows that \( \kappa > 1 \) is a sufficient and necessary condition for successful detection. Interestingly, it is substantially challenging to obtain such a perfect solution when \( k \) becomes larger. For instance, in ordinary graph setting, [43] obtained a region of SNR when \( k \geq 3 \) in which successful detection becomes impossible. Such region is not sharp and a sufficient and necessary condition for successful detection still remains unknown. In Theorem 2.4 below, we explore this question in hypergraph setting. We derive a region for \( \kappa \) such that a successful test does not exist.

For any integers \( m \geq 3, k \geq 2 \), define

\[
\tau_1(m, k) = \begin{cases} \binom{m}{2}^{-1} \sum_{i=1}^{m-2} \frac{1}{k^{i+2}} \binom{m}{i+2}, & m \text{ is even,} \\ \binom{m}{2}^{-1} \sum_{i=1}^{m-1} \frac{1}{k^{i+2}} \binom{m}{i+2}, & m \text{ is odd,} \end{cases}
\]
and define \( \tau_2(m, k) = 1 + \frac{m}{2} \sum_{i=1}^{k-2} \frac{1}{k^i} (\frac{m}{i+2}) \).

**Theorem 2.4.** Suppose that \( m \geq 3, k \geq 2 \) are integers satisfying \( \tau_1(m, k) \leq 1, a > b > 0 \) are fixed constants and \( \alpha = m - 1 \). If

\[
0 < \kappa < \frac{1}{\tau_2(m, k)(k^2 - 1)},
\]

then the probability measures associated with \( H_0 \) and \( H_1 \) are mutually contiguous.

The proof of Theorem 2.4 relies on Janson’s contiguity theory ([32]). Theorem 2.4 says that when \( \kappa \) falls in the range (5), there is no test that can successfully distinguish the hypotheses \( H_0 \) and \( H_1 \), provided that \( m, k \) satisfy \( \tau_1(m, k) \leq 1 \). It should be emphasized that the condition \( \tau_1(m, k) \leq 1 \) for \( \kappa \) covers most of the practical cases (see [23]).

Combining Theorems 2.4 and 2.2, it is still unknown whether \( H_0 \) and \( H_1 \) are distinguishable when \( \frac{\tau_2(m, k)(k^2 - 1)}{\kappa} \leq 1 \). One way to tackle this might be to enhance Janson’s contiguity theory to efficiently handle hypergraph models. We intend to leave this as one future topic.

### 2.3. A Powerful Test for Dense Uniform Hypergraph.

In this section, we consider the problem of testing community structure in dense \( m \)-uniform hypergraphs. Our approach is based on counting the \( l \)-cycles in the observed hypergraph. To make our test successful, \( l \) needs to be correctly selected according to the hyperedge density of the model. Under such correct selection, we derive asymptotic normality for the test as well as analyze its power. We also comment the effect of misspecified \( l \) in Remark 2.1. Our method can be viewed as a generalization of [25, 26] from ordinary graph testing. The substantially different nature of the hypergraph cycles makes our generalization nontrivial.

Throughout the entire section, we will consider the following slightly different hypothesis testing problem (6):

\[
H_0' : A \sim \mathcal{H}_m(n, \frac{a_n + (km-1)b_n}{k^{m-1}n^{m-1}}) \ \text{vs.} \ \ H_1' : A \sim \mathcal{H}_m^k(n, \frac{a_n}{n^{m-1}}, \frac{b_n}{n^{m-1}}),
\]

where \( n^{l-1} < a_n < b_n < n^{l-\frac{3}{2}} \) for an integer \( 1 \leq l \leq \frac{m}{2} \). Note that problem (2) with \( m - l - \frac{3}{2} < \alpha < m - l \) is a special case of (6) with \( a_n = an^{m-1} \) and \( b_n = bn^{m-1} \). Whereas problem (6) allows more general case of \( a_n, b_n \) such as \( a_n = an^{m-1} \log n \) and \( b_n = bn^{m-1} \log n \). Notably, model (6) allows more sparse scenario such as \( 1 \ll a_n \ll b_n \ll n^{1/3} \) (with \( l = 1 \)), compared with spectral algorithm ([23]) which requires \( a_n \gg (\log n)^{2m+1} \).

We consider the following degree-corrected SBM which is more general than (1). Let \( \{W_i, i = 1, \ldots, n\} \) be i.i.d. random variables with \( E(W_i^2) = 1 \) and \( E(W_1) \neq 0 \). Let \( \{\sigma_i, i = 1, \ldots, n\} \) be i.i.d. random variables from multinomial distribution \( \text{Mult}(k, 1/k) \). Assume that \( W_i \)'s and \( \sigma_i \)'s are independent. Given \( W_i \)'s and \( \sigma_i \)'s, the \( A_{i_1i_2\ldots i_m} \)'s, with pairwise distinct \( i_1, \ldots, i_m \), are conditional independent satisfying

\[
\mathbb{P}(A_{i_1i_2\ldots i_m} = 1|W, \sigma) = W_{i_1} \ldots W_{i_m} p_{i_1i_2\ldots i_m}(\sigma),
\]

\[
\mathbb{P}(A_{i_1i_2\ldots i_m} = 0|W, \sigma) = W_{i_1} \ldots W_{i_m} q_{i_1i_2\ldots i_m}(\sigma),
\]

where \( W = (W_1, \ldots, W_n) \),

\[
p_{i_1i_2\ldots i_m}(\sigma) = \left\{ \begin{array}{ll} \frac{a_n}{n^{m-1}}, & \sigma_{i_1} = \cdots = \sigma_{i_m} \\ \frac{b_n}{n^{m-1}}, & \text{otherwise} \end{array} \right., \quad q_{i_1i_2\ldots i_m}(\sigma) = 1 - p_{i_1i_2\ldots i_m}(\sigma).
\]

We call (7) the degree-corrected SBM in hypergraph setting. The degree-correction weights \( W_i \)'s can capture the degree inhomogeneity exhibited in many social networks. When \( m = 2 \), (7) reduces to the classical degree-corrected SBM in ordinary graph regime (see [54, 37, 27, 25]). In the ordinary graph case, [25] proposed a test through counting small subgraphs to distinguish the degree-corrected SBM from an Erdős-Rényi model. In what follows we generalize their results to hypergraph models through counting the \( l \)-cycles.
A hypertriangle in an $m$-uniform hypergraph is an $l$-cycle consisting of three hyperedges. A hypervee consists of two hyperedges with $l$ common vertices. For example, the hyperedge set $\{(v_1, v_2, v_3, v_4), (v_3, v_4, v_5, v_6), (v_5, v_6, v_1, v_2)\}$ is a hypertriangle, and $\{(v_1, v_2, v_3, v_4), (v_3, v_4, v_5, v_6)\}$ is a hypervee, as shown in Figure 2.

![Fig 2: An example of hypervee (left) and hypertriangle (right) with two common vertices between consecutive hyperedges.](image)

Consider the following probabilities of hyperedge, hypervee and hypertriangle in $\mathcal{H}_n^k(n, \frac{a_n}{n^{m-1}}, \frac{a_n}{n^{m-1}})$:

\[
\begin{align*}
E &= \mathbb{P}(A_{i_1i_2\ldots i_m} = 1), \\
V &= \mathbb{P}(A_{i_1i_2\ldots i_m} A_{i_m-i+1\ldots i_{2m-1}} = 1), \\
T &= \mathbb{P}(A_{i_1i_2\ldots i_m} A_{i_m-i+1\ldots i_{2m-1}} A_{i_{2m-2i+1\ldots i_3(m-i)i_{1\ldots i_1}}(m-i)i_{1\ldots i_1} = 1). 
\end{align*}
\]

It follows from direct calculations that

\[
\begin{align*}
E &= (E_1)^m \frac{a_n + (k^{m-1} - 1)b_n}{n^{m-1}k^{m-1}}, \\
V &= (E_1)^2(m-l) \left( \frac{(a_n - b_n)^2}{n^2(m-l)k^{2m-l-1}} + \frac{2(a_n - b_n)b_n}{n^2(m-l)k^{m-1}} + \frac{b_n^2}{n^2(m-l)} \right), \\
T &= (E_1)^3(m-2l) \left( \frac{(a_n - b_n)^3}{n^3(m-l)k^{3m-l-1}} + \frac{3(a_n - b_n)^2b_n}{n^3(m-l)k^{2m-l-1}} + \frac{3(a_n - b_n)b_n^2}{n^3(m-l)k^{m-1}} + \frac{b_n^3}{n^3(m-l)} \right). 
\end{align*}
\]

Define $T = T - \left( \frac{V}{E} \right)^3$. The following result demonstrates a strong relationship between $T$ and $H_0, H_1$.

**Proposition 2.5.** Under $H_0$, $T = 0$. Moreover, if $\mathbb{E}W_1 \neq 0$, then under $H_1$, $T \neq 0$.

Proposition 2.5 says that, if $\mathbb{E}W_1 \neq 0$, then $H_0$ holds if and only if $T = 0$. Hence, it is reasonable to use an empirical version of $T$, namely $\hat{T}$, as a test statistic for (6).

Prior to constructing $\hat{T}$, let us introduce some notation. For convenience, we use $i_1 : i_m$ to represent the ordering $i_1i_2\ldots i_m$. Also define $C_{2m-l}(A, B) = C_{2m-l}(A, A)$ and $C_{3(m-l)}(A) = C_{3(m-l)}(A, A, A)$ for any adjacency tensor $A$. Here, for any adjacency tensors $A, B, C$:

\[
\begin{align*}
C_{2m-l}(A, B) &= A_{i_1:i_m} B_{i_{m-1+l}^2m-1} + A_{i_2:i_{m+1}} B_{i_{m-l+2}^2m-1} + \cdots + A_{i_{2m-l+1}i_{1:m-1}} B_{i_{2m-l-1}^2m-1}, \\
C_{3(m-l)}(A, B, C) &= A_{i_1:i_m} B_{i_{m-1+l}^2m-1} C_{i_2^2m-l+1} + A_{i_2:i_{m+1}} B_{i_{m-l+2}^2m-l+1} C_{i_3^2m-l+1} + \cdots + A_{i_{2m-l+1}i_{1:m-1}} B_{i_{2m-l-1}^2m-l+1} C_{i_3^2m-l+1}.
\end{align*}
\]
Note that \( C_{2m-l}(A, A) \) is the number of hyperedges in the given vertex ordering \( i_1 i_2 \ldots i_{2m-l} \), while \( C_{3(m-l)}(A, A, A) \) counts the number of hypertriangles in the given vertex ordering \( i_1 i_2 \ldots i_{3(m-l)} \). Define \( \hat{E}, \hat{V}, \hat{T} \) as the empirical versions of \( E, V, T \):

\[
\hat{E} = \frac{1}{\binom{n}{m}} \sum_{c(i,m,n)} A_{i_1 : i_m}, \quad \hat{V} = \frac{1}{\binom{n}{2m-l}} \sum_{c(i,2m-l,n)} \frac{C_{2m-l}(A)}{2m-l}, \quad \hat{T} = \frac{1}{\binom{n}{3(m-l)}} \sum_{c(i,3(m-l),n)} \frac{C_{3(m-l)}(A)}{m-1},
\]

(8)

where the summation index set is defined as \( c(i, s, t) = \{ (i_1, \ldots, i_s) : 1 \leq i_1 < \cdots < i_s \leq t \} \). We have the following asymptotic normality result.

**Theorem 2.6.** Suppose \( EW_1^4 = O(1), EW_1 \neq 0 \) and \( n^{l-1} \ll a_n \ll n^{l-\frac{2}{3}} \) for some integer \( 1 \leq l \leq \frac{m}{2} \). Moreover, let

\[
\delta := \frac{\sqrt{\binom{n}{3(m-l)}(m-l)}}{\sqrt{T}} \left( T - \left( \frac{\hat{V}}{\hat{E}} \right)^3 \right) \in [0, \infty).
\]

(9)

Then we have, as \( n \to \infty \),

\[
\frac{\sqrt{\binom{n}{3(m-l)}(m-l)}}{\sqrt{T}} \left( \hat{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^3 \right) - \delta \to N(0, 1),
\]

(10)

\[
2 \sqrt{\binom{n}{3(m-l)}(m-l)} \left( \sqrt{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^{\frac{3}{2}} \right) - \delta \to N(0, 1).
\]

(11)

When \( l = 1 \) and \( m = 2 \), Theorem 2.6 becomes Theorem 2.2 of [25].

Following (10) in Theorem 2.6, we can construct a test statistic for (6) as

\[
\hat{T} = \frac{\sqrt{\binom{n}{3(m-l)}(m-l)}}{\sqrt{T}} \left( \hat{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^3 \right).
\]

(12)

In practice, \( \hat{T} \) might be close to zero which may cause computational instability, an alternative test can be constructed based on (11) as

\[
\hat{T}' = 2 \sqrt{\binom{n}{3(m-l)}(m-l)} \left( \sqrt{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^{\frac{3}{2}} \right).
\]

(13)

Theorem 2.6 proves asymptotic normality for \( \hat{T} \) and \( \hat{T}' \) under both \( H_0' \) and \( H_1' \). Under \( H_0' \), i.e., \( \delta = 0 \), both \( \hat{T} \) and \( \hat{T}' \) become asymptotically standard normal. Under \( H_1' \), both \( \hat{T} \) and \( \hat{T}' \) are asymptotically normal with mean \( \delta > 0 \) and unit variance. When \( \hat{T} \) has a large magnitude, both test statistics can be used to construct valid rejection regions.

The following Theorem 2.7 says that the power of our test tends to one if \( \delta \) goes to infinity.

**Theorem 2.7.** Suppose \( EW_1^4 = O(1), EW_1 \neq 0 \), \( n^{l-1} \ll a_n \ll n^{l-\frac{2}{3}} \) for some integer \( 1 \leq l \leq \frac{m}{2} \). Under \( H_1' \), as \( n, \delta \to \infty \),

\[
\mathbb{P}( |\hat{T}| > z_{\alpha/2} ) \to 1.
\]

The same result holds for \( \hat{T}' \).

**Remark 2.1.** Theorem 2.6 and Theorem 2.7 may fail for misspecified \( l \). For example, if \( m = 4 \) the true value is \( l_0 = 2 \) (corresponding to the authentic hyperedge density), and we count 1-cycle. Then under \( H_0 \), the test statistic in (10) or (11) is of order \( O_p(n^{\frac{1}{2}}) \), i.e., the limiting distribution does not exist. Whereas, if the true value is \( l_0 = 1 \) and we count 2-cycle, then the test statistic in (10) or (11) have the same limiting distribution (if it exists) under \( H_0 \) and \( H_1 \), i.e., the power of the test does not approach one.
2.4. Extentions to Nonuniform Hypergraph

Non-uniform hypergraph can be considered as a superposition of a collection of uniform hypergraphs, firstly introduced by [23] in which the authors proposed a spectral algorithm for community detection. In this section, we study the problem of testing community structure over a nonuniform hypergraph. Interestingly, our results in Section 2.3 can be extended here without much difficulty.

Let \( H^k(n, M) \) be a nonuniform hypergraph over \( n \) vertices, with the vertices uniformly and independently partitioned into \( k \) communities, and \( M \geq 2 \) is an integer representing the maximum length of the hyperedges. Following [23], we can write \( H^k(n, M) = \bigcup_{m=2}^{M} H^k_m \) \( (n, a_{mn}^{k-1}, b_{mn}^{k-1}) \), where \( H^k_m \) \((n, a_{mn}^{k-1}, b_{mn}^{k-1}) \) are independent uniform hypergraphs with degree-corrected vertices introduced in Section 2.3. Assume that, for \( 2 \leq m \leq M \), \( a_{mn}, b_{mn} \) are proxies of the hyperedge densities satisfying \( n^{l_m-1} \ll a_{mn} \ll b_{mn} \ll n^{l_m-\frac{4}{3}} \), for some integer \( 1 \leq l_m \leq \frac{m}{2} \). Correspondingly, define \( H(n, M) = \bigcup_{m=2}^{M} H_m \) \( (n, a_{mn}^{(k-1)} + (k^m-1)b_{mn}) \) as a superposition of Erdős-Rényi models. Clearly, each Erdős-Rényi model in \( H(n, M) \) has the same average degree as its counterpart in \( H^k(n, M) \), and \( H(n, M) \) has no community structure. Let \( A_m \) denote the adjacency tensor for \( m \)-uniform subhypergraph and \( A = \{A_m, m = 2, \ldots, M\} \) is a collection of \( A_m \)'s. We are interested in the following hypotheses:

\[
H_0^m : A \sim H(n, M) \text{ vs. } H_1^m : A \sim H^k(n, M).
\]

For any \( 2 \leq m \leq M \), let \( \hat{T}_{mn} \) and \( \delta_m \) be defined as (12) and (9), respectively, based on the \( m \)-uniform sub-hypergraph. We define a test statistic for (14) as

\[
\hat{T}_n = \sum_{m=2}^{M} c_m \hat{T}_{mn},
\]

where \( c_m \) are constants with normalization \( \sum_{m=2}^{M} c_m^2 = 1 \). As a simple consequence of Theorems 2.6 and 2.7, we get the asymptotic distribution of \( \hat{T}_n \) as follows.

