ALEA, Lat. Am. J. Probab. Math. Stat. 21, 465–489 (2024)
DOI: 10.30757/ALEA.v21-19

Statistical Limits for Testing Correlation of Random Hypergraphs

Mingao Yuan and Zuofeng Shang

Department of Statistics, North Dakota State University
E-mail address: mingao.yuan@ndsu.edu

Department of Mathematical Sciences, New Jersey Institute of Technology
E-mail address: zshang@njit.edu

Abstract. In this paper, we consider the hypothesis testing of correlation between two m-uniform hypergraphs on n unlabelled nodes. Under the null hypothesis, the hypergraphs are independent, while under the alternative hypothesis, the hyperdges have the same marginal distributions as in the null hypothesis but are correlated after some unknown node permutation. We focus on two scenarios: the hypergraphs are generated from the Gaussian-Wigner model and the dense Erdös-Rényi model. We derive the sharp information-theoretic testing threshold. Above the threshold, there exists a powerful test to distinguish the alternative hypothesis from the null hypothesis. Below the threshold, the alternative hypothesis and the null hypothesis are not distinguishable. The threshold involves m and decreases as m gets larger. This indicates testing correlation of hypergraphs ($m \geq 3$) becomes easier than testing correlation of graphs ($m = 2$).

1. Introduction

Graph matching is a fundamental problem in network data analysis. It refers to the problem of identifying a mapping between the nodes of two graphs that preserves as much as possible the relationships between nodes. Graph matching is a powerful technique and is widely used in a variety of scientific fields. For instance, in shape matching and object recognition, graph matching is used to find the correspondence between object graph and its feature graph (Berg et al. (2005); Cho and Lee (2012)); in social network analysis, graph matching identifies all the accounts belonging to the same individual (Korula and Lattanzi (2014)); in computational biology, graph matching can be applied to match brain-graphs (Vogelstein et al. (2015)). Graph matching problem is NP hard in the worst case and various algorithms have been developed to recover the latent mapping (Cour et al. (2006); Vogelstein et al. (2015); Korula and Lattanzi (2014); Cho and Lee (2012); Barak et al. (2019); Berg et al. (2005); Ding et al. (2021); Yu et al. (2021)). In practice, whether there exists a meaningful matching between two graphs is unknown. To solve this issue, Barak et al. (2019); Wu et al. (2023); Mao et al. (2021) initiate the study of testing the correlation of two graphs. Especially,

Received by the editors January 26th, 2023; accepted February 14th, 2024.

2020 *Mathematics Subject Classification.* 62G10, 05C80.

Key words and phrases. statistical limit, uniform hypergraph, Gaussian-Wigner hypergraph, Erdös-Rényi hypergraph, hypergraph correlation.

Wu et al. (2023) derives the sharp information-theoretic threshold for testing correlated Gaussian-Wigner graphs and dense Erdös-Rényi graphs and Mao et al. (2021) propose a test procedure with polynomial-time complexity.

Many complex networks in the real world can be formulated as hypergraphs. Unlike ordinary graphs where the data structure is typically unique, e.g., edges only contain two vertices, hypergraphs demonstrate a number of possibly overlapping data structures so that an edge may contain arbitrarily many vertices. For instance, in coauthorship networks (Estrada and Rodriguez-Velazquez (2005); Ouvrard et al. (2017); Ramasco et al. (2004); Newman (2001); Yuan et al. (2022)), an edge represents a group of arbitrarily many coauthors; in folksonomy network, an edge may represent a triple (user, resource, annotation) structure (Ghoshal et al. (2009)); in login network an edge may represent a (user, remote host, login time, logout time) structure (Ghoshdastidar and Dukkipati (2014)). Recently, there is increasing interest in hypergraph matching problem, that is, to establish the correspondence between nodes of two unlabelled hypergraphs (Zass and Shashua (2008); Duchenne et al. (2011); Lee et al. (2011); Nguyen et al. (2017); Park et al. (2014); Liao et al. (2021); Hou et al. (2023); Wang et al. (2022)). In this paper, we study the hypothesis testing of correlation for hypergraphs and characterize how the sharp testing threshold in Wu et al. (2023) varies in hypergraph.

An *undirected* m-uniform hypergraph is a pair $\mathcal{H}_m = ([n], \mathcal{E})$ in which $[N] := \{1, 2, \ldots, n\}$ is a vertex set and $\mathcal{E}$ is a set of hyperedges. Each hyperedge in $\mathcal{E}$ consists of exactly m vertices in $[n]$. The corresponding adjacency tensor is an m-dimensional symmetric array $A \in (B^n)^{\otimes m}$ satisfying $A_{i_1 i_2 \ldots i_m} \in B$ for $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, in which $B \subset \mathbb{R}$. Here, symmetry means that $A_{i_1 i_2 \ldots i_m} = A_{j_1 j_2 \ldots j_m}$ whenever $i_1, i_2, \ldots, i_m$ is a permutation of $j_1, j_2, \ldots, j_m$. If $|\{i_1, i_2, \ldots, i_m\}| \leq m-1$, then $A_{i_1 i_2 \ldots i_m} = 0$, i.e., no self-loops are allowed. In particular, $B = \{0, 1\}$ corresponds to binary hypergraphs. The general B corresponds to weighted hypergraphs. For convenience, we also denote the hypergraph $\mathcal{H}_m = ([n], \mathcal{E})$ as $\mathcal{H}_m = ([n], A)$.

Let P_n be the permutation group on $[n]$. Two hypergraphs $\mathcal{H}_{m,1} = ([n], A_1)$ and $\mathcal{H}_{m,2} = ([n], A_2)$ are said to be isomorphic, denoted as $\mathcal{H}_{m,1} \cong \mathcal{H}_{m,2}$ if there is a permutation $\pi \in P_n$ such that $A_{1, i_1 i_2 \ldots i_m} = A_{2, \pi_{i_1} \pi_{i_2} \ldots \pi_{i_m}}$ for all $1 \leq i_1 < i_2 < \cdots < i_m \leq n$. Clearly isomorphism defines an equivalence relation and denote the equivalence class of $\mathcal{H}_{m,1}$ as $\overline{\mathcal{H}}_{m,1}$. Each hypergraph $\mathcal{H}_m \in \overline{\mathcal{H}}_{m,1}$ is called an unlabelled hypergraph of $\mathcal{H}_{m,1}$.

For two hypergraphs $\mathcal{H}_{m,1} = ([n], A_1)$ and $\mathcal{H}_{m,2} = ([n], A_2)$, suppose $(A_{1, i_1 i_2 \ldots i_m}, A_{2, i_1 i_2 \ldots i_m})$, $(1 \leq i_1 < i_2 < \cdots < i_m \leq n)$ are independently and identically distributed random variables with $A_{1, i_1 i_2 \ldots i_m}$ and $A_{2, i_1 i_2 \ldots i_m}$ sharing the same marginal distribution. Given two unlabelled hypergraphs (random sample) $\widetilde{A}_1 \in \overline{\mathcal{H}}_{m,1}$ and $\widetilde{A}_2 \in \overline{\mathcal{H}}_{m,2}$, our purpose is to test the following hypergraph correlation hypothesis.

$$
\begin{aligned}
&H_0: \ A_{1, i_1 i_2 \ldots i_m} \textit{ and } A_{2, i_1 i_2 \ldots i_m} \textit{ are independent}, \quad 1 \leq i_1 < i_2 < \cdots < i_m \leq n; \\
&H_1: \ A_{1, i_1 i_2 \ldots i_m} \textit{ and } A_{2, i_1 i_2 \ldots i_m} \textit{ are correlated}, \quad\ \ 1 \leq i_1 < i_2 < \cdots < i_m \leq n. \qquad (1.1)
\end{aligned}
$$

When $m = 2$, (1.1) is just the graph correlation hypothesis testing problem studied in Barak et al. (2019); Wu et al. (2023); Mao et al. (2021). It is not immediately clear what role $m \geq 3$ plays in the hypothesis testing problem (1.1). This motivates us to study (1.1) for general $m \geq 2$.

In this paper, we focus on two scenarios.

(I) Gaussian-Wigner hypergraph: For all $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, $A_{1, i_1 i_2 \ldots i_m}$ and $A_{2, i_1 i_2 \ldots i_m}$ follow the bivariate normal distribution with mean zero, variance one and correlation coefficient $\rho \in [0, 1]$. Then (1.1) is simplified to $H_0 : \rho = 0$, $v.s.$ $H_1 : \rho > 0$. Under H_0, $A_{1, i_1 i_2 \ldots i_m}$ and $A_{2, i_1 i_2 \ldots i_m}$ are independent and follow the standard normal distribution. Under H_1, $A_{1, i_1 i_2 \ldots i_m}$ and $A_{2, i_1 i_2 \ldots i_m}$ follows the standard normal distribution but are correlated with correlation ρ. This model is a natural extension of the correlated

Gaussian-Wigner model proposed in Ding et al. (2021) and has been studied in Ding et al. (2021); Fan et al. (2023a); Ganassali et al. (2022) when $m = 2$.