**Corollary 2.8.** Suppose that the degree-correction weights satisfy the same conditions as in Theorem 2.6, and for any \( 2 \leq m \leq M \), \( n^{l_m-1} \ll a_{mn} \ll b_{mn} \ll n^{l_m-\frac{4}{3}} \), for some integer \( 1 \leq l_m \leq \frac{m}{2} \). Then, as \( n \to \infty \), \( \hat{T}_n - \sum_{m=2}^{M} c_m \delta_m \sim N(0, 1) \). Furthermore, for any constant \( C > 0 \), under \( H_0^m \), \( \mathbb{P}(|\hat{T}_n| > C) \to 1 \), provided that \( \sum_{m=2}^{M} c_m \delta_m \to \infty \) as \( n \to \infty \).

Under \( H_0^m \), i.e., each \( m \)-uniform subhypergraph has no community structure, we have \( \delta_m = 0 \) by Proposition 2.5. Corollary 2.8 says that \( \hat{T}_n \) is asymptotically standard normal. Hence, an asymptotic testing rule at significance \( \alpha \) would be reject \( H_0^m \) if and only if \( |\hat{T}_n| > z_{\alpha/2} \).

The quantity \( \sum_{m=2}^{M} c_m \delta_m \) may represent how \( H_0^m \) separates from \( H_1^m \). By Corollary 2.8, under \( H_1^m \), the test may achieve high power when \( \sum_{m=2}^{M} c_m \delta_m \) is large.

**Remark 2.2.** According to Corollary 2.8, to make \( \hat{T}_n \) have the largest power, we need to maximize the value of \( \sum_{m=2}^{M} c_m \delta_m \) subject to \( \sum_{m=2}^{M} c_m^2 = 1 \). The maximizer is \( c_m^* = \frac{\delta_m}{\sqrt{\sum_{m=2}^{M} \delta_m^2}}, m = 2, 3, \ldots, M \). The corresponding test \( \hat{T}_n^* = \sum_{m=2}^{M} c_{m}^* \hat{T}_{mn} \) becomes asymptotically the most powerful among (15). In particular, \( \hat{T}_n^* \) is more powerful than \( \hat{T}_{mn} \) for a single \( m \). This can be explained by the more hyperedge information involved in the test. This intuition is further confirmed by numerical studies in Section 3. Note that \( \hat{T}_{2n} \) \((m=2)\) is the classic test proposed by [25] in ordinary graph setting. Our test is usually more power than the classic graph-based test.

### 3. Empirical Study

In this section, we provide a simulation study in Section 3.1 and real data analysis in Section 3.2 to assess the finite sample performance of our test.
3.1. Simulation

We generated a nonuniform hypergraph $H^2(n, 3) = H^2_2(n, a_2, b_2) \cup H^2_3(n, a_3, b_3)$, with $n = 100$ under various choices of $(a_m, b_m)$, $m = 2, 3$. In each scenario, we calculated $Z_2 := \hat{T}_2'$ and $Z_3 := \hat{T}_3'$ by (13). Note that $Z_2 = \hat{T}_2'$ is the test for ordinary graph considered in [25]. For testing the community on the nonuniform hypergraph, we calculated the statistic $Z := \hat{T}_{n} = (\hat{T}_2' + \hat{T}_3')/\sqrt{2}$. We examined the the size and power of the test by calculating the rejection proportions based on 500 independent replications at 5% significance level. Let $\delta_m$ denote the quantity defined in (9) which is believed to be the main factor that affects power.

Our study consists of two parts. In the first part, we investigated the power change of the three testing procedures when $\delta_2 = \delta_3 = \delta$ increases from 0 to 10. We designed a setting based on $(a_m, b_m)$ such that $\delta$ is indeed ranging from 0 to 10. Specifically, we set $b_2 = 10b_3$, where $b_3 = 0.01, 0.005, 0.001$ represents the dense, moderately dense and sparse network, respectively; $a_m = r_m b_m$ for $m = 2, 3$ with the values of $r_m$ summarized in Table 2. It can be checked that such choice of $(a_m, b_m)$ indeed makes $\delta$ range from 0 to 10. We examined various ratios of the two community sizes $\gamma = 0.1, 0.3, 0.5$. The rejection proportions under various settings are summarized in Figures 3 through Figure 5. Several interesting findings should be emphasized. First, the rejection proportions of all test statistics at $\delta = 0$ are close to the nominal level 0.05 under different choices of $\gamma$ and $b_3$, which demonstrates that all test statistics are valid. Second, the rejection proportions of the three methods all increase with $\delta$, regardless of the choices of $b_3$ and $\gamma$. Third, the rejection proportions approach one quickly when $\delta$ is close to 10 in the more balanced networks ($\gamma = 0.3, 0.5$), while increase slowly in the extremely unbalanced case ($\gamma = 0.1$). Overall, the testing procedure based on non-uniform hypergraph has larger power than the one only based on 3-uniform hypergraph or 2-uniform graph. This is not surprising since more hyperedge information has been used; see the comments after Corollary 2.8.
In the second part, we investigated the power change along with the hyperedge density. For convenience, we report the results based on the log-scale of $b_3$ which ranges from $-8$ to $-3$. We chose $\delta = 1, 3, \gamma = 0.1, 0.3, 0.5$ and $b_2 = 10b_3$. Similar to the first part, we choose $a_m = r_mb_m$ with $m = 2, 3$ to guarantee that $\log(b_3)$ indeed ranges from $-8$ to $-3$. The values of $r_m$ were summarized in Table 3. Figures 6 and 7 report the rejection proportions for $\delta = 1, 3$ under various hyperedge densities. We notice that larger $b_3$ leads to higher rejection proportion of $Z$. Moreover, $Z$ is more powerful than $Z_2, Z_3$ in the cases $\gamma = 0.3, 0.5$ and $\delta = 3$. In the rest of the scenarios, all procedures have satisfactory performance.

| $\delta$ | $\log(b_3)$ | $-8$ | $-7$ | $-6$ | $-5$ | $-4$ | $-3$ |
|----------|-------------|------|------|------|------|------|------|
| 1        | $r_3$       | 14.18| 6.88 | 3.93 | 2.58 | 1.89 | 1.51 |
|          | $r_2$       | 15.78| 7.03 | 3.72 | 2.37 | 1.74 | 1.42 |
| 3        | $r_3$       | 26.37| 11.51| 5.82 | 3.47 | 2.36 | 1.78 |
|          | $r_2$       | 30.68| 12.54| 5.83 | 3.25 | 2.16 | 1.65 |

Table 3

*Choices of $r_2, r_3, \delta$ for $\log(b_3)$ to range from $-8$ to $-3$.*
3.2. Analysis of Coauthorship Data

In this section, we applied our testing procedure to study the community structure of a coauthorship network dataset. The dataset contains a 2-author graph network and a 3-author hypergraph network. After removing vertices with degrees less than 10 and larger than 20, we obtained a hypergraph (hereinafter referred to as global network) with 58 nodes, 110 edges and 40 hyperedges. The vertex-removal aims to obtain a suitably sparse network so that our testing procedure is applicable. We examined our procedures based on the global network and subnetworks. To do this, we first performed the spectral algorithm proposed by [23] to partition the global network into four subnetworks which consist of 7,13,14,24 vertices, respectively (see Figure 8). In Figure 9, we plot the incidence matrices of the 2- and 3-uniform hypergraphs, denoted 2-UH and 3-UH respectively, as well as their superposition denoted Non-UH. The black dots represent vertices within the same communities. The red crosses represent vertices between different communities. An edge or hyperedge should be drawn between the black dots or red crosses that are vertically aligned. It is observed that the between-community (hyper)edges are more sparse than the within-community ones, indicating the validity of the partitioning.

We conducted testing procedures based on $Z_2, Z_3, Z$ at significance level 0.05 (similar to Section 3.1) to both global network and subnetworks. The values of the test statistics are summarized in Table 4. Observe that $Z_2, Z$ yield very large test values for global network indicating strong rejection of the null hypothesis. For subnetwork testing, $Z_2$ rejects the null hypothesis for subnetwork 3; while $Z_3, Z$ do not reject the null hypotheses for all subnetworks.

4. Proof of Main Results

In this section, we prove the main results of this paper.
4.1. Proof of Theorem 2.1

The proof of Theorem 2.1 is based on one result in Janson ([32]) as below.

Proposition 4.1 (Janson (1995)). Suppose that $L_n = \frac{Q_n}{P_n}$, 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 $\mathbb{E}L = 1$.

Proof of Theorem 2.1: We prove Theorem 2.1 for $k = 2$. The general case follows similarly. In this case, let $\sigma_i \in \{+1, -1\}$ (note that the proof doesn’t rely on this). Define $A_{i_1 : i_m} = A_{i_1 i_2 \ldots i_m}$ and $I[\sigma_{i_1 : i_m}] = I[\sigma_{i_1} = \sigma_{i_2} = \cdots = \sigma_{i_m}]$. Let $d = \frac{(2^m-1)b}{2m-1}$, $p_0 = \frac{d}{n^2}$, $q_0 = 1 - p_0$.

$$p_{i_1 : i_m}(\sigma) = \mathbb{P}(A_{i_1 : i_m} = 1|\sigma) = \left(\frac{a}{n^2}\right)^{I[\sigma_{i_1 : i_m}]} \left(\frac{b}{n^2}\right)^{1-I[\sigma_{i_1 : i_m}]}.$$
and \( q_{i_1:im}(\sigma) = 1 - p_{i_1:im}(\sigma) \). Let \( Y_n \) be ratio of the likelihood function under \( H_1 \) over \( H_0 \). Then
\[
Y_n^2 = 2^{-2n} \sum_{\sigma \eta c(i,m,n)} \prod_{\sigma j_c(i,m,n)} \left( \frac{p_{i_1:im}(\sigma)p_{i_1:im}(\eta)}{p_0^2} \right)^{A_{i_1:im}} \left( \frac{q_{i_1:im}(\sigma)q_{i_1:im}(\eta)}{q_0^2} \right)^{1-A_{i_1:im}}.
\]
The expectation of \( Y_n^2 \) under \( H_0 \) is
\[
\mathbb{E}_0 Y_n^2 = 2^{-2n} \sum_{\sigma \eta c(i,m,n)} \prod_{\sigma j_c(i,m,n)} \left( \frac{p_{i_1:im}(\sigma)p_{i_1:im}(\eta)}{p_0^2} + \frac{q_{i_1:im}(\sigma)q_{i_1:im}(\eta)}{q_0^2} \right). (16)
\]
Define \( s_2 = \{1 \leq i_1 < i_2 < \cdots < i_m \leq n \mid I[\sigma_{i_1} : \sigma_{i_m}] + I[\eta_{i_1} : \eta_{i_m}] = 2 \} \), \( s_1 = \{1 \leq i_1 < i_2 < \cdots < i_m \leq n \mid I[\sigma_{i_1} : \sigma_{i_m}] + I[\eta_{i_1} : \eta_{i_m}] = 1 \} \), \( s_0 = \{1 \leq i_1 < i_2 < \cdots < i_m \leq n \mid I[\sigma_{i_1} : \sigma_{i_m}] + I[\eta_{i_1} : \eta_{i_m}] = 0 \} \). Note that
\[
\frac{1}{p_0} \left( \frac{a}{n^\alpha} \right)^2 + \frac{1}{q_0} \left( 1 - \frac{a}{n^\alpha} \right)^2 = 1 + \frac{(a-d)^2}{dn^\alpha} + \frac{(a-d)^2}{n^{2\alpha}} + O\left( \frac{1}{n^{3\alpha}} \right),
\]
\[
\frac{1}{p_0} \frac{a}{n^\alpha} \frac{b}{n^\alpha} + \frac{1}{q_0} \left( 1 - \frac{a}{n^\alpha} \right) \left( 1 - \frac{b}{n^\alpha} \right) = 1 + \frac{(a-d)(b-d)}{dn^\alpha} + \frac{(a-d)(b-d)}{n^{2\alpha}} + O\left( \frac{1}{n^{3\alpha}} \right),
\]
\[
\frac{1}{p_0} \left( \frac{b}{n^\alpha} \right)^2 + \frac{1}{q_0} \left( 1 - \frac{b}{n^\alpha} \right)^2 = 1 + \frac{(b-d)^2}{dn^\alpha} + \frac{(b-d)^2}{n^{2\alpha}} + O\left( \frac{1}{n^{3\alpha}} \right).
\]
Since \( s_0, s_1, s_2 \) are bounded above by \( n^m \). Then for \( \alpha > \frac{m}{2} \), we have by (16)
\[
\mathbb{E}_0 Y_n^2 = (1 + o(1)) \mathbb{E}_\sigma \left\{ \left( 1 + \frac{(a-d)^2}{dn^\alpha} \right)^{s_2} \left( 1 + \frac{(a-d)(b-d)}{dn^\alpha} \right)^{s_1} \left( 1 + \frac{(b-d)^2}{dn^\alpha} \right)^{s_0} \right\} 
\]
\[
= (1 + o(1)) \mathbb{E}_\sigma \exp \left\{ \frac{(a-d)^2}{dn^\alpha} s_2 + \frac{(a-d)(b-d)}{dn^\alpha} s_1 + \frac{(b-d)^2}{dn^\alpha} s_0 \right\} (17)
\]
If \( \alpha > m, \frac{a}{n^\alpha} \to 0 \), \( i = 0, 1, 2 \), hence \( \mathbb{E}_0 Y_n^2 \to 1 \). Since \( \mathbb{E}_0 Y_n = 1 \), then \( Y_n \) converges to 1 in distribution.

By Proposition 4.1, \( H_0 \) and \( H_1 \) are contiguous. Next we consider \( \alpha = m \). Note that
\[
s_2 = \sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}]I[\eta_{i_1} : \eta_{i_m}],
\]
\[
s_1 = \sum_{c(i,m,n)} (I[\sigma_{i_1} : \sigma_{i_m}] - I[\eta_{i_1} : \eta_{i_m}]) + (1 - I[\sigma_{i_1} : \sigma_{i_m}])I[\eta_{i_1} : \eta_{i_m}],
\]
\[
s_0 = \sum_{c(i,m,n)} (1 - I[\sigma_{i_1} : \sigma_{i_m}]) (1 - I[\eta_{i_1} : \eta_{i_m}]).
\]

Then the numerator of the power in (17) can be written as
\[
(a-d)^2 s_2 + (a-d)(b-d) s_1 + (b-d)^2 s_0
\]
\[
= \frac{n}{m} (b-d)^2 + (a-b)^2 \sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}]
\]
\[
+ (a-b)(b-d) \left( \sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] + \sum_{c(i,m,n)} I[\eta_{i_1} : \eta_{i_m}] \right). (18)
\]

|          | Global Network | SubNetwork 1 | SubNetwork 2 | SubNetwork 3 | SubNetwork 4 |
|----------|----------------|--------------|--------------|--------------|--------------|
| \( n \)  | 58             | 7            | 13           | 14           | 24           |
| \( Z_2 \)| 8.360**       | 0.161        | -0.030       | 2.667*       | 1.661        |
| \( Z_3 \)| 1.451          | -0.100       | -0.211       | -0.289       | -0.052       |
| \( Z \)  | 6.938**        | 0.043        | -0.171       | 1.682        | 1.137        |

Table 4: Values of test statistics based on global network and four subnetworks. Symbols ** and * indicate the strength of rejection, i.e., \( p\text{-value} < 0.001 \) and \( p\text{-value} < 0.05 \) respectively.
For \(s, t = +1, -1\), let

\[
\rho_{st} = \sum_{i=1}^{n} I[\sigma_i = t] I[\eta_i = s], \quad \rho_{t0} = \sum_{i=1}^{n} I[\sigma_i = t], \quad \rho_{0s} = \sum_{i=1}^{n} I[\eta_i = s],
\]

and

\[
\tilde{\rho}_{st} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I[\sigma_i = t] I[\eta_i = s] - \frac{1}{2}), \quad \tilde{\rho}_{t0} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I[\sigma_i = t] - \frac{1}{2}), \quad \tilde{\rho}_{0s} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I[\eta_i = s] - \frac{1}{2}).
\]

It’s easy to verify that \(\sum_{s,t} \tilde{\rho}_{st} = 0, \sum_{s} \tilde{\rho}_{s0} = 0, \sum_{t} \tilde{\rho}_{0t} = 0\) and

\[
\sum_{1 \leq i_1, \ldots, i_m \leq n} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] = m! \sum_{i_1 < i_2 < \ldots < i_m} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] + O(n^{m-1}).
\]

Then we have

\[
\sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] = \frac{1}{m!} \sum_{1 \leq i_1, \ldots, i_m \leq n} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] + O(n^{m-1})
\]

\[
= \frac{1}{m!} \sum_{1 \leq i_1, \ldots, i_m \leq n} \sum_{s,t=-1,+1} \sum_{j=1}^{m} I[\sigma_{i_j} = s] I[\eta_{i_j} = t] + O(n^{m-1})
\]

\[
= \frac{1}{m!} \sum_{s,t=-1,+1} \tilde{\rho}_{st}^m + O(n^{m-1})
\]

\[
= \frac{1}{m!} \sum_{s,t=-1,+1} (\sqrt{n} \tilde{\rho}_{st} + \frac{n}{2})^m + O(n^{m-1})
\]

\[
= \frac{1}{m!} \frac{4n^m}{2^{2m}} + \frac{1}{m!} n^{m-1} \sum_{s,t} \tilde{\rho}_{st}^2 \sum_{k=2}^{m} \binom{m}{k} \frac{1}{2^{2(m-k)}} \left( \frac{\tilde{\rho}_{st}}{\sqrt{n}} \right)^{k-2} + O(n^{m-1})
\]

\[
\sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] = \frac{1}{m!} \sum_{1 \leq i_1, \ldots, i_m \leq n} I[\sigma_{i_1} : \sigma_{i_m}] + O(n^{m-1})
\]

\[
= \frac{1}{m!} \sum_{t=-1,+1} \tilde{\rho}_{t0}^m + O(n^{m-1})
\]

\[
= \frac{1}{m!} \frac{2n^m}{2^m} + \frac{n^{m-1}}{m!} \sum_{t} \tilde{\rho}_{t0}^2 \sum_{k=2}^{m} \binom{m}{k} \frac{1}{2^{2(m-k)}} \left( \frac{\tilde{\rho}_{t0}}{\sqrt{n}} \right)^{k-2} + O(n^{m-1})
\]

\[
\sum_{c(i,m,n)} I[\eta_{i_1} : \eta_{i_m}] = \frac{1}{m!} \sum_{1 \leq i_1, \ldots, i_m \leq n} I[\eta_{i_1} : \eta_{i_m}] + O(n^{m-1})
\]

\[
= \frac{1}{m!} \sum_{s=-1,+1} \tilde{\rho}_{0s}^m + O(n^{m-1})
\]

\[
= \frac{1}{m!} \frac{2n^m}{2^m} + \frac{n^{m-1}}{m!} \sum_{s} \tilde{\rho}_{0s}^2 \sum_{k=2}^{m} \binom{m}{k} \frac{1}{2^{2(m-k)}} \left( \frac{\tilde{\rho}_{0s}}{\sqrt{n}} \right)^{k-2} + O(n^{m-1})
\]

If \(\alpha = m\), by (18), the third equation of (19), the second equation of (20), the second equation of (21) and the law of large number, we have

\[
(a - d)^2 \frac{s_2}{n^m} + (a - d)(b - d) \frac{s_1}{n^m} + (b - d)^2 \frac{s_0}{n^m} \rightarrow (a - b)^2 \frac{4}{2^m} + (a - b)(b - d) \frac{4}{2^m} + (b - d)^2
\]

\[
= \left( \frac{a - b}{2^{m-1}} + (b - d) \right)^2 = 0.
\]
Combining (17) and (22), $E_0 Y_n^2 \to 1$, which implies that $H_0$ and $H_1$ are contiguous by Proposition 4.1.