(II) Erdös-Rényi hypergraph: Let $\mathcal{H}_m$ and $\mathcal{H}'_m$ be independent random Erdös-Rényi m-uniform hypergraphs with hyperedge probability $p \in [0, 1]$. The we can restate (1.1) as follows: H_0 is equivalent to that $\mathcal{H}_{m,1}$ and $\mathcal{H}_{m,2}$ are generated from $\mathcal{H}_m$ and $\mathcal{H}'_m$ respectively by keeping each hyperedge independently with probability $s \in [0, 1]$; H_1 is equivalent to that $\mathcal{H}_{m,1}$ and $\mathcal{H}_{m,2}$ are similarly generated from the same hypergraph $\mathcal{H}_m$. Under H_0, the hypergraphs $\mathcal{H}_{m,1}$ and $\mathcal{H}_{m,2}$ are independently subsampled from two independent hypergraphs $\mathcal{H}_m$ and $\mathcal{H}'_m$. Hence, the correlation is zero. Under H_1, the hypergraphs $\mathcal{H}_{m,1}$ and $\mathcal{H}_{m,2}$ are independently subsampled from the same hypergraph $\mathcal{H}_m$. In this case, the hypergraphs $\mathcal{H}_{m,1}$ and $\mathcal{H}_{m,2}$ are correlated and the correlation between $(A_{1,i_1 i_2 \dots i_m}, A_{2,i_1 i_2 \dots i_m})$, $(1 \leq i_1 < i_2 < \dots < i_m \leq n)$ is $\rho = \frac{s(1-p)}{1-ps}$. When $m = 2$, this model is the correlated Erdös-Rényi model proposed in Pedarsani and Grossglauser (2011) and has been widely studied in the graph matching problem Ding et al. (2021); Wu et al. (2023); Fan et al. (2023b); Ganassali and Massoulié (2020); Mossel and Xu (2019). Figure 1.1-Figure 1.3 provide an illustration of the correlated Erdös-Rényi hypergraphs..

The above two models serve as prototypes of random hypergraph matching. As far as we know, this is the first time that the two models have been studied in hypergraph setting.

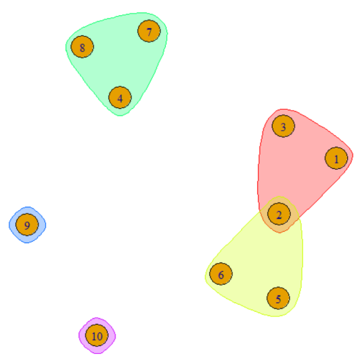

FIGURE 1.1. *A random 3-uniform hypergraph with 3 hyperedges and 10 nodes, denoted as* $\mathcal{H}_3$.

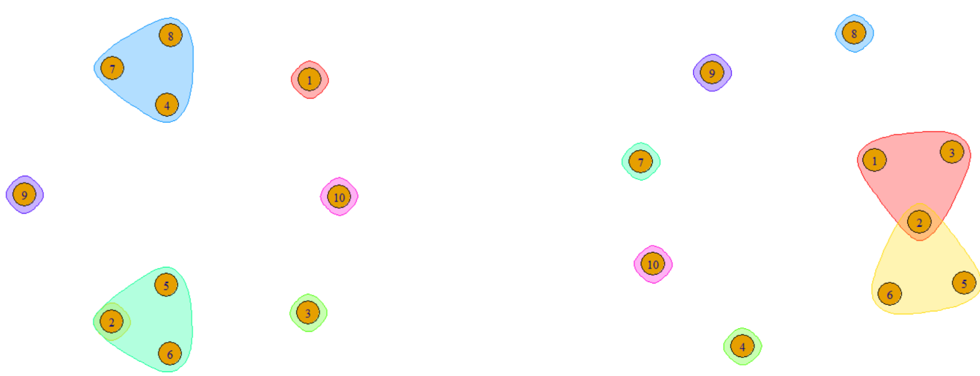

FIGURE 1.2. *Two labeled hypergraphs that are subsampled from* $\mathcal{H}_3$ *in Figure 1.1 with* $s = 0.8$.

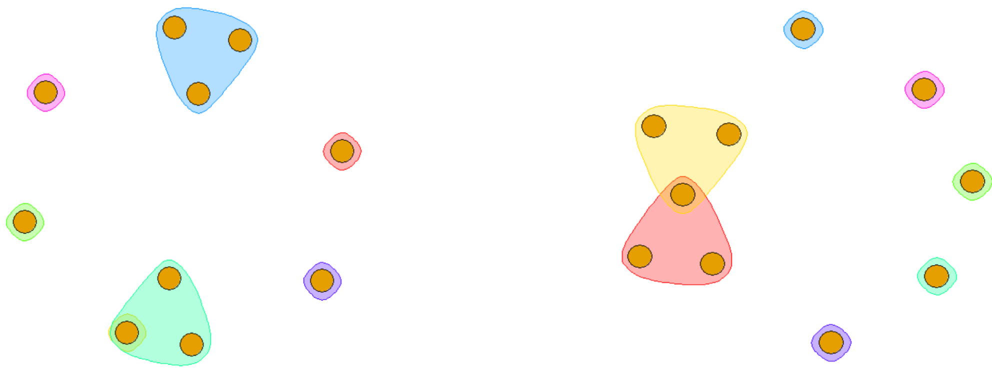

FIGURE 1.3. *The observed hypergraphs that are unlabeled versions of the hypergraphs in Figure 1.2.*

We shall use the total variation distance to measure the difference between H_1 and H_0. The total variation distance between two probability measures P, Q on a sigma-algebra $\mathcal{F}$ of subsets of the sample space Ω is defined as

$$TV(P,Q) = \sup_{E\in\mathcal{F}} |P(E) - Q(E)|.$$

Let P, Q be probability measures under H_0, H_1 respectively. Then H_0 and H_1 are said to be indistinguishable if $TV(P,Q) = o(1)$ and distinguishable if $TV(P,Q) = 1 + o(1)$.

In this paper, we adopt the Bachmann-Landau notation $o(1), O(1)$. For two positive sequences a_n, b_n, denote $a_n \asymp b_n$ or $a_n = \Theta(b_n)$ if $0 < c_1 \leq \frac{a_n}{b_n} \leq c_2 < \infty$ for constants c_1, c_2. Denote $a_n \gg b_n$ or $b_n \ll a_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$. We write $a_n = \Omega(b_n)$ if $a_n \geq cb_n$ for a constant $c > 0$. $I[E]$ denotes the indicator function of event E.

The rest of the paper is organized as follows. In section 2, we present the main result and related proof for Gaussian-Wigner Model. Section 3 provides the main result and proof for Erdös-Rényi Model. Some necessary lemmas are given in section 4.

2. Gaussian-Wigner Hypergraph

In this section, we study the hypergraph correlation test problem under the Gaussian-Wigner model. Denote $\pi \sim Unif(P_n)$ if π is uniformly and randomly selected from P_n. In this case, the hypothesis (1.1) is reformulated as follows.

$$\begin{aligned} H_0 : \begin{pmatrix} A_{1,i_1i_2\ldots i_m} \\ A_{2,i_1i_2\ldots i_m} \end{pmatrix} &\overset{i.i.d.}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right], \\ H_1 : \begin{pmatrix} A_{1,i_1i_2\ldots i_m} \\ A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}} \end{pmatrix} &\overset{i.i.d.}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right], \ \textit{conditional on } \pi \sim Unif(P_n). \end{aligned} \tag{2.1}$$

When $m = 2$, the Gaussian-Wigner model is proposed in Ding et al. (2021) and studied in Ganassali et al. (2022); Fan et al. (2023a); Wu et al. (2023). The following result provides the sharp information-theoretic threshold for hypothesis testing problem (2.1).

Theorem 2.1 (Gaussian-Wigner hypergraph)**.** *Let $m \geq 2$ be any fixed integer. Then H_0 and H_1 in (2.1) are distinguishable if*

$$\rho^2 \geq \frac{2n\log n}{\binom{n}{m}}.$$

H_0 and H_1 in (2.1) are indistinguishable if

$$\rho^2 < \frac{(1-\epsilon)2n\log n}{\binom{n}{m}}, \tag{2.2}$$

for any constant $\epsilon > 0$.