Let $\alpha = m - 1 + \delta, \delta > 0$. Note that $|\frac{\mu_n}{\sqrt{n}}|, |\frac{\rho_{m}}{\sqrt{n}}|, |\frac{\rho_{m}}{\sqrt{n}}|$ are all bounded by 1. Hence, for universal constant $C$, we have

$$
\frac{(a-b)^2}{dm!} \sum_{k=2}^m \left( m \choose k \right) \frac{1}{2^{k(m-k)}} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^{k-2} \leq C,
$$

$$
\frac{(a-b)(b-d)}{dm!} \sum_{k=2}^m \left( m \choose k \right) \frac{1}{2^{k(m-k)}} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^{k-2} \leq C,
$$

$$
\frac{(a-b)(b-d)}{dm!} \sum_{k=2}^m \left( m \choose k \right) \frac{1}{2^{k(m-k)}} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^{k-2} \leq C.
$$

Note that $(b-d)^2 + \frac{1}{2n^2}(a-b)^2 + \frac{1}{2n^2}(a-b)(b-d) = 0$. Then by (17), (18), (19), (20), (21), we have

$$
E_0 Y_n^2 \leq (1 + o(1)) E_{\sigma^2} \exp \left\{ \sum_{st} \frac{C}{n^3} \hat{\rho}_{st}^2 + \sum_t \frac{C}{n^3} \hat{\rho}_{st}^2 + \sum_s \frac{C}{n^3} \hat{\rho}_{os}^2 + O \left( \frac{1}{n^3} \right) \right\}.
$$

(23)

By central limit theorem and Slutsky theorem, $\hat{\rho}_{st}^2, \hat{\rho}_{st}^2, \hat{\rho}_{os}^2$ converge to chi-square in distribution, which implies that $C \hat{\rho}_{st}^2, C \hat{\rho}_{os}^2$ converge to zero in distribution when $\delta > 0$. Note that for any $\gamma > 0$ and $\beta > 0$, by concentration inequality, we have

$$
P\left( \exp \left\{ \frac{C}{n^3} \hat{\rho}_{st}^2 \right\} > \gamma^2 \beta \right) = P\left( \left| \hat{\rho}_{st} \right| > \sqrt{\frac{n^3 \log \gamma^2 \beta}{Cn}} \right) \leq 2 \exp \left\{ - \frac{n^3 \log \gamma^2 \beta}{Cn} \right\} = 2 \gamma^{-\beta \frac{n^3}{C_1}}.
$$

Hence for some large $n_0$, $n \geq n_0$, it follows that

$$
\int_{C_1}^{\infty} P\left( \exp \left\{ \frac{C}{n^3} \hat{\rho}_{st}^2 \right\} > \gamma^2 \beta \right) d\gamma \leq \frac{2}{\beta \mu_n^2} - C_1 = 0, \quad C_1 \to \infty.
$$

Since

$$
P\left( \exp \left\{ \sum_{st} \frac{C}{n^3} \hat{\rho}_{st}^2 + \sum_t \frac{C}{n^3} \hat{\rho}_{st}^2 + \sum_s \frac{C}{n^3} \hat{\rho}_{os}^2 \right\} > t \right)
\leq \sum_{t,s} P\left( \exp \left\{ \frac{C}{n^3} \hat{\rho}_{st}^2 \right\} > t \right) + \sum_t P\left( \exp \left\{ \frac{C}{n^3} \hat{\rho}_{st}^2 \right\} > t \beta \right) + \sum_t P\left( \exp \left\{ \frac{C}{n^3} \hat{\rho}_{os}^2 \right\} > t \right),
$$

then, by uniform integrability and (23), we conclude that $E_0 Y_n^2 \to 1$, hence $H_0$ and $H_1$ are contiguous by Proposition 4.1.

\[\square\]

### 4.2. Proof of Theorem 2.2

The key idea in proving Theorem 2.2 is to count long loose cycles and use Theorem 1 in Gao and Wormald ([28]). Here “long” means that the number of edges of the loose cycle goes to infinity as $n$ diverges. We state Theorem 1 from Gao and Wormald ([28]) below.

**Theorem 4.2** (Gao and Wormald (2004)). Let $s_n > -\frac{1}{\mu_n}$ and $\sigma_n = \sqrt{\mu_n + \mu_n^2 s_n}$, with $0 < \mu_n \to \infty$. Suppose that $\mu_n = o(\sigma_n^3)$ and a sequence $\{X_n\}$ of nonnegative random variables satisfies

$$
E[X_n] \sim \mu_n^k \exp \left( \frac{k^2 s_n}{2} \right)
$$

uniformly for all integers $k$ in the range $c_1 \mu_n / \sigma_n \leq k \leq c_2 \mu_n / \sigma_n$ for some constants $c_2 > c_1 > 0$. Then $(X_n - \mu_n) / \sigma_n$ tends in distribution to the standard normal as $n \to \infty$. 

**Proof of Theorem 2.2:** Let $X_{k_n}$ be the number of $k_n$-edge loose cycle, where $k_n$ is defined before Theorem 2.3. We compute the expectation of $[X_{k_n}]_s$ under $H_1$ with $\alpha = m - 1$. For a tuple of $k_n$-edge loose cycles $(H_{k_n,1}, \ldots, H_{k_n,s})$, let $A$ be the set of tuples with disjoint cycles and $\bar{A}$ be the set of tuples where at least two cycles intersect each other. Then expectation of $[X_{k_n}]_s$ under $H_1$ can be expressed as

$$
\mathbb{E}_1[X_{k_n}]_s = \sum_{A} \mathbb{E}_1 I_{\cup_j \neq 1 H_{k_n,j}} + \sum_{\bar{A}} \mathbb{E}_1 I_{\cup_j \neq 1 H_{k_n,j}}.
$$

Let $\tau$ be a random label assignment. The first term in the right hand side of the above equation is

$$
\mathbb{E}_1 I_{\cup_j \neq 1 H_{k_n,j}} = \mathbb{E}_1 \prod_{i=1}^{s} I_{H_{k_n,i}} = \mathbb{E}_1 \prod_{i=1}^{s} \mathbb{E}_1 I_{H_{k_n,i}} = \mathbb{E}_\tau \prod_{i=1}^{s} \prod_{i_1 \neq \ldots \neq i_m \in \mathcal{E}(H_{k_n,i})} \frac{M_{i_1 i_2 \ldots i_m}(\tau)}{n^{m-1}}
$$

$$
= \sum_{i=1}^{s} \left(\frac{a + (k^{m-1} - 1)b}{(kn)^{m-1}} k_n^{n_i} + (k - 1) \left(\frac{a - b}{(kn)^{m-1}}\right)^{k_n}\right)
$$

$$
= \frac{1}{n^{(m-1)k_n s}} \left(\frac{a + (k^{m-1} - 1)b}{k^{m-1}} k_n^{n_i} + (k - 1) \left(\frac{a - b}{k^{m-1}}\right)^{k_n}\right)^s.
$$

Note that $|A| = \frac{n!}{(n-M_1)!} \left(\frac{1}{(2k_n(m-2)n)^m}\right)^s$, where $M_1 = (m-1)k_n s$. Then

$$
\sum_{A} \mathbb{E}_1 I_{\cup_j \neq 1 H_{k_n,j}} = |A| \mathbb{E}_1 I_{\cup_j \neq 1 H_{k_n,j}}
$$

$$
= \frac{n!}{(n-M_1)!} \left(\frac{1}{2k_n}\right)^m \left(\frac{1}{k^{m-1}(m-2)!}\right)^m + \frac{(k-1)}{2k_n}\left(\frac{a - b}{k^{m-1}(m-2)!}\right)^{k_n}\right)^s.
$$

if $M_1 = o(\sqrt{n})$. To see this, note that by Taylor expansion we have

$$
\log\left(\frac{n!}{(n-M_1)!} n^{-M_1}\right) = \log(1 - \frac{1}{n}) + \log(1 - \frac{2}{n}) + \cdots + \log(1 - \frac{M_1 - 1}{n})
$$

$$
\geq - \frac{M_1(M_1 - 1)}{2n} - \sum_{k=2}^{\infty} \frac{M_1(M_1 - 1)^k}{kn^{k}}\frac{1}{k!}
$$

$$
\geq - \frac{M_1(M_1 - 1)}{2n} - \sum_{k=2}^{\infty} \frac{M_1(M_1 - 1)^k}{k!} \frac{M_1 - 1}{\sqrt{n}}
$$

$$
= - \frac{M_1(M_1 - 1)}{2n} + \frac{(M_1 - 1)^2}{n} + M_1 - 1 \frac{1}{\sqrt{n}} \log\left(1 - \frac{M_1 - 1}{\sqrt{n}}\right) \to 0,
$$

if $M_1 = o(\sqrt{n})$. Note that $|\bar{A}| \leq M_1^2 n^{M_1-1}$ and $\mathbb{E}_1[I_{\cup_j \neq 1 H_{k_n,j}} | \tau] \leq \left(\frac{a}{\sqrt{n}}\right)^{|\mathcal{E}(H)|}$, then

$$
\sum_{\bar{A}} \mathbb{E}_1 I_{\cup_j \neq 1 H_{k_n,j}} \leq M_1^2 n^{M_1-1} \left(\frac{a}{\sqrt{n}}\right)^{|\mathcal{E}(H)|} = M_1^2 \left(\frac{a}{n^{m-1}}\right)^{|\mathcal{E}(H)|} \to 0.
$$

if $M_1 \leq \delta_1 \log_a n$, for some $\delta_1$ with $0 < \delta_1 < 1$.

Define $\mu_{n1} = \frac{1}{2k_n} \left[\frac{a + (k^{m-1} - 1)b}{k^{m-1}(m-2)!}\right]^{k_n} + \frac{(k-1)}{2k_n} \left[\frac{a - b}{k^{m-1}(m-2)!}\right]^{k_n}$ and $\mu_{n0} = \frac{1}{2k_n} \left[\frac{a + (k^{m-1} - 1)b}{k^{m-1}(m-2)!}\right]^{k_n}$. If $M_1 \leq \delta_1 \log_a n$, for some $\delta_1$ with $0 < \delta_1 < 1$, it’s clear that

$$
\mathbb{E}_1[X_{k_n}]_s \sim \mu_{n1}^s.
$$

Similarly, if $M_1 \leq \delta_1 \log_a n$, for some $\delta_1$ with $0 < \delta_1 < 1$, we have

$$
\mathbb{E}_0[X_{k_n}]_s \sim \mu_{n0}^s.
$$

(24)
Note that $\kappa > 1$ implies $\lambda_m > 1$. To see this, let $a = c + (k^{m-1} - 1)d$ and $b = c - d$ for some constants $c > d > 0$. Then it follows from $\kappa > 1$ that $c > (m - 2)!$, which yields $\lambda_m > 1$. Then $\mu_{n1}, \mu_{n0} \to \infty$ as $n \to \infty$. It’s obvious that

$$
\mu_{n1} \leq \left( \log_\gamma n \right) \frac{\delta_0}{k_n}, \quad \mu_{n0} \leq \left( \log_\gamma n \right) \frac{\delta_0}{k_n}.
$$

Let $\sigma_{n1} = \sqrt{\mu_{n1}}, \sigma_{n0} = \sqrt{\mu_{n0}}$. For any constant $c_2 > c_1 > 0$ and $s$ satisfying $c_1 \frac{\sigma_{n1}}{\sigma_{n0}} \leq s \leq c_2 \frac{\sigma_{n1}}{\sigma_{n0}}$ or $c_1 \frac{\sigma_{n0}}{\sigma_{n1}} \leq s \leq c_2 \frac{\sigma_{n0}}{\sigma_{n1}}$, we have for large $n$

$$
M_1 = (m - 1)k_ns = (m - 1)\sqrt{\left( \log_\gamma n \right) \frac{\delta_0}{k_n} \left( \log_\gamma n \right) \frac{\delta_0}{k_n} \leq \delta_1 \log_a n},
$$

which implies (24) and (25) hold. By Theorem 4.2, we conclude that $\frac{X_{kn} - \mu_{n1}}{\sqrt{\mu_{n1}}}$ and $\frac{X_{kn} - \mu_{n0}}{\sqrt{\mu_{n0}}}$ converge in distribution to standard normal distribution under $H_1$ and $H_0$ respectively.

Since $\kappa > 1$, there exists constant $\rho$ satisfying

$$
\sqrt{\frac{a + (k^{m-1} - 1)b}{k^{m-1}(m - 2)!}} < \rho < \frac{a - b}{k^{m-1}(m - 2)!}.
$$

It’s easy to verify that $\mu_{n1} = o(\rho^{2k_n}), \mu_{n0} = o(\rho^{2k_n})$. Define event $A_n = \{X_{kn} \leq \mathbb{E}_0 X_{kn} + \rho^{k_n}\}$. Then we have

$$
P_{n0}(A_n) = P_{n0}\left( \frac{X_{kn} - \mu_{n0}}{\sqrt{\mu_{n0}}} \leq \frac{\rho^{k_n}}{\sqrt{\mu_{n0}}} \right) \to \Phi(\infty) = 1. \tag{26}
$$

Note that $\frac{\mu_{n1} - \mu_{n0}}{\rho^{k_n}} \to \infty$, then for large $n$, we have $\mu_{n1} - \rho^{k_n} \geq \mu_{n0} + \rho^{k_n}$. Then it yields

$$
P_{n1}(A_n) \leq P_{n1}\left( X_{kn} \leq \mathbb{E}_1 X_{kn} - \rho^{k_n} \right) = P_{n1}\left( \frac{X_{kn} - \mu_{n1}}{\sqrt{\mu_{n1}}} \leq - \frac{\rho^{k_n}}{\sqrt{\mu_{n1}}} \right) \to \Phi(-\infty) = 0. \tag{27}
$$

By definition, (26) and (27) shows that $H_0$ and $H_1$ are orthogonal.

\section*{4.3. Proof of Theorem 2.4}

Several lemmas and proposition are presented before we prove Theorem 2.4.

\textbf{Lemma 4.3.} Let $M_0$ be the following $k \times k$ matrix

$$
M_0 = \begin{bmatrix}
(a + (k^{m-2} - 1)b) & k^{m-2}b & \cdots & k^{m-2}b \\
k^{m-2}b & a + (k^{m-2} - 1)b & \cdots & k^{m-2}b \\
\vdots & \vdots & \ddots & \vdots \\
k^{m-2}b & k^{m-2}b & \cdots & a + (k^{m-2} - 1)b
\end{bmatrix}.
$$

Then the trace of $M_0^j$ is

$$
Tr(M_0^j) = (a + (k^{m-2} - 1)b)^j + (k - 1)(a - b)^j,
$$

for any positive integer $j$.

\textbf{Proof of Lemma 4.3:} Note that $M_0 = (a - b)I + k^{m-2}bJ$, where $I$ is $k \times k$ identity matrix and $J$ is $k \times k$ matrix with every entry 1. For any real number $\lambda$, we have

$$
M_0 - \lambda I = (a - b - \lambda)I + k^{m-2}bJ = k^{m-2}b\left( J - \frac{\lambda - a + b}{k^{m-2}b} I \right).
$$
Then \( \det(M_0 - \lambda I) = 0 \) implies that \( \det(J - \frac{\lambda - a + b}{k - 2} I) = 0 \). The eigenvalue of \( J \) are \( k \) and \( 0 \) with multiplicity \( k - 1 \), which implies \( \lambda = a - b, a + (k^{m-2} - 1)b \) and the desired result follows.

\[
\text{Lemma 4.4. For positive integer } j \text{ and } i_1, \ldots, i_{jm-j} \in \{1, 2, \ldots, k\}, \text{ let } M_{i_1i_2\ldots i_m} = (a - b)I[i_1 = i_2 = \ldots = i_m] + b. \text{ Then we have}
\]

\[
\sum_{i_1, \ldots, i_{jm-j} \in \{1, \ldots, k\}} M_{i_1i_2\ldots i_m} M_{i_{m-2m-1}i_{2m-1}\ldots i_{3m-2} \ldots M_{i_{(j-1)m-(j-2)}i_{jm-j}i_1} = Tr(M_0^2),
\]

where \( M_0 \) is the same as in Lemma 4.3.

**Proof of Lemma 4.4:** Let \( I_j = (i_{(j-1)m-j+3}, \ldots, i_{jm-j}) \). Then we have

\[
\sum_{i_1, \ldots, i_{jm-j} \in \{1, \ldots, k\}} M_{i_1i_2\ldots i_m} M_{i_{m-2m-1}i_{2m-1}\ldots i_{3m-2} \ldots M_{i_{(j-1)m-(j-2)}i_{jm-j}i_1} = \sum_{I_1, I_2, \ldots, I_j} Tr(M(I_1)M(I_2) \ldots M(I_j)),
\]

where \( M(I_i) = (M_{I_i})_{i=1}^k \) is a \( k \times k \) matrix. By the definition of \( M_{i_1i_2\ldots i_m} \), it follows that

\[
M(I_i) = \begin{bmatrix}
  a & b & \ldots & b \\
  b & a & \ldots & b \\
  \vdots & \vdots & \ddots & \vdots \\
  b & b & \ldots & a
\end{bmatrix} + \begin{bmatrix}
  b & b & \ldots & b \\
  b & b & \ldots & b \\
  \vdots & \vdots & \ddots & \vdots \\
  b & b & \ldots & b
\end{bmatrix} + \cdots + \begin{bmatrix}
  b & b & \ldots & b \\
  b & b & \ldots & b \\
  \vdots & \vdots & \ddots & \vdots \\
  b & b & \ldots & b
\end{bmatrix} + \sum_{I_1: \text{elements are different}} M(I_i)
\]

\[
= \begin{bmatrix}
  a + (k-1)b & kb & \ldots & kb \\
  kb & a + (k-1)b & \ldots & kb \\
  \vdots & \vdots & \ddots & \vdots \\
  kb & kb & \ldots & a + (k-1)b
\end{bmatrix} + \begin{bmatrix}
  b & b & \ldots & b \\
  b & b & \ldots & b \\
  \vdots & \vdots & \ddots & \vdots \\
  b & b & \ldots & b
\end{bmatrix} = M_0,
\]

which completes the proof.

In this section, we employ the following Proposition 4.5, which was proved by Janson ([32]), to prove our main theorems. For any non-negative integer \( x \), let \( [x]_j \) denote the product \( x(x-1) \cdots (x-j+1) \).

**Proposition 4.5 (Janson (1995)).** Let \( \lambda_i > 0, i = 1, 2, \ldots, \) be constants and suppose that for each \( n \) there are random variables \( X_{i_1}, i = 1, 2, \ldots, \) and \( Y_n \) (defined on the same probability space) such that \( X_{i_1} \) is non-negative integer valued and \( \mathbb{E}[Y_n^2] \neq 0 \) (at least for large \( n \)), and furthermore the following conditions are satisfied:

(A1) \( X_{i_1} \overset{d}{\to} Z_i \) as \( n \to \infty \), jointly for all \( i \), where \( Z_i \sim \text{Poisson}(\lambda_i) \) are independent Poisson random variables;

(A2) \( \mathbb{E}[Y_n[X_{i_1}, \ldots, X_{i_n}]] = \mathbb{E}[Y_n] \to \prod_{i=1}^k \mu_i^{x_i}, \) as \( n \to \infty \), for some \( \mu_i \geq 0 \) and every finite sequence \( j_1, \ldots, j_k \) of non-negative integers;

(A3) \( \sum_{i=1}^\infty \lambda_i \delta_i^2 < \infty \), where \( \delta_i = \mu_i/\lambda_i - 1; \)

(A4) \( \mathbb{E}[Y_n^2]/(\mathbb{E}[Y_n])^2 \to \exp(\sum_{i=1}^\infty \lambda_i \delta_i^2). \)
Then
\[ \frac{Y_n}{\mathbb{E}\{Y_n\}} \xrightarrow{d} W \equiv \prod_{i=1}^{\infty} (1 + \delta_i) Z_i \exp(-\lambda_0 \delta_i), \quad \text{as } n \to \infty, \]
and \( \mathbb{E}W = 1. \)

We also need the following lemma.

**Lemma 4.6.** Let \( X_{hn} \) be the number of \( h \)-edge loose cycle in \( \mathcal{H}_m(n, \frac{d}{n^{m-1}}) \), where \( d = \frac{a+(k^{m-1}-1)b}{k^{m-1}} \) and \( h \geq 2 \). Let \( \lambda_h = \frac{d^h}{2h(m-2)!} \). Then for any integer \( s \geq 2 \), \( \{X_{hn}\}_{h=2}^s \) jointly converges to independent Poisson variables with mean \( \lambda_h \).

**Proof of Lemma 4.6:** Let \( H \) be a graph on a subset of \([n]\) with vertex set \( \mathcal{V}(H) \) and edge set \( \mathcal{E}(H) \). For any sequence of positive integers \( j_2, j_3, \ldots, j_s \), we have
\[
\prod_{h=2}^{s} X_{hn} = \sum_{(H_{hi})} \prod_{h=2}^{s} \prod_{i=1}^{j_h} 1_{H_{hi}}.
\]
Then
\[
\mathbb{E}_0 \prod_{h=2}^{s} X_{hn} = \mathbb{E}_0 \prod_{h=2}^{s} \prod_{i=1}^{j_h} 1_{H_{hi}} = \sum_{(H_{hi})} \mathbb{E}_0 \prod_{h=2}^{s} \prod_{i=1}^{j_h} 1_{H_{hi}} + \sum_{(H_{hi})} \mathbb{E}_0 \prod_{h=2}^{s} \prod_{i=1}^{j_h} 1_{H_{hi}}.
\]
(28)
The summand in the first term of (28) can be calculated as below
\[
\mathbb{E}_0 \prod_{h=2}^{s} \prod_{i=1}^{j_h} 1_{H_{hi}} = \mathbb{E}_0 \prod_{h=2}^{s} \prod_{i=1}^{j_h} 1_{H_{hi}} = \prod_{h=2}^{s} \prod_{i=1}^{j_h} \frac{d^h}{2h(m-2)!}.
\]
Note that \( |A| = \frac{n!}{(n-M_1)!} \prod_{h=2}^{s} (\frac{d^h}{2h(m-2)!})^j_h \), \( M_1 = (m-1) \sum_{h=2}^{s} h j_h \). Hence the first term in the right hand side of (28) by Lemma 4.3 is
\[
|A| \prod_{h=2}^{s} \prod_{i=1}^{j_h} \frac{d^h}{2h(m-2)!} \leq \frac{n!}{(n-M_1)!} \prod_{h=2}^{s} \frac{d^h}{2h(m-2)!} \to \prod_{h=2}^{s} \lambda_h^{j_h}.
\]
For \( (H_{hi}) \in \mathcal{A}, H = \cup H_{hi} \) has at most \( M_1 - 1 \) vertices and \( \sum_{h=2}^{s} h j_h \) hyperedges, and hence \( |\mathcal{V}(H)| < |\mathcal{E}(H)|(m-1) \), and
\[
\mathbb{E}_0 \prod_{h=2}^{s} \prod_{i=1}^{j_h} 1_{H_{hi}} = \prod_{(i_1, \ldots, i_m) \in \mathcal{E}(H)} \left( \frac{a}{n^{m-1}} \right)^{1[\tau_u=\tau_v]} \left( \frac{b}{n^{m-1}} \right)^{1[\tau_u \neq \tau_v]} \leq \left( \frac{a}{n^{m-1}} \right)^{|\mathcal{E}(H)|}.
\]
There are \( \binom{n}{|\mathcal{V}(H)|} \mathbb{E}(|\mathcal{V}(H)|) \) graphs isomorphic to \( H \). Then
\[
\sum_{H': \text{isomorphic to } H} E_1[1_{H'} | \tau] \leq \left( \frac{a}{n^{m-1}} \right)^{|\mathcal{E}(H)|} \left( \frac{n}{|\mathcal{V}(H)|} \right) \mathbb{E}(|\mathcal{V}(H)|) \to 0.
\]
Since the number of isomorphism classes is bounded, the second term in the right hand side of (28) goes to zero. Hence, \( \mathbb{E}_0 \prod_{h=2}^{s} X_{hn} \to \prod_{h=2}^{s} \lambda_h^{j_h} \), which completes the proof by Lemma 2.8 in Wormald([50]).

**Proof of Theorem 2.4:** We prove Theorem 2.4 by Proposition 4.5. Let \( \lambda_h = \frac{1}{2h} \left( \frac{a+(k^{m-1}-1)b}{k^{m-1}(m-2)!} \right)^h \) and \( \delta_h = (k-1) \left( \frac{a-b}{a+(k^{m-1}-1)b} \right)^h \). Condition (A1) follows from Lemma 4.6.
Next, we check condition (A2). Let $S = \{1, 2, \ldots, k\}$. Note that for any sequence of positive integers $j_2, \ldots, j_s$, we have

$$E_0Y_n[X_{2n}]j_2 \ldots [X_{sn}]j_s = \sum_{H \in A} E_0Y_n1_H + \sum_{H \in A} E_0Y_n1_H.$$  

(29)

Direct computation yields

$$E_0Y_n1_H = \frac{1}{k^n} \sum_{\sigma \in S^n} E_01_H \prod_{e(i, m, n)} \left( \frac{p_{i1:im}(\sigma)}{p_0} \right)^{A_{i1:im}} \left( \frac{q_{i1:im}(\sigma)}{q_0} \right)^{1-A_{i1:im}}$$

$$= \frac{1}{k^n} \sum_{\sigma \in S^n} \sum_{(i, \ldots, m) \in E(H)} \prod_{e(i, m, n)} \left( \frac{p_{i1:im}(\sigma)}{p_0} \right)^{A_{i1:im}} \left( \frac{q_{i1:im}(\sigma)}{q_0} \right)^{1-A_{i1:im}}$$

$$= k^{-|V(H)|} \sum_{\sigma \in S^{\mid V(H)\mid}} \sum_{(i, \ldots, m) \in E(H)} \prod_{e(i, m, n)} \left( \frac{p_{i1:im}(\sigma)}{p_0} \right)^{A_{i1:im}} \left( \frac{q_{i1:im}(\sigma)}{q_0} \right)^{1-A_{i1:im}}.$$

where the second equality follows by the independence of $A_{i1:im}$. Define $\sigma^{1hi}$ and $\sigma^{2hi}$ to be the restrictions of $\sigma$ on $V(H_{m_i})$ and $[n]\setminus V(H_{m_i})$. Similarly, $\sigma^1$ and $\sigma^2$ are the restrictions of $\sigma$ on $V(H)$ and $[n]\setminus V(H)$. Then by the above equation, we have

$$E_0Y_n1_H = k^{-M_1} \sum_{\sigma \in S^{\mid V(H)\mid}} \prod_{(i, \ldots, m) \in E(H)} p_{i1:im}(\sigma^1)$$

$$= E_{\sigma^1} \prod_{(i, i) \in E(H)} p_{i1:im}(\sigma^1)$$

$$= \prod_{h=2}^s j_h \prod_{i=1}^h E_{\sigma^{1hi}} \prod_{(i, \ldots, m) \in E(H^{hi})} p_{i1:im}(\sigma^{1hi})$$

$$= \prod_{h=2}^s j_h \prod_{i=1}^h E_{\sigma^{1hi}} \prod_{(i, \ldots, m) \in E(H^{hi})} M_{g_{i1:im}}^{1hi}$$

$$= \prod_{h=2}^s j_h \prod_{i=1}^h E_{\sigma^{1hi}} \prod_{(i, \ldots, m) \in E(H^{hi})} \frac{M_{g_{i1:im}}^{1hi}}{\eta^{h(m-1)}}$$

$$= \prod_{h=2}^s j_h \prod_{i=1}^h \frac{Tr(M_0^h)}{k^{h(m-1)}\eta^{h(m-1)}}.$$

where we used Lemma 4.4 for the last equality. Note $|A| = \frac{n!}{(n-M_1)!} \prod_{h=2}^s (\frac{1}{2k(n-2)^h})^{j_h}$, $M_1 = (m - 1) \sum_{h=2}^s h j_h$. Hence the first term in (29) by Lemma 4.3 is
\[
|A| \prod_{h=2}^{s} \prod_{i=1}^{h} \frac{Tr(M_{ih}^n)}{k_h(m-1)^{m_h(h-1)}} = \frac{n!}{(n-M_1)!^{M_1}} \prod_{h=2}^{s} \left( \frac{1}{2h(m-2)^h} \right)^{j_h} \left( d^h + \frac{(k-1)(a-b)^h}{k^{m-1}} \right)^{j_h} \\
= \frac{n!}{(n-M_1)!^{M_1}} \prod_{h=2}^{s} [\lambda_h(1 + \delta_h)]^{j_h} \to \prod_{h=2}^{s} [\lambda_h(1 + \delta_h)]^{j_h}.
\]

For \( H \in \mathcal{A} \), one has

\[
\mathbb{E}_0 Y_n 1_H = k^{-n} \sum_{\sigma \in S_n} \mathbb{E}_0 1_H \prod_{(i_1, \ldots, i_m) \in E(H)} \left( \frac{p_{i_1 : \ldots : i_m}(\sigma)}{p_0} \right)^{A_{uv}} \left( \frac{q_{i_1 : \ldots : i_m}(\sigma)}{q_0} \right)^{1-A_{uv}}
\]

\[
\leq k^{-n} p_0^{ |V(H)| } \sum_{\sigma \in S_n} p_0^{- |V(H)|} \left( \frac{a^{n^{m-1}}}{n^{|V(H)|}} \right)^{ |V(H)| !} \to 0,
\]

and \( \sum_{H \in \mathcal{A}} \mathbb{E}_0 Y_n 1_H \to 0 \). Hence, \( \mathbb{E}_0 Y_n [X_{2n}]_{j_2} \cdots [X_{sn}]_{j_s} \to \prod_{h=2}^{s} [\lambda_h(1 + \delta_h)]^{j_h} \).

Lastly, we check condition (A4). Note that

\[
\mathbb{E}_0 Y_n^2 = (1+o(1)) \exp \left\{ \frac{1}{m-1} \left( \binom{n}{m} (b-d)^2 + (a-b)^2 \sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] \right) \right\}
\]

\[
+(a-b)(b-d) \left( \sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] + \sum_{c(i,m,n)} I[\eta_{i_1} : \eta_{i_m}] \right) \}
\]

Let \( C = \{(i_1, \ldots, i_m)| \exists i, i', i'' \neq i, i'' if s', t' \notin \{s, t\} \} \). Then

\[
\sum_{i_1, i_2, \ldots, i_m} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] = m! \sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] + \sum_{C} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] + O(n^{m-2}).
\]

Direct computation yields

\[
\sum_{c(i,m,n)} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}]
\]

\[
= \frac{1}{m!} \sum_{i_1, i_2, \ldots, i_m} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] - \frac{1}{m!} \sum_{C} I[\sigma_{i_1} : \sigma_{i_m}] I[\eta_{i_1} : \eta_{i_m}] + O(n^{m-2})
\]

\[
= \frac{1}{m!} \sum_{s,t=1}^{k} (\sqrt{n} \tilde{\rho}_{st} + \frac{n}{k^2})^{m-1} \frac{1}{m!} \left( \sum_{i=1}^{m} \left( \frac{\tilde{\rho}_{st}}{\sqrt{n}} \right)^i \right)
\]

\[
= \frac{1}{m!} \frac{m^m}{k^2} \sum_{s,t=1}^{k} \tilde{\rho}_{st} \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^{2i}} \left( \frac{\tilde{\rho}_{st}}{\sqrt{n}} \right)^i \right]
\]

\[
= \frac{1}{m!} \frac{m^m}{k^2} \sum_{s,t=1}^{k} \tilde{\rho}_{st} \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^{2i}} \left( \frac{\tilde{\rho}_{st}}{\sqrt{n}} \right)^i \right]
\]

\[
= O(n^{m-2}).
\]
Similarly, one gets
\[
\sum_{c(i,m,n)} I[\sigma_{i1} : \sigma_{in}] = \frac{1}{m!} \frac{n^m}{k^{m-1}} + \frac{1}{m!} \frac{m}{k^{m-2}} \sum_{s=1}^{k} \rho_{s0}^2 \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^i} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^i \right]
\]
\[
- \frac{(m/2)}{k^{m-1}} \sum_{s=1}^{m-1} \left( \begin{array}{c} m-1 \\ i \end{array} \right) \frac{1}{k^{(m-1-i)}} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^i + O(n^{m-2}).
\]
and
\[
\sum_{c(i,m,n)} I[\eta_{i1} : \eta_{im}] = \frac{1}{m!} \frac{n^m}{k^{m-1}} + \frac{1}{m!} \frac{m}{k^{m-2}} \sum_{t=1}^{k} \rho_{0t}^2 \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^i} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^i \right]
\]
\[
- \frac{(m/2)}{k^{m-1}} \sum_{t=1}^{m-1} \left( \begin{array}{c} m-1 \\ i \end{array} \right) \frac{1}{k^{(m-1-i)}} \left( \frac{\hat{\rho}_{0t}}{\sqrt{n}} \right)^i + O(n^{m-2}).
\]