For Gaussian-Wigner model, a phase transition phenomenon occurs at the threshold $\frac{2n\log n}{\binom{n}{m}}$: H_1 and H_0 are distinguishable if and only if the correlation is above the threshold. Theorem 2.1 generalizes Theorem 1 in Wu et al. (2023) to m-uniform hypergraph with $m \geq 2$. When $m = 2$, we have

$$\frac{2n\log n}{\binom{n}{2}} = \frac{4\log n}{n-1}, \qquad \frac{(1-\epsilon)2n\log n}{\binom{n}{2}} = \frac{(4-4\epsilon)\log n}{n-1}. \tag{2.3}$$

Since ϵ is an arbitrary positive constant, the thresholds (2.3) are the same as those in Theorem 1 of Wu et al. (2023). When $m = 3$, we have

$$\frac{2n\log n}{\binom{n}{3}} = \frac{12\log n}{(n-1)(n-2)}, \quad \frac{(1-\epsilon)2n\log n}{\binom{n}{3}} = \frac{(12-12\epsilon)\log n}{(n-1)(n-2)}. \tag{2.4}$$

For large n, the thresholds in (2.4) is smaller than that in (2.3), which implies the indistinguishable region of 3-uniform hypergraph ($m = 3$) is smaller than graph ($m = 2$). Table 2.1 summarizes the thresholds for $m = 2, 3, 4, 5, 6$. In general, the threshold decreases at rate $\frac{\log n}{n^{m-1}}$ as a function of m. This indicates that testing correlated Gaussian-Wigner hypergraphs ($m \geq 3$) is easier than testing correlated Gaussian-Wigner graphs (see result for $m = 2$ in Wu et al. (2023)).

m	2	3	4	5	6
threshold	$\frac{4\log n}{n-1}$	$\frac{12\log n}{(n-1)(n-2)}$	$\frac{48\log n}{(n-1)(n-2)(n-3)}$	$\frac{240\log n}{(n-1)(n-2)(n-3)(n-4)}$	$\frac{1440\log n}{(n-1)(n-2)(n-3)(n-4)(n-5)}$

TABLE 2.1. The testing threshold for m-uniform Gaussian Wigner hypergraph.

The proof of Theorem 2.1 follows the same strategy as Wu et al. (2023). For the positive part, we show that the generalized maximum likelihood estimator can achieve asymptotic power approaching one. For the negative part, the truncated second moment method is used to show that no test can achieve high power. However, technical derivations are nontrivial and more involved for general m. In particular, cares need to be taken w.r.t. the design of generalized MLE and the study of truncated second moment for arbitrary m.

Proof of Theorem 2.1: (Positive Part). Note that one minus the total variation distance is less than or equal to the sum of type I error and type II error of any test. Hence we only need to construct a test with type I error and type II error convergent to zero. We shall construct a powerful test statistic based on the maximum likelihood method. Since the testing problem is easier for larger ρ^2, then we can assume $\rho^2 = \frac{2n\log n}{\binom{n}{m}}$. For convenience, let $t_n = \rho\binom{n}{m} - \sqrt{\binom{n}{m}}n^{0.25}$.

Let π be a uniformly and randomly selected permutation on $[n]$ such that $A_{1,i_1i_2\ldots i_m}$ and $A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}$ follow the bivariate normal distribution with mean zero, variance one and correlation coefficient $\rho \in [0, 1]$.

Under H_1, the likelihood ratio given π is equal to

$$\begin{aligned}\frac{Q(A_1,A_2|\pi)}{P(A_1,A_2)} &= \frac{1}{\sqrt{1-\rho^2}^{\binom{n}{m}}}\exp\left\{-\frac{\rho^2}{2(1-\rho^2)}\sum_{1\le i_1<\cdots<i_m\le n}(A^2_{1,i_1i_2\ldots i_m}+A^2_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}})\right\}\\ &\quad\times\exp\left\{\frac{\rho}{1-\rho^2}\sum_{1\le i_1<\cdots<i_m\le n}A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}\right\}. \qquad (2.5)\end{aligned}$$

Hence, to maximize the likelihood ratio with respect to π is equivalent to maximizing $T(\pi)$ given by

$$T(\pi)=\sum_{1\le i_1<\cdots<i_m\le n}A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}.$$

Then we define the test statistic as $T_n=\max_\pi T(\pi)$.

Under the alternative hypothesis, we shall show $\mathbb{P}(T_n\ge t_n)=1+o(1)$. By the Hanson-Wright inequality in Lemma 4.1, it is easy to verify that

$$\mathbb{P}(T_n\le t_n)\le\mathbb{P}(T(\pi)\le t_n)\le e^{-cn^{\frac{m+0.5}{2}}}+e^{-c\sqrt{n}},$$

for some constant $c>0$. Then $\mathbb{P}(T_n\le t_n)=o(1)$ and hence $\mathbb{P}(T_n\ge t_n)=1+o(1)$.

Under the null hypothesis, we show $\mathbb{P}(T_n\ge t_n)=o(1)$. Note that $A_{1,i_1i_2\ldots i_m}$ and $A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}$ are independent for any π and they follow the standard normal distribution. For $\lambda:=\frac{t_n}{\binom{n}{m}}=o(1)$, the Chernoff bound in Lemma 4.2 yields

$$\begin{aligned}\mathbb{P}(T(\pi)\ge t_n) &= \mathbb{P}(e^{T(\pi)}\ge e^{t_n})\le\exp\left\{-\lambda t_n-\frac{\binom{n}{m}}{2}\log(1-\lambda^2)\right\}\\ &= \exp\left\{-2n\log n-n^{0.5}+2\sqrt{2n\log n}n^{0.25}+n\log n+\frac{n^{0.5}}{2}+o(n)\right\}.\end{aligned}$$

Note that $n!\le en^{n+0.5}e^{-n}$. Then by the union bound, it follows that

$$\mathbb{P}(T_n\ge t_n)\le n!\mathbb{P}(T(\pi)\ge t_n)=\exp(-n+o(n))=o(1).$$

Then the proof is complete. □

Proof of Theorem 2.1: (Negative Part). It is well-known that if the second moment of the likelihood ratio converges to one under H_0, then the total variation distance between the two probability measures converges to zero. Therefore, to prove the negative result, it suffices to prove that

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right]\le 1+o(1),$$

under H_0. The details are given in the following Proposition 2.2 and Proposition 2.4. □

Before presenting Proposition 2.2 and Proposition 2.4, we provide some basic facts about permutation. Each permutation $\pi\in P_n$ can be decomposed into product of disjoint cycles. Each cycle forms an orbit of any element in the cycle. Let K_m be the complete m-uniform hypergraph on $[n]$. Then π induces a permutation π^K on the hyperedge set of K_m by

$$\pi^K(i_1,i_2,\ldots,i_m)=(\pi_{i_1},\pi_{i_2},\ldots,\pi_{i_n}),\quad i_1<i_2<\cdots<i_m.$$

We call π node permutation and π^K hyperedge permutation. Let n_k denote the number of cycles (orbits) in π with length k and N_k the number of hyperedge cycles (hyperedge orbits) with length k. Note that N_k can be expressed as a function of $n_t,(t\le k)$. For example, let $m=3$. Then $N_1=\binom{n_1}{3}+n_1n_2+n_3$.

Proposition 2.2. *For any fixed integer $m \geq 2$, if $\rho^2 < \frac{(1-\epsilon)n\log n}{\binom{n}{m}}$ for any constant $\epsilon \in [0,1)$, then H_0 and H_1 are indistinguishable for both Gaussian Wigner model and Erdos-Renyi model.*

Proof of Proposition 2.2: We only need to focus on $m \geq 3$, since the result for $m = 2$ is given in Wu et al. (2023). Denote $\tilde{\pi}$ be an independent copy of π. Firstly, we consider Gaussian Wigner model. Define

$$\begin{aligned} & L_1(A_{1,i_1i_2\ldots i_m}, A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}) \\ = \;& \frac{1}{\sqrt{1-\rho^2}} \exp\left\{ \frac{-\rho^2(A^2_{1,i_1i_2\ldots i_m} + A^2_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}) + 2\rho A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}}{2(1-\rho^2)} \right\}, \end{aligned}$$

and

$$L_{i_1i_2\ldots i_m} = L_1(A_{1,i_1i_2\ldots i_m}, A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}})L_1(A_{1,i_1i_2\ldots i_m}, A_{2,\tilde{\pi}_{i_1}\tilde{\pi}_{i_2}\ldots\tilde{\pi}_{i_m}}).$$

By (2.5), the second moment of the likelihood ratio under H_0 is equal to

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right] = \mathbb{E}_{\pi,\tilde{\pi}}\left(\mathbb{E}\left[\frac{Q(A_1,A_2|\pi)}{P(A_1,A_2)}\frac{Q(A_1,A_2|\tilde{\pi})}{P(A_1,A_2)}\right]\right) = \mathbb{E}_{\pi,\tilde{\pi}}\left(\mathbb{E}\Big[\prod_{1\leq i_1<\cdots<i_m\leq n} L_{i_1i_2\ldots i_m}\Big]\right). \tag{2.6}$$

Denote $\sigma = \pi^{-1}\circ\tilde{\pi}$. For a hyperedge orbit O induced by σ, define

$$L_O = \prod_{\{i_1,\ldots,i_m\}\in O} L_{i_1i_2\ldots i_m}.$$

Since $\tilde{\pi}(e) = \pi\circ\sigma(e)$ for any hyperedge e, then L_O only depends on $A_{1,e}, A_{2,\pi_e}$ for $e \in O$.