Notice that \( \left( \frac{\rho_{st}}{\sqrt{n}} \right) = \frac{n^m}{m!} - \frac{n^m}{m!} \rho_{st0} + O(n^{m-2}) \) and
\[
\frac{n^m}{m!} \left( \frac{(a-b)^2}{k^{2(m-2)}} + \frac{2(a-b)(b-d)}{k^{m-1}} + \frac{(b-d)^2}{k^{m-1}} \right) = \frac{n^m}{m!} \left( \frac{(a-b)^2}{k^{m-1}} + \frac{(b-d)^2}{k^{m-1}} \right) = 0.
\]
\[
\frac{(m/2)^{n-1}}{m!} \left( \frac{k^2(a-b)^2}{k^{2(m-1)}} + \frac{2k(a-b)(b-d)}{k^{m-1}} + \frac{(b-d)^2}{k^{m-1}} \right) = \frac{(m/2)^{n-1}}{m!} \left( \frac{(k-1)^2(a-b)^2}{k^{2(m-1)}} \right).
\]

Let \( c_1 = \frac{(m/2)}{m!} \left( \frac{(a-b)(b-d)}{k^{2(m-2)}} \right) \) and \( c_2 = \frac{(m/2)}{m!} \left( \frac{(a-b)(b-d)}{k^{2(m-2)}} \right). \) Since \( |\hat{\rho}_{st}/\sqrt{n}| \leq 1, |\hat{\rho}_{st}/\sqrt{n}| \leq 1, |\hat{\rho}_{0t}/\sqrt{n}| \leq 1 \) and \( |\hat{\rho}_{0t}/\sqrt{n}| \to 0, |\hat{\rho}_{st}/\sqrt{n}| \to 0, \)
\( |\hat{\rho}_{st}/\sqrt{n}| \to 0 \) in probability. Hence,
\[
\hat{Z}_n = c_2 \sum_{s,t=1}^{k} \rho_{st}^2 \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^i} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^i \right]
\]
\[
+ c_1 \left( \sum_{s,t=1}^{k} \rho_{0t}^2 \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^i} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^i \right] + \sum_{s=1}^{k} \rho_{s0}^2 \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^i} \left( \frac{\hat{\rho}_{0t}}{\sqrt{n}} \right)^i \right] \right)
\]
and \( Z_n = c_2 \sum_{s,t=1}^{k} \rho_{st}^2 + c_1 \left( \sum_{s,t=1}^{k} \rho_{0t}^2 + \sum_{s=1}^{k} \rho_{s0}^2 \right) \) are asymptotically equivalent.

If \( \tau_1(m,k) \leq 1, \) then \( 1 + \sum_{i=1}^{m-2} \frac{1}{k^i} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^i \geq 0, \) hence
\[
\hat{Z}_n \leq c_2 \sum_{s,t=1}^{k} \rho_{st}^2 \left[ 1 + \sum_{i=1}^{m-2} \frac{1}{k^i} \left( \frac{\hat{\rho}_{st}}{\sqrt{n}} \right)^i \right] \leq c_2 \tau_2(m) \sum_{s,t=1}^{k} \rho_{st}^2.
\]

Let \( f_j = \frac{1}{\sqrt{n}} \sum_{u=1}^{j} \left( 1_{[\sigma_{u}=1][\eta_{u}=1]} - \frac{1}{k^2} \right), \ldots, \left( 1_{[\sigma_{u}=k][\eta_{u}=k]} - \frac{1}{k^2} \right) \right)^T \) and \( d_j = f_j - f_{j-1}. \) Then \( ||d_j||^2 = \frac{1}{n} \frac{k^2-1}{k^2} \) and \( b_s^2 = \sum_{j=1}^{n} ||d_j||^2 = \frac{k^2-1}{k^2}. \) By Theorem 3.5 in Pinelis([45]), for any \( t > 0, \)
\[
P \left( \exp \left\{ c_2 \tau_2(m) ||f_n||^2 > t \right\} \right) = P \left( c_2 \tau_2(m) ||f_n||^2 > \log(t) \right) = P \left( ||f_n|| > \frac{\log(t)}{c_2 \tau_2(m)} \right) \leq 2 \exp \left( -\frac{1}{t^{\frac{1}{n(k^2-1)\tau_2(m)}}} \right).
\]

Hence, then condition \( \kappa(k^2-1)\tau_2(m,k) < 1 \) implies that \( \{\exp(\hat{Z}_n)\}_{n=1}^{\infty} \) is uniformly integrable.
By Lemma 4.7 below, we conclude that \( Z_n \) converges to \( \frac{1}{k} \chi^2_{(k-1)^2} \). Since \( \kappa(k-1)\tau_2(m,k) < 1 \) implies \( \kappa(k-1)^2 < 1 \) and \( \frac{c_2}{k^2} \leq \frac{1}{2} \), then it follows that

\[
\begin{align*}
\mathbb{E}_0 Y_n^2 & \rightarrow \exp \left\{ - \left( \frac{m}{m!d} \right) \frac{(k-1)^2(a-b)^2}{k^2(m-1)} \right\} \exp \left\{ \frac{c_2}{k^2} \chi^2_{(k-1)^2} \right\} \\
& = \exp \left\{ - \left( \frac{m}{m!d} \right) \frac{(k-1)^2(a-b)^2}{k^2(m-1)} \right\} \exp \left\{ - \frac{(k-1)^2}{2} \log \left( 1 - \frac{c_2}{k^2} \right) \right\} = \exp \left\{ \sum_{h=2}^{\infty} \lambda_h \delta_h^2 \right\},
\end{align*}
\]

where we used the fact that

\[
\frac{(k-1)^2}{2} \left( \frac{2c_2}{k^2} \right) \frac{1}{h^2} = \frac{(k-1)^2}{2h} \left( a + \frac{(m-1)b}{k^2(m-2)} \right)^h \left( a + \frac{(m-1)b}{k^2(m-2)} \right)^h = \lambda_h \delta_h^2.
\]

Obviously, \( \mathbb{E}_0 Y_n = 1 \). Hence, \( H_0 \) and \( H_1 \) are contiguous.

Let \( \sigma_u = (1_{[\sigma_u=1]}, \ldots, 1_{[\sigma_u=k]})^T, \tau = (1_{[\tau_u=1]}, \ldots, 1_{[\tau_u=k]})^T \). Clearly, \( \sigma_u, \tau_u \sim \text{Multinomial}(1,k,p) \) with \( p = \frac{1}{k} \). Let \( A \) be a \((k^2 + 2k) \times (k^2 + 2k)\) order diagonal matrix, with the first \( 2k \) diagonal elements \( c_1 \), the last \( k \) diagonal elements \( c_2 \). Then \( Z_n = \hat{\rho}A\hat{\rho}^T \). By the central limit theorem, \( \hat{\rho} \) converges to \( N(0,\Sigma) \), where \( \Sigma \) is the covariance matrix of \( (\sigma_u^T, \tau_u^T, \sigma_u^T \otimes \tau_u^T) \).

**Lemma 4.7.** Let \( \sigma_u, \tau_u \sim \text{Multinomial}(1,k,p) \) with \( p = \frac{1}{k} \). The covariance matrix \( \Sigma \) of \( (\sigma_u^T, \tau_u^T, \sigma_u^T \otimes \tau_u^T) \) is given by

\[
\Sigma = \begin{bmatrix}
V & 0 & V \otimes p^T \\
0 & V & p^T \otimes V \\
V \otimes p & p \otimes V & V_2
\end{bmatrix},
\]

where \( V = \text{Var}(\sigma_u) = pI - p^2J, p = E(\sigma_u), V_2 = p^2I_{k^2} - p^4J_{k^2} \), \( J_{k^2} \) is an \( k^2 \times k^2 \) order matrix with all elements 1. Besides, \( V^2 = pV, V_2 = p^2V_2 \). Let

\[
R = \begin{bmatrix}
I & 0 & -I \otimes p^T \\
0 & I & -p^T \otimes I \\
0 & 0 & I_2
\end{bmatrix}, \quad \Lambda = \begin{bmatrix}
V & 0 & 0 \\
0 & V & 0 \\
0 & 0 & \Omega_2
\end{bmatrix}, \quad \Lambda_1 = \begin{bmatrix}
\frac{1}{\sqrt{p}}V & 0 & 0 \\
0 & \frac{1}{\sqrt{p}}V & 0 \\
0 & 0 & \frac{1}{p}\Omega_2
\end{bmatrix}, \quad A = \begin{bmatrix}
c_1I & 0 & 0 \\
0 & c_1I & 0 \\
0 & 0 & c_2I_2
\end{bmatrix}
\]

where \( \Omega_2 = V_2 - p^2V \otimes J - p^2J \otimes V \) with \( \Omega_2^2 = p^2\Omega_2 \). Then \( R^T \Sigma R = \Lambda \) and

\[
R^{-1} = \begin{bmatrix}
I & 0 & I \otimes p^T \\
0 & I & p^T \otimes I \\
0 & 0 & I_2
\end{bmatrix}, \quad \Lambda_1 R^{-1} A (R^{-1})^T \Lambda_1 = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & c_2\Omega_2
\end{bmatrix}.
\]

Hence, \( Z_n \rightarrow c_2p^2\chi^2_{(k-1)^2} \). Furthermore, \( \{\exp(Z_n)\}_n \) is uniformly integrable if \( \kappa(k-1)^2 < 1 \).

**Proof:** We only need to find \( \text{Cov}(\sigma_u, \sigma_u \otimes \tau_u), \text{Cov}(\tau_u, \sigma_u \otimes \tau_u) \) and \( \text{Var}(\sigma_u \otimes \tau_u) \).

\[
\text{Cov}(\sigma_u, \sigma_u \otimes \tau_u) = E[(\sigma_u - p)\sigma_u^T \otimes \tau_u^T]
\]

\[
= E\begin{bmatrix}
(\sigma_1 - p)\sigma_1 \tau_u^T & (\sigma_2 - p)\sigma_2 \tau_u^T & \cdots & (\sigma_1 - p)\sigma_1 \tau_u^T \\
(\sigma_2 - p)\sigma_1 \tau_u^T & (\sigma_2 - p)\sigma_2 \tau_u^T & \cdots & (\sigma_2 - p)\sigma_2 \tau_u^T \\
\vdots & \vdots & \ddots & \vdots \\
(\sigma_s - p)\sigma_1 \tau_u^T & (\sigma_s - p)\sigma_2 \tau_u^T & \cdots & (\sigma_s - p)\sigma_2 \tau_u^T \\
(p - p^2)p^T & -p^2p^T & \cdots & -p^2p^T \\
-p^2p^T & (p - p^2)p^T & \cdots & -p^2p^T \\
\vdots & \vdots & \ddots & \vdots \\
-p^2p^T & -p^2p^T & \cdots & (p - p^2)p^T
\end{bmatrix} = V \otimes \text{p^T}.
\]
Similarly one can get $Cov(\tau_u, \sigma_u \otimes \tau_u) = p^T \otimes V$. The variance of $\sigma_u \otimes \tau_u$ can be calculated as

$$
Cov(\sigma_u \otimes \tau_u, \sigma_u \otimes \tau_u) = E[(\sigma_u \otimes \tau_u - p \otimes p)(\sigma_u \otimes \tau_u - p \otimes p)^T]
$$

$$
= E \begin{bmatrix}
(\sigma_1 \tau_u - p \sigma_u \otimes \tau_u) & \cdots & (\sigma_1 \tau_u - p \sigma_u \otimes \tau_u) \\
(\sigma_2 \tau_u - p \sigma_u \otimes \tau_u) & \cdots & (\sigma_2 \tau_u - p \sigma_u \otimes \tau_u) \\
\vdots & \cdots & \vdots \\
(\sigma_s \tau_u - p \sigma_u \otimes \tau_u) & \cdots & (\sigma_s \tau_u - p \sigma_u \otimes \tau_u)
\end{bmatrix}
$$

$$
= \begin{bmatrix}
p^2 I - p^4 J & -p^4 J & \cdots & -p^4 J \\
- p^4 J & p^2 I - p^4 J & \cdots & -p^4 J \\
\vdots & \vdots & \cdots & \vdots \\
- p^4 J & -p^4 J & \cdots & p^2 I - p^4 J
\end{bmatrix} = p^2 I_{s^2} - p^4 J_{s^2}.
$$

Note that $(I \otimes p)V = V \otimes p$, $V(I \otimes p^T) = V \otimes p^T$, $(p \otimes I)V = p \otimes V$, $V(p^T \otimes I) = p^T \otimes V$. Direct computation yields $R^T \Sigma R = \Lambda$ and

$$
\Lambda_1 R^{-1} A R^{-1}^T \Lambda_1 = \Lambda_1 \begin{bmatrix} I & 0 & I \otimes p^T \end{bmatrix} \begin{bmatrix} c_1 I & 0 & 0 \\
0 & 0 & c_2 I_{s^2} \end{bmatrix} \begin{bmatrix} I & 0 & 0 \end{bmatrix} \Lambda_1
$$

$$
= \Lambda_1 \begin{bmatrix} I_{s^2} & 0 & c_2 I_{s^2} \end{bmatrix} \begin{bmatrix} c_1 I & 0 & 0 \\
0 & 0 & c_2 I_{s^2} \end{bmatrix} \begin{bmatrix} I_{s^2} \end{bmatrix} \Lambda_1
$$

$$
= \Lambda_1 \begin{bmatrix} (c_1 + c_2 p) I & c_2 p^2 J & c_2 I \otimes p^T \\
c_2 p^2 J & (c_1 + c_2 p) I & c_2 P^T \otimes I \\
c_2 I \otimes p & c_2 P \otimes I & c_2 I_{s^2} \end{bmatrix} \Lambda_1
$$

$$
= \begin{bmatrix} \frac{1}{\sqrt{p}} & 0 & 0 \\
0 & \frac{1}{\sqrt{p}} & 0 \\
0 & 0 & \frac{1}{p} \end{bmatrix} \begin{bmatrix} (c_1 + c_2 p) V^2 & c_2 p^2 V^2 J & c_2 V(I \otimes p^T) \Omega_2 \\
c_2 p^2 V^2 J & (c_1 + c_2 p) V^2 & c_2 V(p^T \otimes I) \Omega_2 \\
c_2 \Omega_2(I \otimes p)V & c_2 \Omega_2(p \otimes I)V & c_2 \Omega_2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{p}} & 0 & 0 \\
0 & \frac{1}{\sqrt{p}} & 0 \\
0 & 0 & \frac{1}{p} \end{bmatrix}.
$$

Notice that $V J = JV = 0$,

$$
c_1 + c_2 p = \frac{(m^2)}{m d} \frac{(b - d) (a - b)}{k m^2} + \frac{(m^2)}{m d} \frac{(a - b)^2}{k^2 (m^2 - 2)} = 0,
$$

$$
\Omega_2(I \otimes p)V = (V_2 - p^2 V \otimes J - p^2 J \otimes V)(I \otimes p)V
$$

$$
= V_2(I \otimes p)V - p(V \otimes p)V = p^2 (V \otimes p) - p(V \otimes p)(p I - p^2 J) = 0,
$$

and $V(p^T \otimes I) \Omega_2 = V(I \otimes p^T) \Omega_2 = \Omega_2(p \otimes I)V = 0$, which yields the desired result.