Let $\mathcal{O}$ be the set of hyperedge orbits of σ. Note that the hyperedge orbits are mutually disjoint and $A_{1,i_1i_2\ldots i_m}$ and $A_{2,i_1i_2\ldots i_m}$ are i.i.d. under H_0. Then by (2.6), we have

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right] = \mathbb{E}_{\pi,\tilde{\pi}}\left(\prod_{O\in\mathcal{O}}\mathbb{E}(L_O)\right) = \mathbb{E}_{\pi,\tilde{\pi}}\left[\prod_{k=1}^{\binom{n}{m}}\left(\frac{1}{1-\rho^{2k}}\right)^{N_k}\right], \tag{2.7}$$

where the second equality follows from Proposition 1 in Wu et al. (2023) and N_k is the number of hyperedge orbits with length k.

Note that $\sum_{k=2}^{\binom{n}{m}} N_k \leq n^m$. According to (2.2), $\rho^4 n^m = O\left(\frac{\log n}{n^{m-2}}\right) = o(1)$ for $m \geq 3$. Consequently,

$$\prod_{k=2}^{\binom{n}{m}}\left(\frac{1}{1-\rho^{2k}}\right)^{N_k} \leq \left(\frac{1}{1-\rho^4}\right)^{\binom{n}{m}} \leq \exp\left(\frac{n^m\rho^4}{1-\rho^4}\right) = 1+o(1).$$

Then

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right] \leq (1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left[\left(\frac{1}{1-\rho^2}\right)^{N_1}\right] \leq (1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left[\exp\left(\frac{N_1\rho^2}{1-\rho^2}\right)\right] \leq 1+o(1). \tag{2.8}$$

where the last step follows from the following Lemma 2.3. Then the proof is complete for Gaussian Wigner model.

For Erdos-Renyi model, by a similar argument and using Proposition 1 in Wu et al. (2023), we have

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right] = \mathbb{E}_{\pi,\tilde{\pi}}\left[\prod_{k=1}^{\binom{n}{m}}\left(1+\rho^{2k}\right)^{N_k}\right].$$

By the condition $\rho^2 < \frac{(1-\epsilon)n\log n}{\binom{n}{m}}$ and $m \geq 3$, it follows that

$$\prod_{k=2}^{\binom{n}{m}} \left(1+\rho^{2k}\right)^{N_k} \leq \left(1+\rho^4\right)^{\binom{n}{m}} \leq \exp\left(n^m\rho^4\right) = 1+o(1).$$

Hence

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right] \leq (1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left[\left(1+\rho^2\right)^{N_1}\right] \leq (1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left[\exp\left(N_1\rho^2\right)\right] \leq 1+o(1).$$

Here the last inequality follows from Lemma 2.3. □

Lemma 2.3. *Let N_1 be the number of hyperedge orbits of $\sigma = \pi^{-1}\circ\tilde{\pi}$ with length one. If $\rho^2 < \frac{(1-\epsilon)n\log n}{\binom{n}{m}}$ for any positive constant ϵ, then*

$$\mathbb{E}_{\pi,\tilde{\pi}}\left[\exp\left(\frac{N_1\rho^2}{1-\rho^2}\right)\right] \leq 1+o(1).$$

Proof of Lemma 2.3: Let n_k be the number of k-nodes cycles of permutation σ. Since the cycles of σ are disjoint, then $n_k \leq n$. Note that 1-hyperedge orbit is just a single hyperedge and this hyperedge can only involve nodes in k-nodes cycles with $k \leq m$. Hence, $N_1 = R(n_1,n_2,\dots,n_m)$, where $R(n_1,n_2,\dots,n_m)$ is a polynomial in $n_1,n_2,\dots,n_m$. If a hyperedge contains a k-node cycle, then we only need to select $m-k$ nodes to form a hyperedge. Hence, any terms in $R(n_1,n_2,\dots,n_m)$ involving k-node cycles are bounded by $n_kn^{m-k} = O(n^{m-k+1})$. Since $\rho^2 < \frac{(1-\epsilon)n\log n}{\binom{n}{m}}$, then $\rho^2n^{m-k+1} = O\left(\frac{\log n}{n^{k-2}}\right) = o(1)$ for $k \geq 3$. If a term in $R(n_1,n_2,\dots,n_m)$ contains n_2^k, then it is bounded by $\rho^2n_2^kn^{m-2k} = O\left(\frac{\log n}{n^{k-1}}\right) = o(1)$ for $k \geq 2$. Hence, we have

$$\rho^2N_1 = \rho^2\left[\binom{n_1}{m} + n_2\binom{n_1}{m-2}\right] + o(1).$$

Then

$$\begin{aligned}
\mathbb{E}_{\pi,\tilde{\pi}}\left[\exp\left(\frac{N_1\rho^2}{1-\rho^2}\right)\right] &= (1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\left[\binom{n_1}{m} + n_2\binom{n_1}{m-2}\right]\right)\right] \\
&= (1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\left[\binom{n_1}{m} + n_2\binom{n_1}{m-2}\right]\right)I[0 \leq n_1 < \sqrt{n}]\right] \\
&\quad + (1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\left[\binom{n_1}{m} + n_2\binom{n_1}{m-2}\right]\right)I[\sqrt{n} \leq n_1 \leq n]\right] \\
&= (a) + (b).
\end{aligned}$$

If $n_1 < \sqrt{n}$, then

$$\rho^2\left[\binom{n_1}{m} + n_2\binom{n_1}{m-2}\right] = O\left(\frac{n^{\frac{m}{2}}\log n}{n^{m-1}} + \frac{n^{1+\frac{m-2}{2}}\log n}{n^{m-1}}\right) = o(1), \quad m \geq 3.$$

Hence $(a) = 1+o(1)$.

Next, we show $(b) = o(1)$ if $\rho^2 < \frac{(1-\epsilon)n\log n}{\binom{n}{m}}$. Let Z_t, $(1 \leq t \leq k)$ be independent Poisson variables with $Z_t \sim Poi(\frac{1}{t})$. By Lemma 4.3, we have

$$\begin{aligned}(b) &\leq (1+o(1))\mathbb{E}_{Z_1,Z_2}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\left[\binom{Z_1}{m} + Z_2\binom{Z_1}{m-2}\right]\right) I[\sqrt{n} \leq Z_1 \leq n]\right] e^{\frac{3}{2}} \\ &= (1+o(1))e^{\frac{3}{2}}\mathbb{E}_{Z_1}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\binom{Z_1}{m}\right) I[\sqrt{n} \leq Z_1 \leq n]\mathbb{E}_{Z_2}\left(\exp\left(\frac{\rho^2}{1-\rho^2}Z_2\binom{Z_1}{m-2}\right)\middle| Z_1\right)\right]. \end{aligned} \tag{2.9}$$

By the moment generating function of Poisson distribution, we have

$$\mathbb{E}_{Z_2}\left(\exp\left(\frac{\rho^2}{1-\rho^2}Z_2\binom{Z_1}{m-2}\right)\middle| Z_1\right) \quad = \quad \exp\left[\frac{1}{2}\left(e^{\frac{\rho^2}{1-\rho^2}\binom{Z_1}{m-2}} - 1\right)\right].$$

On the event $\sqrt{n} \leq Z_1 \leq n$, it follows that

$$\frac{\rho^2}{1-\rho^2}\binom{Z_1}{m-2} = O\left(\frac{\log n}{n^{m-1}}n^{m-2}\right) = o(1).$$

Hence, by (2.9), $k! \geq \left(\frac{k}{e}\right)^k$ and $\rho^2 < \frac{(1-\epsilon)n\log n}{\binom{n}{m}}$, we have

$$\begin{aligned}(b) &\leq (1+o(1))e^{\frac{3}{2}}\mathbb{E}_{Z_1}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\binom{Z_1}{m}\right) I[\sqrt{n} \leq Z_1 \leq n]\right] \\ &= (1+o(1))e^{\frac{3}{2}-1}\sum_{k=\sqrt{n}}^{n}\exp\left(\frac{\rho^2}{1-\rho^2}\binom{k}{m}\right)\frac{1}{k!} \\ &\leq (1+o(1))e^{\frac{3}{2}-1}\sum_{k=\sqrt{n}}^{n}\exp\left(\frac{\rho^2}{1-\rho^2}\binom{k}{m} - k\log k - k\right) \\ &\leq (1+o(1))e^{\frac{3}{2}-1}\sum_{k=\sqrt{n}}^{n}\exp\left(k\left((1-\epsilon)\frac{\log n}{n^{m-1}}k^{m-1} - \log k\right) - k\right). \end{aligned} \tag{2.10}$$

Define $f(k) = (1-\epsilon)\frac{\log n}{n^{m-1}}k^{m-1} - \log k$. The derivative of $f(k)$ is equal to

$$f'(k) = (1-\epsilon)(m-1)\frac{\log n}{n^{m-1}}k^{m-2} - \frac{1}{k}.$$

Solving $f'(k) = 0$ yields $k_0 = \frac{n}{((1-\epsilon)(m-1)\log n)^{\frac{1}{m-1}}}$. Then $f(k)$ is decreasing for $k \leq k_0$ and increasing for $k \geq k_0$. Hence,

$$\begin{aligned} f(k) \leq \max\left\{f(\sqrt{n}), f(n)\right\} &= \max\left\{(1-\epsilon)\frac{\log n}{n^{m-1}}\sqrt{n}^{m-1} - \log\sqrt{n}, (1-\epsilon)\frac{\log n}{n^{m-1}}n^{m-1} - \log n\right\} \\ &= \max\left\{-\frac{1}{2}\log n(1+o(1)), -\epsilon\log n\right\}. \end{aligned}$$

By (2.10), for a positive consant c, we have

$$(b) \quad \leq \quad (1+o(1))e^{\frac{3}{2}-1}e^{\left(\log n - \sqrt{n} - c\sqrt{n}\log n\right)} = o(1).$$

Then the proof is complete. □

The bound in Proposition 2.2 is not sharp. The conditional second moment method will be used to close the gap. The result is summarized in the following Proposition 2.4.

Proposition 2.4. *If* $\frac{n\log n}{\binom{n}{m}} \leq \rho^2 < \frac{(1-\epsilon)2n\log n}{\binom{n}{m}}$ *for any positive constant* ϵ*, then* H_0 *and* H_1 *are indistinguishable.*

Proof of Proposition 2.4: We use the conditional second moment method as in Wu et al. (2023) to prove Proposition 2.4.

Let I be the set of fixed points of σ and $\mathcal{O}_1$ be the set of subsets in I with cardinality m. Then for any $\{i_1,\ldots,i_m\}\in\mathcal{O}_1$, $\{\pi_{i_1},\ldots,\pi_{i_m}\}=\{\tilde{\pi}_{i_1},\ldots,\tilde{\pi}_{i_m}\}$. For $S\subset[n]$ and a positive constant C, define event E_S as

$$\begin{aligned} E_S &= \Bigg\{\sum_{\{i_1,i_2,\ldots i_m\}\subset S} A^2_{1,i_1i_2\ldots i_m}\geq\binom{|S|}{m}-t_S, \sum_{\{i_1,i_2,\ldots i_m\}\subset S} A^2_{2,\pi_{i_1}\ldots\pi_{i_m}}\geq\binom{|S|}{m}-t_S, \\ &\qquad \sum_{\{i_1,i_2,\ldots i_m\}\subset S} A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\ldots\pi_{i_m}}\leq\rho\binom{|S|}{m}+t_S\Bigg\}, \end{aligned}$$

where t_S is of order $n^{\frac{m+1}{2}}$. Let

$$E=\cap_{S\subset[n],|S|\geq\frac{n}{2^{\frac{1}{m-1}}}}E_S.$$

We shall use E to truncate the second moment. By Lemma 2.5, $\mathbb{P}(E)=1-o(1)$ under H_1. Hence, we have

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right]=\mathbb{E}_{\pi,\tilde{\pi}}\left(\prod_{O\in\mathcal{O}}\mathbb{E}(L_OI[A_1,A_2,\pi\in E]I[A_1,A_2,\tilde{\pi}\in E])\right). \tag{2.11}$$

For $n_1\leq\frac{n}{2^{\frac{1}{m-1}}}$, by a similar argument as in the proof of Lemma 2.3, one has

$$\begin{aligned} &\prod_{O\in\mathcal{O}}\mathbb{E}(L_OI[A_1,A_2,\pi\in E]I[A_1,A_2,\tilde{\pi}\in E])\leq\mathbb{E}\left(\prod_{O\in\mathcal{O}}L_O\right)=\prod_{O\in\mathcal{O}}\frac{1}{1-\rho^{2|O|}} \\ =&\prod_{O\in\mathcal{O}_1}\frac{1}{1-\rho^{2|O|}}\prod_{O\notin\mathcal{O}_1}\frac{1}{1-\rho^{2|O|}} \\ =&\left(\frac{1}{1-\rho^2}\right)^{\binom{n_1}{m}+n_2\binom{n_1}{m-2}}\left(\frac{1}{1-\rho^4}\right)^{\binom{n}{m}}(1+o(1)) \\ \leq&\exp\left(\frac{\rho^2}{1-\rho^2}\left[\binom{n_1}{m}+n_2\binom{n_1}{m-2}\right]\right)(1+o(1)). \end{aligned}$$

Suppose $n_1\geq\frac{n}{2^{\frac{1}{m-1}}}$. Since $\rho^2\geq\frac{n\log n}{\binom{n}{m}}$, then $n^{\frac{m+1}{2}}=o(\rho\binom{n_1}{m})$. In this case, on event E_I, we get

$$\sum_{\{i_1,i_2,\ldots i_m\}\subset S}A^2_{1,i_1i_2\ldots i_m}\geq\binom{|S|}{m}(1+o(1)),\quad\sum_{\{i_1,i_2,\ldots i_m\}\subset S}A^2_{2,\pi_{i_1}\ldots\pi_{i_m}}\geq\binom{|S|}{m}(1+o(1)),$$

$$\sum_{\{i_1,i_2,\ldots i_m\}\subset S}A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\ldots\pi_{i_m}}\leq\rho\binom{|S|}{m}(1+o(1)).$$

Then it follows that

$$
\begin{aligned}
& \mathbb{E}\left[\prod_{O\in\mathcal{O}} L_O I[A_1,A_2,\pi\in E] I[A_1,A_2,\tilde{\pi}\in E]\right] \\
\leq\ & \mathbb{E}\left[\prod_{O\in\mathcal{O}} L_O I[A_1,A_2,\pi\in E_I]\right] \\
=\ & \mathbb{E}\left[\prod_{\{i_1,i_2,\dots i_m\}\subset I} L_{i_1 i_2\dots i_m} I[A_1,A_2,\pi\in E_I]\right] \prod_{O\notin\mathcal{O}_1} \frac{1}{1-\rho^{2|O|}}.
\end{aligned}
$$

Further, on event E_I, the following inequalities hold.

$$
\begin{aligned}
& \mathbb{E}\Big[\prod_{\{i_1,i_2,\dots i_m\}\subset I} L_{i_1 i_2\dots i_m} I[A_1,A_2,\pi\in E_I]\Big] \\
& \leq \frac{1}{(1-\rho^2)^{\binom{n_1}{m}}}\mathbb{E}\left[\exp\Big\{-\frac{\rho^2}{(1-\rho^2)}\sum_{\{i_1,i_2,\dots i_m\}\subset I}(A^2_{1,i_1 i_2\dots i_m}+A^2_{2,\pi_{i_1}\pi_{i_2}\dots\pi_{i_m}})\Big\}\right. \\
& \qquad\qquad \left.\times\exp\left\{\frac{2\rho}{1-\rho^2}\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m}A_{2,\pi_{i_1}\pi_{i_2}\dots\pi_{i_m}}\right\} I[A_1,A_2,\pi\in E_I]\right] \\
& \leq \frac{1}{(1-\rho^2)^{\binom{n_1}{m}}}\exp\left\{-\frac{(1+o(1))2\rho^2}{(1-\rho^2)}\binom{n_1}{m}\right\} \\
& \times\mathbb{E}\left[\exp\Big\{\frac{2\rho}{1-\rho^2}\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m}A_{2,\pi_{i_1}\pi_{i_2}\dots\pi_{i_m}}\Big\} I\left[\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m}A_{2,\pi_{i_1}\pi_{i_2}\dots\pi_{i_m}}\leq\rho\binom{n_1}{m}\right]\right] \\
& \leq \frac{1}{(1-\rho^2)^{\binom{n_1}{m}}}\exp\left\{-\frac{(1+o(1))2\rho^2}{(1-\rho^2)}\binom{n_1}{m}\right\} \\
& \times\mathbb{E}\left[\exp\left\{\frac{2\rho}{1-\rho^2}\left(\frac{1-\rho^2}{2}\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m}A_{2,\pi_{i_1}\pi_{i_2}\dots\pi_{i_m}}+(1-\frac{1-\rho^2}{2})\rho\binom{n_1}{m}\right)\right\}\right] \\
& = \frac{1}{(1-\rho^2)^{\binom{n_1}{m}}}\exp\left\{-\frac{(1+o(1))2\rho^2}{(1-\rho^2)}\binom{n_1}{m}\right\}\exp\left\{\frac{\rho^2(1+\rho^2)}{1-\rho^2}\binom{n_1}{m}\right\}\exp\left\{-\frac{1}{2}\binom{n_1}{m}\log(1-\rho^2)\right\} \\
& = \exp\left\{\frac{(1+o(1))\rho^2}{2}\binom{n_1}{m}\right\}.
\end{aligned}
$$

In the second last equality we used the fact that $\mathbb{E}[e^{\lambda XY}]=\frac{1}{1-\lambda^2}$ for independent standard normal random variables X,Y and $|\lambda|<1$.