Let $Q = (\Lambda_1 R^{-1})^T$ and $Z \sim N(0, I_{s^2})$. Then the covariance matrix $\Sigma$ can be decomposed as

$$
\Sigma = (R^{-1})^T \Lambda R^{-1} = (\Lambda_1 R^{-1})^T (\Lambda_1 R^{-1}) = QQ^T.
$$

Hence

$$
\tilde{\rho} A \rho^T \rightarrow Z^T Q^T A Q Z = Z^T \Lambda_1 R^{-1} A (R^{-1})^T \Lambda_1 Z = c_2 Z^T \Omega_2 Z.
$$

Note $\Omega_2 = p^2 \Omega_2$ implies the eigenvalues of $\Omega_2$ are either 0 or $p^2$ and

$$
\text{Tr}(\Omega_2) = \text{Tr} \left( V_2 - p^2 V \otimes J - p^2 J \otimes V \right)
$$

$$
= \text{Tr} \left( p^2 I_{k^2} - p^3 I \otimes J - p^3 J \otimes I + p^4 J_{k^2} \right)
$$

$$
= k^2 p^2 - p^3 k^2 - p^3 k^2 + p^4 k^2 = \frac{(k - 1)^2}{k^2}.
$$

Hence $\Omega_2$ has $(k - 1)^2$ eigenvalues $p^2$ with other eigenvalues 0. Then $c_2 Z^T \Omega_2 Z \sim c_2 p^2 \chi^2_{(k-1)^2}$. 
Note that we can rewrite $Z_n$ as
\[
Z_n = \frac{(m)}{m!d} \frac{(a-b)^2}{k^{2(m-2)}} \left( \sum_{s,t} \rho_{st}^2 - \frac{1}{k} \sum_{s=1}^{k} \rho_{s0}^2 + \sum_{t=1}^{k} \rho_{0t}^2 \right)
\]
\[
= \frac{1}{2(m-2)!d} \frac{(a-b)^2}{k^{2(m-2)}} \sum_{s,j=1}^{k} \left( \frac{1}{\sqrt{m}} \sum_{u=1}^{n} (I[\sigma_u = s] - \frac{1}{k}) (I[\eta_u = t] - \frac{1}{k}) \right)^2.
\]

Let $f_j = \frac{1}{\sqrt{m}} \sum_{u=1}^{n} \left( (1[\sigma_u = s] - \frac{1}{k}) (1[\eta_u = t] - \frac{1}{k}) \right) \cdots (1[\sigma_u = s] - \frac{1}{k}) (1[\eta_u = k] - \frac{1}{k}) \right)^T$ and $d_j = f_j - f_{j-1}$. Then
\[
||d_j||^2 = \frac{1}{n} \frac{(k-1)^2}{k^2} \quad \text{and} \quad b_j^2 = \sum_{j=1}^{n} ||d_j||^2 = \frac{(k-1)^2}{k^2}.
\]
By Theorem 3.5 in [45], we have for any $t > 0$,
\[
P\left( \exp \left\{ \frac{1}{2(m-2)!d} \frac{(a-b)^2}{k^{2(m-2)}} ||f_n||^2 > \log(t) \right\} \right)
\]
\[
= P \left( ||f_n|| > \sqrt{\frac{1}{2(m-2)!d} \frac{(a-b)^2}{k^{2(m-2)}}} \log(t) \right)
\]
\[
\leq 2 \exp \left( -\frac{\log(t)}{\kappa(k-1)^2} \right) = 2t^{-\frac{1}{\kappa(k-1)^2}}.
\]
Hence, if $\kappa(k-1)^2 < 1$, $\{\exp(Z_n)\}_{n=1}^{\infty}$ is uniformly integrable.

\[\square\]

4.4. Proof of Proposition 2.5

In the proof of this section and the next section, for convenience, we denote $a_1 = \frac{a_n}{n^{m-3}}$ and $b_1 = \frac{b_n}{n^{m-3}}$.

**Proof of Proposition 2.5:** Under $H_0$, we have $a_1 = b_1$, and then
\[
\mathcal{T} = (E\mathcal{W}_1)^{3(m-2)} \left( b_1^3 - \left( \frac{b_1^2}{b_1} \right)^3 \right) = 0.
\]
Under $H_1$, $k \geq 2$ and $a_1 > b_1$. For $l = 1$, direct computation yields
\[
\mathcal{T} = (E\mathcal{W}_1)^{3(m-2)} \frac{(k-1)(a_1-b_1)^3}{k^{3(m-1)}} \neq 0.
\]
Next we assume $l \geq 2$, let $E_1 = (E\mathcal{W}_1)^{-m} E$, $V_1 = (E\mathcal{W}_1)^{-2(m-1)} V$ and $T_1 = (E\mathcal{W}_1)^{-3(m-2)} T$. Then
\[
\mathcal{T} = (E\mathcal{W}_1)^{3(m-2)} \left( T_1 - \left( \frac{V_1}{E_1} \right)^3 \right).
\]
We calculate $T_1 E_1^3 - V_1^3$ to get the following
\[
T_1 E_1^3 - V_1^3 = (a_1 - b_1)^6 \frac{1-k^{-1}}{k^{6(m-3/-2)}} + 3(a_1 - b_1)^5 b_1 \left( \frac{k^l - 2}{k^{3(m-3/-l-2)}} + \frac{1}{k^{3(m-3/-l-4)}} \right)
\]
\[
+ 3(a_1 - b_1)^4 b_1^2 \left( \frac{k^l - 1 - k^{-2l+1}}{k^{3(m-3/-l-2)}} + \frac{1}{k^{4(m-1/-2)}} \right)
\]
\[
+ (a_1 - b_1)^3 b_1^3 \left( \frac{1-3k^{-2l+1}}{k^{3(m-3/-l-2)}} + \frac{2}{k^{3(m-3/-l-4)}} \right).
\]
Clearly, if $k \geq 2$, $a_1 > b_1 > 0$ and $l \geq 2$, each term in the right hand side of (30) is positive, which implies that $T_1 E_1^3 - V_1^3 > 0$ and hence $\mathcal{T} \neq 0$.

\[\square\]
4.5. Proof of Theorem 2.6 and Theorem 2.7

Proof of Theorem 2.6: It’s easy to have the following expansion
\[
\hat{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^3 = T - \left( \frac{V}{E} \right)^3 + (\hat{T} - T)
\]
\[
+ \left( \frac{V}{E} - \frac{\hat{V}}{\hat{E}} \right)^3 - 3 \left( \frac{V}{E} \right)^2 \left( \frac{\hat{V}}{\hat{E}} \right) + 3 \left( \frac{V}{E} \right)^2 \left( \frac{1}{E} - \frac{1}{\hat{E}} \right) (V - \hat{V})
\]
\[-3 \left( \frac{V}{E} \right)^2 \left( \frac{1}{E} - \frac{1}{\hat{E}} \right) V - 3 \left( \frac{V}{E} \right)^2 \left( \frac{1}{E} - \frac{1}{\hat{E}} \right) (V - \hat{V}).
\]

(31)

By Lemma 4.8 below, the first two terms in (31) are the leading terms and hence we have
\[
\sqrt{\left( \frac{n}{3(m-l)} \right) (m-l)} \left( \hat{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^3 \right)
\]
\[
\sqrt{T} \rightarrow N(0, 1).
\]

Since \( \hat{T} = T + o_P(1) \), then
\[
\sqrt{\left( \frac{n}{3(m-l)} \right) (m-l)} \left( \hat{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^3 \right) - \delta \rightarrow N(0, 1),
\]
which completes the proof.

Proof of Theorem 2.7: We rewrite the statistic as
\[
2 \sqrt{\left( \frac{n}{3(m-l)} \right) (m-l)} \left( \sqrt{T} - \left( \frac{\hat{V}}{\hat{E}} \right)^3 \right)
\]
\[
= 2 \sqrt{\left( \frac{n}{3(m-l)} \right) (m-l)} \left( T - \left( \frac{V}{E} \right)^3 \right)
\]
\[
+ 2 \sqrt{\left( \frac{n}{3(m-l)} \right) (m-l)} \left( \frac{\hat{T} - T}{\sqrt{T} + \left( \frac{\hat{V}}{\hat{E}} \right)^3} \right) + o_P(1).
\]

The first term is of the same order as \( \delta \), while the second term is bounded in probability. Hence, we get the desired result.

Lemma 4.8. Under the condition of Theorem 2.6, we have
\[
\mathbb{E}(\hat{E} - E)^2 = O\left( \frac{a_1^2}{n} \right),
\]
\[
\mathbb{E}(\hat{V} - V)^2 = O\left( \frac{a_1^4}{n} \right),
\]
\[
\mathbb{E}(\hat{T} - T)^2 = O\left( \frac{a_1^3}{n^{3(m-l)}} \right),
\]
\[
\sqrt{\left( \frac{n}{3(m-l)} \right) (m-l)} \left( \hat{T} - T \right)
\]
\[
\sqrt{T} \rightarrow N(0, 1).
\]

(32) (33) (34) (35)
To get the asymptotic distribution, we need the martingale central limit theorem in Hall and Heyde([31]).

**Theorem 4.9** (Hall and Heyde (2014)). Suppose that for every \( n \in \mathbb{N} \) and \( k_n \to \infty \) the random variables \( X_{n,1}, \ldots, X_{n,k_n} \) are a martingale difference sequence relative to an arbitrary filtration \( \mathcal{F}_{n,1} \subset \mathcal{F}_{n,2} \subset \ldots \mathcal{F}_{n,k_n} \). If (1) \( \sum_{i=1}^{k_n} \mathbb{E}(X_{n,i}^2 | \mathcal{F}_{n,i-1}) \to 1 \) in probability, (2) \( \sum_{i=1}^{k_n} \mathbb{E}(X_{n,i}^2 I[|X_{n,i}| > \epsilon] | \mathcal{F}_{n,i-1}) \to 0 \) in probability for every \( \epsilon > 0 \), then \( \sum_{i=1}^{k_n} X_{n,i} \to N(0,1) \) in distribution.

**Proof of Lemma 4.8:** Let \( W_{i_1:i_m} = W_{i_1}W_{i_2} \ldots W_{i_m} \), \( \eta_{i_1:i_m} = (a_1 - b_1)I[\sigma_{i_1} = \sigma_{i_2} = \ldots = \sigma_{i_m}] + b_1 \) and \( \theta_{i_1:i_m} = \eta_{i_1:i_m}W_{i_1:i_m} \). Clearly \( \mathbb{E}(A_{i_1:i_m} | W, \sigma) = \theta_{i_1:i_m} \).

Firstly, we show equation (32). Write \( \hat{E} - E \) as

\[
\hat{E} - E = \left( \hat{E} - \mathbb{E}(\hat{E} | W, \sigma) \right) + \left( \mathbb{E}(\hat{E} | W, \sigma) - \mathbb{E}(\hat{E} | \sigma) \right) + \left( \mathbb{E}(\hat{E} | \sigma) - E \right).
\]

Note that the three terms in the right hand side are mutually uncorrelated. Hence

\[
\mathbb{E}(\hat{E} - E)^2 = \mathbb{E}(\hat{E} - \mathbb{E}(\hat{E} | W, \sigma))^2 + \mathbb{E}(\mathbb{E}(\hat{E} | W, \sigma) - \mathbb{E}(\hat{E} | \sigma))^2 + \mathbb{E}(\mathbb{E}(\hat{E} | \sigma) - E)^2. \tag{36}
\]

It’s easy to check that \( A_{i_1:i_m} \) and \( A_{j_1:j_m} \) are conditionally independent if \( i_1 : i_m \neq j_1 : j_m \). For the first term, we have

\[
\mathbb{E}(\hat{E} - \mathbb{E}(\hat{E} | W, \sigma))^2 = \mathbb{E}\left( \frac{1}{n_m} \sum_{c(i,m,n)} (A_{i_1:i_m} - \theta_{i_1:i_m}) \right)^2
\]

\[
= \frac{1}{n_m^2} \sum_{c(i,m,n)c(j,m,n)} \mathbb{E}(A_{i_1:i_m} - \theta_{i_1:i_m}) (A_{j_1:j_m} - \theta_{j_1:j_m})
\]

\[
= \frac{1}{n_m^2} \sum_{c(i,m,n)} \mathbb{E}(A_{i_1:i_m} - \theta_{i_1:i_m})^2
\]

\[
= \frac{1}{n_m^2} \sum_{c(i,m,n)} \mathbb{E} \theta_{i_1:i_m} (1 - \theta_{i_1:i_m})
\]

\[
\leq \frac{1}{n_m^2} \sum_{c(i,m,n)} \mathbb{E} \theta_{i_1:i_m}
\]

\[
= \frac{1}{n_m^2} \sum_{c(i,m,n)} (\mathbb{E} W_1)^m (a_1 + (k^{m-1} - 1)b_1)
\]

\[
= \frac{(\mathbb{E} W_1)^m}{n_m} (a_1 + (k^{m-1} - 1)b_1) = O \left( \frac{a_1}{n_m} \right). \tag{37}
\]

For the third term in (36), one has

\[
\mathbb{E}(\mathbb{E}(\hat{E} | \sigma) - E)^2
\]

\[
= \mathbb{E}\left( \frac{1}{n_m} \sum_{c(i,m,n)} (\mathbb{E} W_1)^m (\eta_{i_1:i_m} - \mathbb{E} \eta_{i_1:i_m}) \right)^2
\]

\[
= (\mathbb{E} W_1)^{2m} \mathbb{E}\left( \frac{1}{n_m} \sum_{c(i,m,n)} (a_1 - b_1) (I[\sigma_{i_1} : \sigma_{i_m}] - P[\sigma_{i_1} : \sigma_{i_m}]) \right)^2
\]

\[
\leq (\mathbb{E} W_1)^{2m} 2(a_1^2 + b_1^2) \mathbb{E}\left( \frac{1}{n_m} \sum_{c(i,m,n)} (I[\sigma_{i_1} : \sigma_{i_m}] - P[\sigma_{i_1} : \sigma_{i_m}])^2 \right). \tag{38}
\]
Note that
\[ E\left(\frac{1}{\binom{n}{m}} \sum_{c(i,m,n)} (I[\sigma_{i_1} : \sigma_{i_m}] - \mathbb{P}[\sigma_{i_1} : \sigma_{i_m}])^2 \right) = \frac{1}{\binom{n}{m}^2} \sum_{c(i,m,n);c(j,m,n)} E[I[\sigma_{i_1} : \sigma_{i_m}] - \mathbb{P}[\sigma_{i_1} : \sigma_{i_m}]) (I[\sigma_{j_1} : \sigma_{j_m}] - \mathbb{P}[\sigma_{j_1} : \sigma_{j_m}]) \right) \] (39)

If there is no repeated index in \(i_1 : i_m\) and \(j_1 : j_m\), then
\[ E(I[\sigma_{i_1} : \sigma_{i_m}] - \mathbb{P}[\sigma_{i_1} : \sigma_{i_m}]) (I[\sigma_{j_1} : \sigma_{j_m}] - \mathbb{P}[\sigma_{j_1} : \sigma_{j_m}]) = 0. \]

If there is only one repeated index in \(i_1 : i_m\) and \(j_1 : j_m\), say, \(i_1 = j_1\) and other indexes are different, then
\[ E(I[\sigma_{i_1} : \sigma_{i_m}] - \mathbb{P}[\sigma_{i_1} : \sigma_{i_m}]) (I[\sigma_{j_1} : \sigma_{j_m}] - \mathbb{P}[\sigma_{j_1} : \sigma_{j_m}]) = -\frac{k}{k^2m-1} - 2\frac{k}{km-1} + \frac{1}{k^2(m-1)} = 0. \]

If there are two or more indexes in \(i_1 : i_m\) and \(j_1 : j_m\) are the same, it is easy to verify that
\[ E(I[\sigma_{i_1} : \sigma_{i_m}] - \mathbb{P}[\sigma_{i_1} : \sigma_{i_m}]) (I[\sigma_{j_1} : \sigma_{j_m}] - \mathbb{P}[\sigma_{j_1} : \sigma_{j_m}]) \neq 0. \]

The leading term in (39) is when there are exactly two indexes in \(i_1 : i_m\) and \(j_1 : j_m\) are the same. Hence, by (38) and (39), we have
\[ E\left(E(\hat{E}|\sigma) - E\right)^2 = O\left((a_1^2 + b_1^2) \frac{1}{\binom{n}{m}^2} \binom{n}{m} \binom{n}{m-2}\right) = O\left(\frac{a_1^2}{n^2}\right). \] (40)

For the second term in (36), we have
\[ E\left(E(\hat{E}|W,\sigma) - E(\hat{E}|\sigma)\right)^2 = E\left(\frac{1}{\binom{n}{m}} \sum_{c(i,m,n)} \eta_{i_1:i_m}(W_{i_1:i_m} - EW_{i_1:i_m})\right)^2. \] (41)

Note that for some constants \(c_{s_1}, c_{s_1s_2}, \ldots, c_{s_1s_{m-1}}\) dependent on \(EW_1, 1 \leq s_1, \ldots, s_{m-1} \leq m\), one has
\[ W_{i_1:i_m} - EW_{i_1:i_m} = \sum_{s_1=1}^{m} c_{s_1}(W_{i_1} - EW_{i_1}) + \sum_{1 \leq s_1 \neq s_2 \leq m} c_{s_1s_2}(W_{i_1} - EW_{i_1})(W_{i_2} - EW_{i_2}) + \cdots + (W_{i_1} - EW_{i_1})(W_{i_2} - EW_{i_2}) \cdots (W_{i_m} - EW_{i_m}). \] (42)

Clearly, the summation terms in (42) are mutually uncorrelated. And for \(W_{i_1} - EW_{i_1}\), we have
\[ E\left(\frac{1}{\binom{n}{m}} \sum_{c(i,m,n)} \eta_{i_1:i_m}(W_{i_1} - EW_{i_1})\right)^2 = \frac{1}{\binom{n}{m}^2} \sum_{c(i,m,n);c(j,m,n)} E\left(\eta_{i_1:i_m} \eta_{j_1:j_m}(W_{i_1} - EW_{i_1})(W_{j_1} - EW_{j_1})\right) \]
\[ = \frac{1}{\binom{n}{m}^2} O\left(\frac{a_1^2}{n}\right) \binom{n}{m} \binom{n}{m-1} = O\left(\frac{a_1^2}{n}\right). \] (43)

It’s easy to verify that the terms \(\prod_{s=1}^{t}(W_{i_s} - EW_{i_s})(t \geq 2)\) are of higher order. By equation (41),
\[ E\left(E(\hat{E}|W,\sigma) - E(\hat{E}|\sigma)\right)^2 = O\left(\frac{a_1^2}{n}\right). \] (44)

Combining (37), (40) and (44) yields (32).