Then we can bound the second moment of the likelihood ratio under H_0 as

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right] &= (1+o(1))\mathbb{E}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\left[\binom{n_1}{m}+n_2\binom{n_1}{m-2}\right]\right) I[n_1\leq\frac{n}{2^{\frac{1}{m-1}}}]\right] \\
&\quad +(1+o(1))\mathbb{E}\left[\exp\left\{\frac{\rho^2}{2}\binom{n_1}{m}+\frac{\rho^2}{1-\rho^2}n_2\binom{n_1}{m-2}\right\} I[n_1\geq\frac{n}{2^{\frac{1}{m-1}}}]\right] \\
&= (c)+(d).
\end{aligned}
$$

By the proof of Lemma 2.3 and $\rho^2 < \frac{(1-\epsilon)2n\log n}{\binom{n}{m}}$, we have

$$
\begin{aligned}
(c) &= (1+o(1))\mathbb{E}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\binom{n_1}{m}\right) I[n_1 \le \sqrt{n}]\right] \\
&\quad + (1+o(1))\mathbb{E}\left[\exp\left(\frac{\rho^2}{1-\rho^2}\binom{n_1}{m}\right) I[\sqrt{n} < n_1 \le \frac{n}{2^{\frac{1}{m-1}}}]\right] \\
&\le 1 + o(1) + e^{\frac{3}{2}-1}\sum_{k=\sqrt{n}}^{\frac{n}{2^{\frac{1}{m-1}}}} \exp\left(k\left((1-\epsilon)\frac{2\log n}{n^{m-1}}k^{m-1} - \log k\right) - k\right).
\end{aligned}
$$

Let $f(k) = (1-\epsilon)\frac{2\log n}{n^{m-1}}k^{m-1} - \log k$. Similar to the proof of Lemma 2.3, it is easy to verify

$$
f(k) \le \max\left\{f(\sqrt{n}), f\left(\frac{n}{2^{\frac{1}{m-1}}}\right)\right\} = \max\left\{-\frac{1}{2}\log n(1+o(1)), -\epsilon\log n(1+o(1))\right\}.
$$

Hence, $(c) = 1 + o(1)$.

For (d), by Lemma 4.3, one has

$$
\begin{aligned}
(d) &= (1+o(1))\mathbb{E}\left[\exp\left(\frac{\rho^2}{2}\binom{n_1}{m}\right) I[n_1 \ge \frac{n}{2^{\frac{1}{m-1}}}]\right] \\
&\le e^{\frac{3}{2}-1}\sum_{k=\frac{n}{2^{\frac{1}{m-1}}}}^{n} \exp\left(k\left((1-\epsilon)\frac{\log n}{n^{m-1}}k^{m-1} - \log k\right) - k\right).
\end{aligned}
$$

Let $f(k) = (1-\epsilon)\frac{\log n}{n^{m-1}}k^{m-1} - \log k$. Similar to the proof of Lemma 2.3, it is easy to verify

$$
f(k) \le \max\left\{f(n), f\left(\frac{n}{2^{\frac{1}{m-1}}}\right)\right\} = \max\left\{-\frac{1}{2}\log n(1+o(1)), -\frac{\epsilon+1}{2}\log n\right\}.
$$

Hence, $(d) = o(1)$. Then it follows that

$$
\mathbb{E}\left[\left(\frac{Q(A_1, A_2)}{P(A_1, A_2)}\right)^2\right] \le 1 + o(1).
$$

The proof is complete. $\square$

Lemma 2.5. *Under H_1, $\mathbb{P}(E) = 1 - o(1)$.*

Proof of Lemma 2.5: For integer k with $\frac{n}{2^{\frac{1}{m-1}}} \le k \le n$, let $\delta_k = 2^{-k}\binom{n}{k}^{-1}$, S be a subset with $|S| = k$ and $t_S = C\left(\sqrt{\binom{k}{m}\log\frac{1}{\delta_k}} + \log\frac{1}{\delta_k}\right) = Cn^{\frac{m+1}{2}}(1+o(1))$. By Hanson-Wright inequality in Lemma 4.1, we have $\mathbb{P}(E_S^c) \le 6\delta_k$. Hence,

$$
\mathbb{P}(E^c) \le 6\sum_{k=\frac{n}{2^{\frac{1}{m-1}}}}^{n}\binom{n}{k}\delta_k \le 6n2^{-\frac{n}{2^{\frac{1}{m-1}}}} = o(1).
$$

Then the proof is complete. $\square$

3. Erdös-Rényi Hypergraph

In this section, we study the hypergraph correlation test under the Erdös-Rényi model. In this case, the hypothesis (1.1) is reformulated as follows.

$$H_0: A_{1,i_1 i_2 \dots i_m}, A_{2,i_1 i_2 \dots i_m} \overset{i.i.d.}{\sim} Bern(ps),$$
$$H_1: A_{1,i_1 i_2 \dots i_m} \overset{i.i.d.}{\sim} Bern(ps),\ A_{2,\pi_{i_1}\pi_{i_2}\dots\pi_{i_m}} \overset{i.i.d.}{\sim} Bern\left(sA_{1,i_1 i_2 \dots i_m} + (1 - A_{1,i_1 i_2 \dots i_m})\frac{ps(1-s)}{1-ps}\right),$$
$$\textit{conditional on } \pi \sim Unif(P_n). \tag{3.1}$$

It is easy to verify the correlation between $A_{1,i_1 i_2 \dots i_m}$ and $A_{2,\pi_{i_1}\pi_{i_2}\dots\pi_{i_m}}$ under H_1 is

$$\rho = \frac{s(1-p)}{1-ps}.$$

When $p = o(1)$, $\rho = s(1 + o(1))$. In this case, s measures the scale of correlation. For $m = 2$, the correlated Erdös-Rényi graph model is proposed in Pedarsani and Grossglauser (2011) and widely studied in graph matching problem (Barak et al. (2019); Mossel and Xu (2019); Ding et al. (2021); Wu et al. (2023); Mao et al. (2021)).

The following theorem provides a sharp testing threshold for hypothesis (3.1) when the Erdös-Rényi hypergraphs are dense.

Theorem 3.1 (Erdös-Rényi model). *Let $m \geq 2$ be a fixed integer. Then H_0 and H_1 in (3.1) are distinguishable if*

$$s^2 \geq \frac{n \log n}{\binom{n}{m}\left(\log \frac{1}{p} - 1 + p\right)p}.$$

Suppose p is bounded away from one and $\log \frac{1}{p} = o(\log n)$. Then H_0 and H_1 in (3.1) are indistinguishable if

$$s^2 < \frac{(1-\epsilon) n \log n}{\binom{n}{m}\left(\log \frac{1}{p} - 1 + p\right)p}, \tag{3.2}$$

for any constant $\epsilon > 0$.

Theorem 3.1 generalizes the result of the dense regime in Theorem 2 of Wu et al. (2023) to m-uniform hypergraph with $m \geq 2$. For $m = 2$, we have

$$\frac{n \log n}{\binom{n}{2}\left(\log \frac{1}{p} - 1 + p\right)p} = \frac{2 \log n}{(n-1)p\left(\log \frac{1}{p} - 1 + p\right)}, \quad \frac{(1-\epsilon) n \log n}{\binom{n}{2}\left(\log \frac{1}{p} - 1 + p\right)p} = \frac{(2-2\epsilon) \log n}{(n-1)p\left(\log \frac{1}{p} - 1 + p\right)}. \tag{3.3}$$

Since ϵ is an arbitrary positive constant, the thresholds in (3.3) coincide with that in Theorem 2 of Wu et al. (2023). When $m = 3$, we have

$$\frac{n \log n}{\binom{n}{3}\left(\log \frac{1}{p} - 1 + p\right)p} = \frac{6 \log n}{(n-1)(n-2)p\left(\log \frac{1}{p} - 1 + p\right)}, \tag{3.4}$$

$$\frac{(1-\epsilon) n \log n}{\binom{n}{3}\left(\log \frac{1}{p} - 1 + p\right)p} = \frac{(6-6\epsilon) \log n}{(n-1)(n-2)p\left(\log \frac{1}{p} - 1 + p\right)}. \tag{3.5}$$

The thresholds in (3.4) and (3.5) are smaller than that in (3.3), which implies the indistinguishable region of 3-uniform hypergraph ($m = 3$) is smaller than graph ($m = 2$). Table 3.2 summarizes the thresholds for $m = 2, 3, 4, 5$. Generally, the sharp testing boundary $\frac{n \log n}{\binom{n}{m}\left(\log \frac{1}{p} - 1 + p\right)p}$ decreases as

m gets larger. This shows that testing correlated Erdös-Rényi hypergraph ($m \geq 3$) is easier than testing correlated Erdös-Rényi graphs (see result for $m = 2$ in Wu et al. (2023)).

m	2	3	4	5
threshold	$\frac{2\log n}{(n-1)p\left(\log\frac{1}{p}-1+p\right)}$	$\frac{6\log n}{(n-1)(n-2)p\left(\log\frac{1}{p}-1+p\right)}$	$\frac{24\log n}{(n-1)(n-2)(n-3)p\left(\log\frac{1}{p}-1+p\right)}$	$\frac{120\log n}{(n-1)(n-2)(n-3)(n-4)p\left(\log\frac{1}{p}-1+p\right)}$

TABLE 3.2. The testing threshold for m-uniform Erdös-Rényi hypergraph.