Next we prove (33). We can similarly decompose the mean square as
\[ E(\tilde{V} - V)^2 = E\left(\tilde{V} - E(\tilde{V}|W,\sigma)\right)^2 + E\left(E(\tilde{V}|W,\sigma) - E(\tilde{V}|\sigma)\right)^2 + E\left(E(\tilde{V}|\sigma) - V\right)^2. \] (45)
Firstly we have the following decomposition

\[ A_{i_1:i_m}A_{i_{m-l+1}:i_{2m-l}} - \theta_{i_1:i_m}\theta_{i_{m-l+1}:i_{2m-l}} = (A_{i_1:i_m} - \theta_{i_1:i_m})(A_{i_{m-l+1}:i_{2m-l}} - \theta_{i_{m-l+1}:i_{2m-l}}) + (A_{i_1:i_m} - \theta_{i_1:i_m})\theta_{i_{m-l+1}:i_{2m-l}} + \theta_{i_1:i_m}(A_{i_{m-l+1}:i_{2m-l}} - \theta_{i_{m-l+1}:i_{2m-l}}), \]

from which it follows

\[
\begin{align*}
\hat{V} - \mathbb{E}(\hat{V}|W, \sigma) &= \frac{1}{(2m-l)} \sum_{c(i,m,n)} C_{2m-l}(A) - C_{2m-l}(\theta) \\
&= \frac{1}{(2m-l)} \sum_{c(i,m,n)} C_{2m-l}(A - \theta) + \frac{1}{(2m-l)} \sum_{c(i,m,n)} C_{2m-l}(A - \theta, \theta) + C_{2m-l}(\theta, A - \theta).
\end{align*}
\]

In the last equation of (46), the first summation and the second summation are uncorrelated. Hence

\[
\mathbb{E}\left(\left(\hat{V} - \mathbb{E}(\hat{V}|W, \sigma)\right)^2\right) = \mathbb{E}\left(\left(\frac{1}{(2m-l)} \sum_{c(i,m,n)} C_{2m-l}(A - \theta)\right)^2\right) + \mathbb{E}\left(\frac{1}{(2m-l)} \sum_{c(i,m,n)} C_{2m-l}(A - \theta, \theta) + C_{2m-l}(\theta, A - \theta)\right)^2.
\]

The terms in \( C_{2m-l}(A - \theta) \) are also uncorrelated and

\[
\begin{align*}
\mathbb{E}\left(\frac{1}{(2m-l)} \sum_{c(i,m,n)} (A_{i_1:i_m} - \theta_{i_1:i_m})(A_{i_{m-l+1}:i_{2m-l}} - \theta_{i_{m-l+1}:i_{2m-l}})^2\right) &= \frac{1}{(2m-l)^2} \sum_{c(i,m,n)} \mathbb{E}(A_{i_1:i_m} - \theta_{i_1:i_m})^2(A_{i_{m-l+1}:i_{2m-l}} - \theta_{i_{m-l+1}:i_{2m-l}})^2 (2m-l)^2 \\
&= \frac{1}{(2m-l)^2} O\left(\frac{a_1^2}{n}\right) = O\left(\frac{a_1^2}{n^2m-l}\right),
\end{align*}
\]

which is the order of the first term in (47). For the second summand term in (47), one has

\[
\begin{align*}
\mathbb{E}\left(\frac{1}{(2m-l)} \sum_{c(i,m,n)} (A_{i_1:i_m} - \theta_{i_1:i_m})\theta_{i_{m-l+1}:i_{2m-l}}\right)^2 &= \frac{1}{(2m-l)^2} \sum_{c(i,m,n)} \mathbb{E}(A_{i_1:i_m} - \theta_{i_1:i_m})^2\theta_{i_{m-l+1}:i_{2m-l}}^2 (2m-l)^2 \\
&= \frac{1}{(2m-l)^2} O\left(a_1^2\left(\frac{n}{2m-l}\right)\right) = O\left(\frac{a_1^2}{n^2m-l}\right).
\end{align*}
\]

Hence, it follows from (48) and (49) that

\[
\mathbb{E}\left(\hat{V} - \mathbb{E}(\hat{V}|W, \sigma)\right)^2 = O\left(\frac{a_1^2}{n^2m-l}\right).
\]

For middle term in (45), by definition, it’s equal to

\[
\mathbb{E}\left(\mathbb{E}(\hat{V}|W, \sigma) - \hat{V}\right)^2 = \mathbb{E}\left(\frac{1}{(2m-l)} \sum_{c(i,2m-l,m)} C_{2m-l}(\theta) - \mathbb{E}(C_{2m-l}(\theta)|\sigma)\right)^2.
\]
The first term in $C_{2m-l}(\theta) - \mathbb{E}(C_{2m-l}(\theta)|\sigma)$ is
\[
(W_{i_1:i_{m-l}}^2 W_{i_{m-l+1}:i_{m}} W_{i_{m+1}:i_{2m-l}} - (\mathbb{E}W_{i_1}^2)^{l}(\mathbb{E}W_{i_1})^{2(m-l)} \eta_{i_1:i_m \eta_{m-l+1:2m-l}}.
\]
and all the terms in it are uncorrelated. Let $\delta_s = 2$ if $s = m-l+1, \ldots, m$ and $\delta_s = 1$ otherwise. For constants $c_{s_1}, c_{s_1 s_2}, \ldots, c_{s_1 \ldots s_{2m-l-1}}$, the following expansion is true.
\[
W_{i_1:i_{m-l}} W_{i_{m-l+1}:i_{m}} W_{i_{m+1}:i_{2m-l}} - (\mathbb{E}W_{i_1}^2)^{l}(\mathbb{E}W_{i_1})^{2(m-l)}
= \sum_{s_1=1}^{2m-l} c_{s_1} (W_{i_{s_1}} - \mathbb{E}W_{i_{s_1}}) + \sum_{1 \leq s_1 \neq s_2 \leq 2m-l} c_{s_1 s_2} (W_{i_{s_1}} - \mathbb{E}W_{i_{s_1}})(W_{i_{s_2}} - \mathbb{E}W_{i_{s_2}})
+ \cdots + \prod_{s_1=1}^{2m-l} (W_{i_{s_1}} - \mathbb{E}W_{i_{s_1}})
\]
(51)

Clearly, the three summation terms in (51) are mutually uncorrelated. For any $s_1$,
\[
\mathbb{E}\left( \frac{1}{n} \sum_{c(i,2m-l,n)} \frac{(W_{i_{s_1}} - \mathbb{E}W_{i_{s_1}})(\eta_{i_1:i_m \eta_{m-l+1:2m-l}}}{2m-l} \right)^2
= \frac{1}{n^{2m-l}c(i,2m-l,n)} \mathbb{E}(W_{i_{s_1}} - \mathbb{E}W_{i_{s_1}})^2 \mathbb{E}(W_{i_{s_1}} - \mathbb{E}W_{i_{s_1}})^2 \left( \frac{n}{2m-l} \right) \left( \frac{n}{2m-l-1} \right) = O\left( \frac{a^4}{n} \right).
\]

It’s easy to verify that the product terms of $W_{i_{s_1}} - \mathbb{E}W_{i_{s_1}}$ are of higher order. Hence
\[
\mathbb{E}\left( \left( \mathbb{E}(\tilde{W}|W, \sigma) - \mathbb{E}(\tilde{W}|\sigma) \right)^2 \right) = O\left( \frac{a^4}{n} \right).
\]
(52)

The last term in (45) can be expressed as
\[
\mathbb{E}\left( \left( \mathbb{E}(\tilde{W}|\sigma) - \mathbb{V} \right)^2 \right) = \mathbb{V} \left( \frac{1}{n^{2m-l}c(i,2m-l,n)} \right)^2 \sum_{c(i,2m-l,n)} \frac{C_{2m-l}(\eta)}{2m-l}
= O\left( \mathbb{V} \left( \frac{1}{n^{2m-l}c(i,2m-l,n)} \right)^2 \sum_{c(i,2m-l,n)} \frac{\eta_{i_1:i_m \eta_{m-l+1:2m-l}}}{2m-l} \right).
\]
(53)

To find the variance, let $H \subset [k]^{2m-l}$. We have
\[
\mathbb{E}\left( \sum_{c(i,2m-l,n)} \sum_{(h_{s_1}) \in H} \left( \prod_{s=1}^{2m-l} I[\sigma_{s_1} = h_{s_1}] - \mathbb{E} \prod_{s=1}^{2m-l} I[\sigma_{s_1} = h_{s_1}] \right)^2 \right)
\leq |H| \sum_{(h_{s_1}) \in H} \mathbb{E} \left( \sum_{c(i,2m-l,n)} \left( \prod_{s=1}^{2m-l} I[\sigma_{s_1} = h_{s_1}] - \mathbb{E} \prod_{s=1}^{2m-l} I[\sigma_{s_1} = h_{s_1}] \right)^2 \right).
\]
Since
\[
\prod_{s=1}^{2m-l} I[\sigma_{s} = h_{s}] - E \prod_{s=1}^{2m-l} I[\sigma_{s} = h_{s}]
= \sum_{s_1=1}^{2m-l} c_{s_1} \left(I[\sigma_{s_1} = h_{s_1}] - E I[\sigma_{s_1} = h_{s_1}]\right) \\
+ \sum_{s_1,s_2=1}^{2m-l} c_{s_1,s_2} \left(I[\sigma_{s_1} = h_{s_1}] - E I[\sigma_{s_1} = h_{s_1}]\right) \left(I[\sigma_{s_2} = h_{s_2}] - E I[\sigma_{s_2} = h_{s_2}]\right) \\
+ \cdots + \prod_{s=1}^{2m-l} \left(I[\sigma_{s} = h_{s}] - E I[\sigma_{s} = h_{s}]\right),
\]
and
\[
E\left( \sum_{c(i,2m-l,n)} \left(I[\sigma_{i} = h_{i}] - E I[\sigma_{i} = h_{i}]\right) \right)^2 \\
= \sum_{c(i,2m-l,n) \neq (j,2m-l,n)} E\left( I[\sigma_{i} = h_{i}] - E I[\sigma_{i} = h_{i}] \right) \left( I[\sigma_{j} = h_{j}] - E I[\sigma_{j} = h_{j}] \right) \\
= O\left(n^{2(2m-l)-1}\right),
\]
then
\[
E\left( \sum_{c(i,2m-l,n)} \sum_{(h_{s}) \in H} \left( \prod_{s=1}^{2m-l} I[\sigma_{s} = h_{s}] - E \prod_{s=1}^{2m-l} I[\sigma_{s} = h_{s}] \right) \right)^2 = O\left(n^{2(2m-l)-1}\right). \tag{54}
\]
Note that
\[
\eta_{i_{1}:m} \eta_{m-i_{1}:m} = (a_1 - b_1)^2 I[\sigma_{i_1} : \sigma_{2m-l}] + (a_1 - b_1) b_1 I[\sigma_{i_1} : \sigma_m] \\
+ (a_1 - b_1) b_1 I[\sigma_{m-i_1+1} : \sigma_{2m-l}] + b_1^2.
\]
Then by (54) we have
\[
\text{Var}\left( \frac{1}{n} \sum_{c(i,2m-l,n)} \eta_{i_{1}:m} \eta_{m-i_{1}:m} \right) \left[ (2m-l) \right] \\
\times \text{Var}\left( \frac{1}{n} \sum_{c(i,2m-l,n)} (a_1 - b_1)^2 I[\sigma_{i_1} : \sigma_{2m-l}] \right) \left[ (2m-l) \right] \\
+ \text{Var}\left( \frac{1}{n} \sum_{c(i,2m-l,n)} (a_1 - b_1) b_1 I[\sigma_{i_1} : \sigma_m] \right) \left[ (2m-l) \right] \\
+ \text{Var}\left( \frac{1}{n} \sum_{c(i,2m-l,n)} (a_1 - b_1) b_1 I[\sigma_{m-i_1+1} : \sigma_{2m-l}] \right) \left[ (2m-l) \right] \\
\times \frac{a_1^4}{n} \left[ n^{2(2m-l)-1} \right] + \frac{a_1^4}{n} \left[ n^{2(2m-l)-1} \right] + \frac{a_1^4}{n} \left[ n^{2(2m-l)-1} \right] = O\left( \frac{a_1^4}{n} \right). \tag{55}
\]
By (53) and (55), one gets
\[
E\left( E(V|\sigma) - V \right)^2 = O\left( \frac{a_1^4}{n} \right). \tag{56}
\]
From (50), (52), (56) and the condition \( n^{l-1} \ll a_n \ll b_n \), we conclude (33).
In the following, we prove (34). Similar to the previous proof, we have
\[
\hat{T} - T = \left( \hat{T} - \mathbb{E}(\hat{T}|W, \sigma) \right) + \left( \mathbb{E}(\hat{T}|W, \sigma) - \mathbb{E}(T|\sigma) \right) + \left( \mathbb{E}(T|\sigma) - T \right),
\]
and
\[
\mathbb{E}(\hat{T} - T)^2 = \mathbb{E}(\hat{T} - \mathbb{E}(\hat{T}|W, \sigma))^2 + \mathbb{E}(\mathbb{E}(\hat{T}|W, \sigma) - \mathbb{E}(T|\sigma))^2 + \mathbb{E}(\mathbb{E}(T|\sigma) - T)^2. \tag{57}
\]
For the first expectation, one has
\[
\mathbb{E}(\mathbb{E}(\hat{T}|W, \sigma) - \mathbb{E}(T|\sigma))^2 = \mathbb{E}\left( \frac{1}{n} \sum_{i(3(m-l))} C_{3(m-l)}(\theta) - EC_{3(m-l)}(\theta) \right)^2.
\]
The first term in \(C_{3(m-l)}(\theta) - EC_{3(m-l)}(\theta)\) is
\[
\eta_{1:i_m} \eta_{m-1:2m-l} \eta_{2m-2l-1:3(m-l)} \delta_{i:i_l}
\times \left( W_{i_1:i_m} W_{i_{m-1}:i_{2m-l}} W_{i_{2m-2l-1:3(m-l)}:i_{1:i_l}} - \mathbb{E}W_{i_1:i_m} W_{i_{m-1}:i_{2m-l}} W_{i_{2m-2l-1:3(m-l)}:i_{1:i_l}} \right),
\]
and the \(m-1\) terms are uncorrelated. Let \(\delta_s = 2\) if \(s = m-l+1, \ldots, m\) or \(s = 2m-2l+1, \ldots, 2m-l\) and \(\delta_s = 1\) otherwise. Then following decomposition holds.
\[
W_{i_1:i_m} W_{i_{m-1}:i_{2m-l}} W_{i_{2m-2l-1:3(m-l)}:i_{1:i_l}} = \sum_{s_1=1}^{3(m-l)} c_{s_1}(W_{i_{s_1}}^\delta_{s_1} - \mathbb{E}W_{i_{s_1}}^\delta_{s_1}) + \sum_{s_1, s_2=1}^{3(m-l)} c_{s_1 s_2}(W_{i_{s_1}}^\delta_{s_1} - \mathbb{E}W_{i_{s_1}}^\delta_{s_1})(W_{i_{s_2}}^\delta_{s_2} - \mathbb{E}W_{i_{s_2}}^\delta_{s_2})
+ \cdots + \prod_{s_1=1}^{3(m-l)} (W_{i_{s_1}}^\delta_{s_1} - \mathbb{E}W_{i_{s_1}}^\delta_{s_1}).
\]
Note that
\[
\mathbb{E}\left( \frac{1}{n} \sum_{i(3(m-l))} \eta_{1:i_m} \eta_{m-1:2m-l} \eta_{2m-2l-1:3(m-l)} \delta_{i:i_l} (W_{i_{s_1}}^\delta_{s_1} - \mathbb{E}W_{i_{s_1}}^\delta_{s_1}) \right)^2 = \frac{1}{n} \sum_{i(3(m-l)), c(j,3(m-l),n)} \frac{O(a)^6}{m-l} \mathbb{E}(W_{i_{s_1}}^\delta_{s_1} - \mathbb{E}W_{i_{s_1}}^\delta_{s_1}) (W_{i_{s_2}}^\delta_{s_2} - \mathbb{E}W_{i_{s_2}}^\delta_{s_2}) = O\left( \frac{a^6}{n} \right),
\]
and the product terms of \(W_{i_{s_1}}^\delta_{s_1} - \mathbb{E}W_{i_{s_1}}^\delta_{s_1}\) are of higher order. Hence,
\[
\mathbb{E}(\mathbb{E}(\hat{T}|W, \sigma) - \mathbb{E}(T|\sigma))^2 = O\left( \frac{a^6}{n} \right). \tag{58}
\]
For the third expectation in (57), similar to (53), one has
\[
\mathbb{E}(\mathbb{E}(\hat{T}|\sigma) - T)^2 = Var\left( \frac{1}{n} \sum_{i(3(m-l))} \eta_{1:i_m} \eta_{m-1:2m-l} \eta_{2m-2l-1:3(m-l)} (W_{i_{s_1}}^\delta_{s_1} - \mathbb{E}W_{i_{s_1}}^\delta_{s_1}) \right) = O\left( \frac{a^6}{n} \right). \tag{59}
\]
For the second expectation in (57), notice that
\[
\hat{T} - \mathbb{E}(\hat{T}|W, \sigma) = \frac{1}{n} \sum_{i(3(m-l))} \frac{C_{3(m-l)}(\theta) - C_{3(m-l)}(\theta)}{m-l}.
\]
The $m - l$ terms in $C_{3(m-l)}(A) - C_{3(m-l)}(\theta)$ are uncorrelated and the first term in it can be decomposed as

\[ A_{i_1;i_m}A_{i_m+1;i_{2m-1}}A_{i_{2m-2l-1};A} = (A_{i_1;i_m} - \theta_{i_1;i_m})\theta_{i_m+1;i_{2m-1}} - \theta_{i_m+1;i_{2m-1}}\theta_{i_1;i_m} + (A_{i_{2m-2l-1};A} - \theta_{i_{2m-2l-1};A})\theta_{i_1;i_m} + \ldots \]

Note that

\[
\mathbb{E}\left(\frac{1}{n^{3(m-l)}} \sum_{c(i,3(m-l),n)} (A_{i_1;i_m} - \theta_{i_1;i_m})\theta_{i_m+1;i_{2m-1}} - \theta_{i_m+1;i_{2m-1}}\theta_{i_1;i_m})\theta_{i_2m-2l-1;i_{3(m-l)}i_1;i_l} \right)^2 = O\left(\frac{a_1^5}{n}\right), \tag{60}
\]

and

\[
\mathbb{E}\left(\frac{1}{n^{3(m-l)}} \sum_{c(i,3(m-l),n)} (A_{i_1;i_m} - \theta_{i_1;i_m})\theta_{i_m+1;i_{2m-1}} - \theta_{i_m+1;i_{2m-1}}\theta_{i_1;i_m})\theta_{i_2m-2l-1;i_{3(m-l)}i_1;i_l} \right)^2 = O\left(\frac{a_1^5}{n}\right). \tag{61}
\]

Let

\[ G_{i_1;i_{3(m-l)}} = (A_{i_1;i_m} - \theta_{i_1;i_m})\theta_{i_m+1;i_{2m-1}} - \theta_{i_m+1;i_{2m-1}}\theta_{i_1;i_m} \]

Then

\[
\mathbb{E}\left(\frac{1}{n^{3(m-l)}} \sum_{c(i,3(m-l),n)} G_{i_1;i_{3(m-l)}} \right)^2 = \frac{1}{n^{3(m-l)}} \sum_{c(i,3(m-l),n)} \mathbb{E}G_{i_1;i_{3(m-l)}}^2 \leq \frac{1}{n^{3(m-l)}} \sum_{c(i,3(m-l),n)} \mathbb{E}G_{i_1;i_{3(m-l)}}^2 = \mathcal{O}\left(\frac{a_1^3}{n^{3(m-l)}}\right). \tag{62}
\]

Under the condition $a_n \asymp b_n \ll n^{\frac{3}{2}}$, by (57), (58), (59), (60), (61) and (62), we get (34).