The proof of Theorem 3.1 follows the same proof strategy as in Wu et al. (2023). For the positive result, we show that the generalized maximum likelihood estimator can achieve asymptotic power one. For the negative result, the truncated second moment method is used to show that no test can achieve high power. However the proof is not trivial and straightforward. How to incorporate m in the proof needs special care.

Proof of Theorem 3.1: (Positive Part). Note that one minus the total variation distance is less than or equal to the sum of type I error and type II error of any test. Hence we only need to construct a test with type I error and type II error convergent to zero. Similar to the Gaussian Wigner model, we shall use the maximum likelihood method to construct a powerful test statistic. The likelihood ratio given π is equal to

$$\begin{aligned}
\frac{Q(A_1, A_2|\pi)}{P(A_1, A_2)} \quad = \quad & \prod_{1\leq i_1<\cdots<i_m\leq n} (1-s)^{A_{1,i_1i_2\ldots i_m}} \left(\frac{1-2ps+ps^2}{1-ps}\right)^{1-A_{1,i_1i_2\ldots i_m}} \\
& \times \prod_{1\leq i_1<\cdots<i_m\leq n} \frac{1}{1-ps}\left(\frac{(1-ps)(1-s)}{1-2ps+ps^2}\right)^{A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}} \\
& \times \left(\frac{1-2ps+ps^2}{p(1-s)^2}\right)^{\sum_{1\leq i_1<\cdots<i_m\leq n} A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}}.
\end{aligned}$$

Let $T_n = \max_\pi T(\pi)$ with $T(\pi) = \sum_{1\leq i_1<\cdots<i_m\leq n} A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}$.

The correlation coefficient ρ for Erdos-Renyi model is given by

$$\rho = \frac{s(1-p)}{1-ps}.$$

Larger s implies larger correlation ρ. Hence, it is easier to test the correlation. Then we can assume

$$s^2 = \frac{n\log n}{\binom{n}{m}\left(\log\frac{1}{p}-1+p\right)p}, \tag{3.6}$$

which implies $p \gg \frac{1}{n^{m-1}}$ and $\binom{n}{m}ps^2 \gg n$. Let $t_n = \binom{n}{m}ps^2(1-\tau_n)$ with $\left(\binom{n}{m}ps^2\right)^{-0.5} \ll \tau_n < 1$.

Under H_1, we show $\mathbb{P}(T_n \geq t_n) = 1 + o(1)$. Note that the product $A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}$ are independent and follow Bernoulli(ps^2). Hence $T(\pi) \sim Binomial(\binom{n}{m}, ps^2)$. By Chenorff bound in Lemma 4.2, it is easy to get

$$\mathbb{P}(T_n \leq t_n) \leq \mathbb{P}(T(\pi) \leq t_n) \leq e^{-\frac{\tau_n^2}{2}\binom{n}{m}ps^2} = o(1).$$

Next, we show under H_0, $\mathbb{P}(T_n \geq t_n) = o(1)$. In this case, $A_{1,i_1i_2\ldots i_m}A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}$ are independent and follow Bernoulli(p^2s^2). Hence $T(\pi) \sim Binomial(\binom{n}{m}, p^2s^2)$. By the multiplicative Chernoff bound in Lemma 4.2, we have

$$\begin{aligned}
\mathbb{P}(T_n \geq t_n) &\leq n!\mathbb{P}(T(\pi) \geq t_n) \\
&\leq n! \exp\left(\binom{n}{m} p^2 s^2 \left[\frac{1-\tau_n}{p} \log \frac{1-\tau_n}{p} + 1 - \frac{1-\tau_n}{p}\right]\right) \\
&= n! \exp\left(\binom{n}{m} ps^2 (1-\tau_n) \log \frac{1-\tau_n}{ep} - \binom{n}{m} p^2 s^2\right) \\
&\leq n! \exp\left[-\binom{n}{m} ps^2 \left(\log \frac{1}{p} - 1 + p\right) + \tau_n \binom{n}{m} ps^2 \log \frac{1}{p}\right] \\
&\leq e \exp\left[-n + \tau_n \binom{n}{m} ps^2 \log \frac{1}{p} + 0.5 \log n\right].
\end{aligned}$$

If p is bounded away from one, then $\binom{n}{m}ps^2 = O(n \log n)$. Taking $\tau_n = \left(\binom{n}{m}ps^2\right)^{-0.5} \log n$ and noting that $\log \frac{1}{p} = o(\log n)$ yields $\mathbb{P}(T_n \geq t_n) = o(1)$.

Suppose $p = 1 + o(1)$. Take $\tau_n = \left(\binom{n}{m}ps^2\right)^{-\epsilon}$ with $\frac{m-1}{2m-1} < \epsilon < 0.5$. Note that for some positive constant $c > 0$, by (3.6) it follows that

$$\log \frac{1}{p} - 1 + p = \frac{(1-p)^2}{p}(1 - \frac{1}{2p})(1 + o(1)) \geq c \frac{\log n}{n^{m-1}}.$$

Besides, $\log \frac{1}{p} < \frac{1-p}{p}$. Hence

$$\tau_n \binom{n}{m} ps^2 \log \frac{1}{p} = O\left(\frac{(n \log n)^{1-\epsilon}}{(1-p)^{2(1-\epsilon)}}(1-p)\right) = O\left(\log n)^{\epsilon} n^{1+(m-1)-\epsilon(2m-1)}\right) = o(n).$$

Then $\mathbb{P}(T_n \geq t_n) = o(1)$. The proof is complete. □

Proof of Theorem 3.1: (Negative Part). It is well-known that if the second moment of the likelihood ratio converges to one under H_0, then the total variation distance between the two probability measures converges to zero. To prove the negative result, we only need to prove

$$\mathbb{E}\left[\left(\frac{Q(A_1, A_2)}{P(A_1, A_2)}\right)^2\right] \leq 1 + o(1), \tag{3.7}$$

under H_0.

We shall use the truncation method as in Wu et al. (2023) to prove (3.7). Assume $p \in (0, 1 - \epsilon_0)$ for a constant $\epsilon_0 \in (0, 1)$ and $\log \frac{1}{p} = o(\log n)$. Note that the smaller s is, the harder it is to test the correlation. Hence we assume

$$s^2 = \frac{(1-\epsilon) n \log n}{\binom{n}{m}\left(\log \frac{1}{p} - 1 + p\right) p}, \tag{3.8}$$

for any constant $\epsilon \in (0, 1)$. In this case,

$$s^2 = o(1), \quad s^2 \gg \frac{n}{\binom{n}{m}} = n^{-(m-1)}, \quad \binom{n}{m} ps^2 \gg n. \tag{3.9}$$

Now we define the truncation event. Let $w(x)$ be the solution of equation $w(x)e^{w(x)} = x$ for $x \geq -\frac{1}{e}$. Define $\zeta(k)$ as

$$\zeta(k) = \binom{k}{m} ps^2 \exp\left(1 + w\left(\frac{k \log \frac{2en}{k}}{eps^2 \binom{k}{m}} - \frac{1}{e}\right)\right), \quad k \geq m.$$

Let

$$\alpha_p = \left(\log \frac{1}{p} - 1 + p\right) p.$$

Clearly, $\alpha_p \geq cn^{-o(1)}$ for some constant $c > 0$ and $n\alpha_p^{\frac{1}{m-1}} \geq cn^{1-o(1)}$. Define event E as

$$E = \cap_{n\alpha_p^{\frac{1}{m-1}} \leq |S| \leq n, S \subset [n]} E_S,$$

where E_S is given by

$$\begin{aligned} E_S \quad = \quad & \Bigg\{ \sum_{\{i_1,i_2,\ldots i_m\}\subset S} A_{1,i_1i_2\ldots i_m} \geq \binom{|S|}{m} ps - \sqrt{2\binom{|S|}{m} ps|S| \log \frac{2en}{|S|}}, \\ & \sum_{\{i_1,i_2,\ldots i_m\}\subset S} A_{2,\pi_{i_1}\ldots\pi_{i_m}} \geq \binom{|S|}{m} ps - \sqrt{2\binom{|S|}{m} ps|S| \log \frac{2en}{|S|}}, \\ & \sum_{\{i_1,i_2,\ldots i_m\}\subset S} A_{1,i_1i_2\ldots i_m} A_{2,\pi_{i_1}\ldots\pi_{i_m}} \leq \zeta(|S|) \Bigg\} \end{aligned}$$

Lemma 3.2. *Under H_1, $\mathbb{P}(E) = 1 - e^{-\Omega(n\alpha_p^{\frac{1}{m-1}})}$.*

Proof of Lemma 3.2: For $S \subset [n]$ with $|S| = k$, let $\delta_k = \left(\frac{k}{2en}\right)^k$ and $t_n = \sqrt{2\binom{|S|}{m} ps \log \frac{1}{\delta_k}}$ and

$$v_n = \binom{k}{m} ps^2 \exp\left(1 + w\left(\frac{\log \frac{1}{\delta_k}}{eps^2\binom{k}{m}} - \frac{1}{e}\right)\right).$$