In the end, we show the asymptotic normality by using Theorem 4.9. Let

\[ \mathcal{W}_n = \left\{ \left| \frac{1}{n} \sum_{i=1}^{n} W_i^2 - EW_1^2 \right| \leq n^{-\frac{3}{4}} \right\}, \quad \Theta_n = \sqrt{\mathbb{E}\left( \sum_{c(i,3(m-l),n)} G_{i_1;i_{3(m-l)}} \right)^2} \]

Clearly, $\Theta_n \asymp n^{\frac{3}{2}} a_1^2 \rightarrow \infty$ if $n^{-1} \ll a_n \ll b_n$. Define

\[ S_{n,t} = \frac{\sum_{c(i,3(m-l),t)} G_{i_1;i_{3(m-l)}}}{\Theta_n}, \]

and let $X_{n,t} = S_{n,t} - S_{n,t-1}$. We show the asymptotic normality by applying the martingale central limit theorem to $X_{n,t}$ conditioning on $W$ and $\sigma$. Simple calculation yields that

\[ X_{n,t} = \frac{\sum_{c(i,3(m-l)-1,t-1)} G_{i_1;i_{3(m-l)-1,t}}}{\Theta_n} \]
and \( \mathbb{E}(X_{n,t}|\mathcal{F}_{n,t-1}) = 0 \). Hence, \( X_{n,t} \) is martingale difference. Note that

\[
\mathbb{E} \left( \sum_{t=1}^{n} \mathbb{E}(S_{n,t} - S_{n,t-1})^2 | \mathcal{F}_{n,t-1}, W, \sigma \right)
\]

(63)

\[
= \sum_{t=1}^{n} \left( \mathbb{E}(S_{n,t}^2 | W, \sigma) - \mathbb{E}(S_{n,t-1}^2 | W, \sigma) \right) = \mathbb{E}(S_{n,n}^2 | W, \sigma) = 1,
\]

and

\[
\text{Var} \left( \sum_{t=1}^{n} \mathbb{E}[(S_{n,t} - S_{n,t-1})^2 | \mathcal{F}_{n,t-1}, W, \sigma] \right)
\]

\[
= \frac{1}{\Theta_n^2} \text{Var} \left( \sum_{t=1}^{n} \mathbb{E} \left( \sum_{c(i,3(m-l)-1,t-1)} G_{i_1:i_3(m-l)-1} \right)^2 | \mathcal{F}_{n,t-1}, W, \sigma \right)
\]

\[
= \frac{1}{\Theta_n^4} \text{Var} \left( \sum_{t=1}^{n} \sum_{s,t=1}^{n} (A_{i_1:i_3m} - \theta_{i_1:i_3m})^2 (A_{i_3m-i_1'i_3's} - \theta_{i_3m-i_1'i_3's})^2 O(a_1) | W, \sigma \right)
\]

\[
= \frac{O(a_1^2)}{\Theta_n^4} \sum_{s,t=1}^{n} (3(m-l)-1) \left( 2m-3l - 1 \right)
\]

\[
= \frac{O(a_1^2)}{\Theta_n^4} \sum_{s=1}^{n} (3(m-l)-1) \sum_{s=1}^{n} (2m-3l - 1)
\]

\[
= \frac{O(a_1^2)}{\Theta_n^4} n^{3(m-l)} n^{2m-3l} = \frac{O(a_1^2)}{n^{6(m-l)} a_1^5} n^{3(m-l)} n^{2m-3l} = \frac{1}{a_1 n^m} \to 0.
\]

Equations (63) and (64) implies that

\[
\sum_{t=1}^{n} \mathbb{E} \left( (S_{n,t} - S_{n,t-1})^2 | \mathcal{F}_{n,t-1}, W, \sigma \right) \to 1,
\]

which is condition (1) in Theorem 4.9.

Next we check the Lindeberg condition. For any \( \epsilon > 0 \), we have

\[
\sum_{t=1}^{n} \mathbb{E} \left( (S_{n,t} - S_{n,t-1})^2 I[|S_{n,t} - S_{n,t-1}| > \epsilon] | \mathcal{F}_{n,t-1}, W, \sigma \right)
\]

\[
\leq \sum_{t=1}^{n} \sqrt{\mathbb{E} \left( (S_{n,t} - S_{n,t-1})^4 | \mathcal{F}_{n,t-1}, W, \sigma \right)} \sum_{t=1}^{n} \sqrt{\mathbb{P}[|S_{n,t} - S_{n,t-1}| > \epsilon] | \mathcal{F}_{n,t-1}, W, \sigma}
\]

\[
\leq \frac{1}{\epsilon^2} \sum_{t=1}^{n} \mathbb{E} \left( (S_{n,t} - S_{n,t-1})^4 | \mathcal{F}_{n,t-1}, W, \sigma \right)
\]

\[
= \frac{1}{\epsilon^2 \Theta_n^4} \sum_{t=1}^{n} \sum_{c(i,3(m-l)-1,t-1)} G_{i_1:i_3(m-l)-1}^4 | \mathcal{F}_{n,t-1}, W, \sigma \right).
\]

(64)

For convenience, let \( c(i) = c(i,3(m-l)-1,t-1), D_{1i} = A_{i_1:i_3m} - \theta_{i_1:i_3m}, D_{2i} = A_{i_3m-i_1'i_3's} - \theta_{i_3m-i_1'i_3's} \).
and $D_{3i} = A_{i2m-2l-1;i3(m-l)}^{i1;i1} - \theta_{i2m-2l-1;i3(m-l)}^{i1;i1}$. Then
\[
\mathbb{E}\left(\sum_{c(i,3(m-l)-1,t-1)} G_{i1;i3(m-l)-1}^{i1} \mid \mathcal{F}_{n,t-1}, W, \sigma\right) = \mathbb{E}\left(\sum_{c(i),c(j),c(r),c(s)} D_{1i} D_{2i} D_{3i}^2 D_{1j} D_{2j} D_{3j}^2 D_{1r} D_{2r} D_{3r}^2 \mid \mathcal{F}_{n,t-1}, W, \sigma\right).
\]
(65)

For indexes $i_{2m-2l-1} : j_{3(m-l)-1} : i_1$, $j_{2m-2l-1} : j_1 : j_1$, $r_{3(m-l)-1} : r_1$ and $s_{3(m-l)-1} : s_1 : s_1$, either all of them are the same or two of them are the same and the other two are the same. Otherwise, the conditional expectation in (65) given $W, \sigma$ vanishes. The same is true for the other two sets of indexes. We consider the case $i_1 : i_{3(m-l)-1} = j_1 : j_3(m-l)-1$ and $r_1 : r_{3(m-l)-1} = s_1 : s_3(m-l)-1$ for example. Then by (65), (64) is equal to
\[
\frac{1}{\epsilon^2} \sum_{t=1}^{n} \mathbb{E}\left(\sum_{c(i),c(j),c(r)} \mathbb{E}D_{1i} D_{2i} D_{3i}^2 D_{1j} D_{2j} D_{3j}^2 D_{1r} D_{2r} D_{3r}^2 \mid \mathcal{F}_{n,t-1}, W, \sigma\right) = \frac{O(\epsilon^6)}{\epsilon^2 n^6(m-l)\alpha_1^6} n^{3(m-l)-1} n^{3(m-l)-1} \rightarrow 0.
\]
The other cases can be similarly proved. Hence,
\[
\sum_{t=1}^{n} \mathbb{E}\left(\mathbb{E}(S_{n,t} - S_{n,t-1})^2 \mid \mathbb{F}_{n,t-1}, W, \sigma\right) \rightarrow 0,
\]
which is condition (2) in Theorem 4.9. Then we conclude that conditional on $W \in W_n$ and $\sigma$,
\[
\sum_{c(i,3(m-l)-n)} G_{i1;i3(m-l)}^{i1} = \sum_{t=1}^{n} (S_{n,t} - S_{n,t-1}) \rightarrow N(0,1).
\]
(66)

Since $\Theta_n \propto \sqrt{\frac{n}{3(m-l)}} T$, and
\[
\left(\begin{array}{c}
n \\
3(m-l)
\end{array}\right) (m-l) (T - T) = \sum_{c(i,3(m-l)-n)} G_{i1;i3(m-l)} + \cdots + \sum_{c(i,3(m-l)-n)} G_{i_{m-l};i3(m-l)} + o(1),
\]
then (35) follows from (66). \hfill \Box

References

[1] Abbe, E. (2017). Community detection and stochastic block models: recent developments. Journal of Machine Learning Research, 18, 1-86.
[2] 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.
[3] Agarwal, S., Branson, K. and Belongie, S. (2006). Higher order learning with graphs. Proceedings of the International Conference on Machine Learning, 17-24.
[4] 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.
[5] Angelini, M., Caltagirone,F., Krzakala, F. and Zdeborova, L. (2015). Spectral detection on sparse hypergraphs. Allerton Conference on Communication, Control, and Computing, 6673.
[6] Banerjee, D. (2018). Contiguity and non-reconstruction results for planted partition models: the dense case. Electronic Journal of Probability, 23, 1-28.
[7] Bolla, M. (1993). Spectra, euclidean representations and clusterings of hypergraphs. Discrete Mathematics, 117(1), 19-39.
[8] Bolllobás, B. (2001). Random Graphs. Cambridge University Press, second edition.
[9] Bollobás, B. and Erdős, P. (1976). Cliques in random graphs. Mathematical Proceedings of the Cambridge Philosophical Society, 80, 419-427.
[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] Banks, J., Moore, C., Neeman, J. and Netrapalli, P. (2016). Information-theoretic thresholds for community detection in sparse networks. JMLR: Workshop and Conference Proceedings 49, 1-34.
[12] 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.
[13] Chertok, M. and Keller, Y. (2010). Efficient high order matching. IEEE Trans. on Pattern Analysis and Machine Intelligence, 32(12), 2205-2215.
[14] Chen, J. and Yuan, B. (2006). Detecting functional modules in the yeast protein-protein interaction network. Bioinformatics, 22(18), 2283-2290.
[15] Decelle, A., Krzakala, F., Moore, C., and Zdeborová, F. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. Physics Review E, 84, 066-106.
[16] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. Publ. Math. Inst. Hungar. Acad. Sci., 5, 17-61.
[17] Estrada, E. and Rodriguez-velasquez, J. (2005). Complex networks as hypergraphs. https://arxiv.org/ftp/physics/papers/0505/0505137.pdf.
[18] Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486 (3-5), 75-174.
[19] Fosdick, B. K. and Hoff, P. D. (2015). Testing and modeling dependencies between a network and nodal attributes. Journal of the American Statistical Association, 110, 1047-1056.
[20] Frieze, A. and Karonski, M. (2015). Introduction to random graphs. Cambridge University Press.
[21] Govindu, V. M. (2005). A tensor decomposition for geometric grouping and segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1150-1157.
[22] Ghoshdastidar, D. and Dukkipati, A. (2014). Consistency of spectral partitioning of uniform hypergraphs under planted partition model. Advances in Neural Information Processing Systems (NIPS), 397-405.
[23] Ghoshdastidar, D. and Dukkipati, A. (2017). Consistency of spectral hypergraph partitioning under planted partition model. The Annals of Statistics, 45(1), 289-315.
[24] Gibson, D., Kleinberg, J. and Raghavan, P. (2000). Clustering categorical data: An approach based on dynamical systems. VLDB Journal, 8, 222-236.
[25] Gao, C. and Lafferty, J. (2017a). Testing for global network structure using small subgraph statistics. https://arxiv.org/pdf/1710.00862.pdf.
[26] Gao, C. and Lafferty, J. (2017b). Testing network structure using relations between small subgraph probabilities. https://arxiv.org/pdf/1704.06742.pdf.
[27] Gao, C., Ma, Z., Zhang, A.Y. and Zhou, H. H. (2016). Community detection in degree-corrected block models. https://arxiv.org/pdf/1607.06993.pdf.
[28] Gao, Z. and Wormald, N. (2004). Asymptotic normality determined by high moments, and submap counts of random maps. Probab. Theory Relat. Fields, 130(3), 368-376.
[29] Ghoshal, G., Zlatic, V., Caldarelli, G. and Newman, M. E. J. (2009). Random hypergraphs and their applications. Physical Review E 79.
[30] Goldenberg, A., Zheng, A. X. S., Fienberg, E., and Airoldi, E. M. (2010). A survey of statistical network models. Foundations and Trends in Machine Learning 2, 129-233.
[31] Hall, P. and Heyde, C. C. (2014). Martingale limit theory and its application. Academic press.
[32] Janson, S. (1995). Random regular graphs: asymptotic distributions and contiguity. Combinatorics, Probability and Computing, 4, 369-405.
[33] Kim, C., Bandeira, A. and Goemans, M. (2017). Community detection in hypergraphs, spiked tensor models, and sum-of-squares. 2017 International Conference on Sampling Theory and Applications (SampTA), 124-128.
[34] Lei, J. (2016). A goodness-of-fit test for stochastic block models. Annals of Statistics, 44, 401-424.
[35] Lin, C., Chien, I. and Wang, I. (2017). On the fundamental statistical limit of community detection in random hypergraphs. Information Theory (ISIT), 2017 IEEE International Symposium , 2178-2182.
[36] Leskovec, J., Lang, K. L., Dasgupta, A. and Mahoney, M. W. (2008). Statistical properties of community structure in large social and information networks. Proceeding of the 17th international conference on World Wide Web, 695-704. ACM.
[37] Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. The Annals of Statistics, 43(1), 215-237.
[38] Michoel, T. and Nachtergaele, B. (2012). Alignment and integration of complex networks by hypergraph-based spectral clustering. Physical Review E 86.

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

[40] Mossel, E., Neeman, J. and Sly, A. (2017). A proof of the block model threshold conjecture. Combinatorica, 1-44.

[41] 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.

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

[43] Neeman, J. and Netrapalli, P. (2014). Non-reconstructability in the stochastic block model. https://arxiv.org/abs/1404.6304

[44] Ouvrard, X., Goff, J. and Marchand-Maillet, S. (2017). Networks of Collaborations: hypergraph modeling and visualisation. https://arxiv.org/pdf/1707.00115.pdf.

[45] Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. The Annals of Probability, 22(4), 1679-1706.

[46] Rodriguez, J. A. (2009). Laplacian eigenvalues and partition problems in hypergraphs. Applied Mathematics Letters, 22(6), 916-921.

[47] Ramasco, J., Dorogovtsev, S. N. and Pastor-Satorras, R. (2004). Self-organization of collaboration networks, Phys. Rev. E 70, 036-106.

[48] Rota Bulo, S. and Pelillo, M. (2013). A game-theoretic approach to hypergraph clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(6), 1312-1327.

[49] Shi, J. and Malik, J. (1997). Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888-905.

[50] Wormald, N. C. (1999). Models of random regular graphs. London Mathematical Society Lecture Note Series, 239-298. Cambridge University Press.

[51] Yuan, M., Feng, Y. and Shang, Z. (2018a). A likelihood-ratio type test for stochastic block models with bounded degrees. https://arxiv.org/pdf/1807.04426.pdf

[52] Yuan, M., Feng, Y. and Shang, Z. (2018b). Inference on multi-community stochastic block models with bounded degree. Manuscript.

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

[54] 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.