By the multiplicative Chernoff bound in Lemma 4.2, we have

$$\mathbb{P}\left(\sum_{\{i_1,i_2,\ldots i_m\}\subset S} A_{1,i_1i_2\ldots i_m} \leq \binom{|S|}{m} ps - t_n\right) \leq \exp\left(-\log \frac{1}{\delta_k}\right) = \left(\frac{k}{2en}\right)^k,$$

$$\mathbb{P}\left(\sum_{\{i_1,i_2,\ldots i_m\}\subset S} A_{2,\pi_{i_1}\ldots\pi_{i_m}} \leq \binom{|S|}{m} ps - t_n\right) \leq \exp\left(-\log \frac{1}{\delta_k}\right) = \left(\frac{k}{2en}\right)^k,$$

$$\mathbb{P}\left(\sum_{\{i_1,i_2,\ldots i_m\}\subset S} A_{1,i_1i_2\ldots i_m} A_{2,\pi_{i_1}\ldots\pi_{i_m}} \geq v_n\right) \leq \exp\left(-\log \frac{1}{\delta_k}\right) = \left(\frac{k}{2en}\right)^k.$$

Hence,

$$\mathbb{P}(E^c) \leq \sum_{k=n\alpha_p^{\frac{1}{m-1}}}^{n} \binom{n}{k} 3\delta_k \leq 3 \sum_{k=n\alpha_p^{\frac{1}{m-1}}}^{n} \frac{1}{2^k} = e^{-\Omega(n\alpha_p^{\frac{1}{m-1}})}.$$

Then the proof is complete. □

We use E to truncate the second moment. Let

$$L_{i_1i_2\ldots i_m} = L_1(A_{1,i_1i_2\ldots i_m}, A_{2,\pi_{i_1}\pi_{i_2}\ldots\pi_{i_m}}) L_1(A_{1,i_1i_2\ldots i_m}, A_{2,\tilde{\pi}_{i_1}\tilde{\pi}_{i_2}\ldots\tilde{\pi}_{i_m}}),$$

with

$$L_1(x, y) = \frac{1-\eta}{1-ps}\left(\frac{1-s}{1-\eta}\right)^{x+y}\left(\frac{s(1-\eta)}{\eta(1-s)}\right)^{xy}, \quad \eta = \frac{ps(1-s)}{1-ps}.$$

Then by Lemma 3.2 we have

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right]=(1+o(1))\mathbb{E}_{\pi,\tilde{\pi}}\left(\prod_{O\in\mathcal{O}}\mathbb{E}(L_O I[A_1,A_2,\pi\in E]I[A_1,A_2,\tilde{\pi}\in E])\right). \tag{3.10}$$

Next we show the expectation in the right-hand side of (3.10) is less than or equal to $1+o(1)$.

If $n_1\le n\alpha_p^{\frac{1}{m-1}}$, then

$$\prod_{O\in\mathcal{O}}\mathbb{E}(L_O I[A_1,A_2,\pi\in E]I[A_1,A_2,\tilde{\pi}\in E])\le\prod_{O\in\mathcal{O}}\mathbb{E}(L_O)=\prod_{O\in\mathcal{O}}(1+\rho^{2|O|}).$$

If $n_1> n\alpha_p^{\frac{1}{m-1}}$, then

$$\begin{aligned}&\prod_{O\in\mathcal{O}}\mathbb{E}(L_O I[A_1,A_2,\pi\in E]I[A_1,A_2,\tilde{\pi}\in E])\le\prod_{O\in\mathcal{O}}\mathbb{E}(L_O I[E_I]))\\ =&\prod_{O\notin\mathcal{O}_1}\mathbb{E}(L_O)\prod_{O\in\mathcal{O}_1}\mathbb{E}(L_O I[E_I]))\\ =&\prod_{O\notin\mathcal{O}_1}(1+\rho^{2|O|})\prod_{\{i_1,\dots,i_m\}\subset I}\mathbb{E}(L_{i_1i_2\dots i_m}I[E_I])\end{aligned}$$

Note that for $n_1>n\alpha_p^{\frac{1}{m-1}}\ge cn^{1-o(1)}$,

$$\frac{t_n^2}{\binom{n_1}{m}^2p^2s^2}=O\left(\frac{\log\frac{n}{n_1}}{n_1^{m-1}ps}\right)=o\left(\frac{\log n}{n^{\frac{m-1}{2}-o(1)}}\right)=o(1).$$

Hence, on E_I with $n_1>n\alpha_p^{\frac{1}{m-1}}$, one has

$$\sum_{\{i_1,i_2,\dots i_m\}\subset I}A_{1,i_1i_2\dots i_m}\ge\binom{n_1}{m}ps(1+o(1)),\qquad\sum_{\{i_1,i_2,\dots i_m\}\subset I}A_{2,\pi_{i_1}\dots\pi_{i_m}}\ge\binom{n_1}{m}ps(1+o(1)),$$

$$\sum_{\{i_1,i_2,\dots i_m\}\subset I}A_{1,i_1i_2\dots i_m}A_{2,\pi_{i_1}\dots\pi_{i_m}}\le\zeta(n_1).$$

Next, we consider the order of $\zeta(n_1)$. Define

$$\gamma=\frac{n_1\log\frac{2en}{n_1}}{\binom{n_1}{m}ps^2}.$$

Lemma 3.3. $[I]$. *If $\gamma=o(1)$, then $\zeta(n_1)=\binom{n_1}{m}ps^2(1+o(1))$.*

$[II]$. *If $\gamma=\Theta(1)$, then $\zeta(n_1)=\Theta\left(\binom{n_1}{m}ps^2\right)$. For $n_1>n\alpha_p^{\frac{1}{m-1}}$, $\zeta(n_1)=o\left(\binom{n_1}{m}s^2\right)$.*

$[III]$. *If $\gamma\gg1$, then $\zeta(n_1)=(e+o(1))\binom{n_1}{m}ps^2\frac{\gamma}{\log\gamma}$. For $n_1>n\alpha_p^{\frac{1}{m-1}}$, $\zeta(n_1)=o\left(\binom{n_1}{m}s^2\right)$.*

Proof of Lemma 3.3: $[I]$ follows from the fact that $w(\frac{\gamma-1}{e})=-1+\sqrt{2\gamma}+O(\gamma)$ if $\gamma=o(1)$.

For $[II]$, if $\gamma=\Theta(1)$, it is obvious that $\zeta(n_1)=\Theta\left(\binom{n_1}{m}ps^2\right)$. Suppose $n_1>n\alpha_p^{\frac{1}{m-1}}$ and $p\ge c$ for some constant $c>0$. Then $s^2=\Theta\left(\frac{\log n}{n^{m-1}}\right)$, $n_1=\Theta(n)$ and

$$\gamma=O\left(\frac{\log\frac{n}{n_1}}{n^{m-1}s^2}\right)=\frac{O(1)}{\log n}=o(1),$$

which contradicts $\gamma=\Theta(1)$. Hence, $p=o(1)$.

For $[III]$, note that $w(x) = \log x - \log\log x + o(1)$ if $x \gg 1$. Then $\zeta(n_1) = (e + o(1))\binom{n_1}{m} ps^2 \frac{\gamma}{\log\gamma}$. If $n_1 > n\alpha_p^{\frac{1}{m-1}}$, then

$$p\gamma = O\left(\frac{\log\frac{n}{n_1}}{\alpha_p n^{m-1}s^2}\right) = O\left(\frac{\log\frac{1}{p}}{\log n}\right) = o(1).$$

The proof is complete. □

By Lemma 3.3, we get $\zeta(n_1) = \binom{n_1}{m} ps^2 + o\left(\binom{n_1}{m} s^2\right)$ for $n_1 > n\alpha_p^{\frac{1}{m-1}}$. Then

$$\begin{aligned}
&\prod_{\{i_1,\dots,i_m\}\subset I} \mathbb{E}(L_{i_1 i_2\dots i_m} I[E_I]) \leq \left(\frac{1-\eta}{1-ps}\right)^{2\binom{n_1}{m}} \left(\frac{1-s}{1-\eta}\right)^{4(1+o(1))\binom{n_1}{m}ps} \\
&\times \mathbb{E}\left[\left(\frac{s(1-\eta)}{\eta(1-s)}\right)^{2\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m} A_{2,\pi_{i_1}\dots\pi_{i_m}}} I[\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m} A_{2,\pi_{i_1}\dots\pi_{i_m}} \leq \zeta(n_1)]\right] \\
&= \exp\left(-2(1+o(1))\binom{n_1}{m} ps^2(1-p)\right) \\
&\times \mathbb{E}\left[\lambda^{\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m} A_{2,\pi_{i_1}\dots\pi_{i_m}}} I[\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m} A_{2,\pi_{i_1}\dots\pi_{i_m}} \leq \zeta(n_1)]\right],
\end{aligned}$$

where $\lambda = \left(\frac{s(1-\eta)}{\eta(1-s)}\right)^2 = (1+o(1))\frac{1}{p^2}$.

Note that for any $t \in [0,1]$,

$$\begin{aligned}
& \mathbb{E}\left[\lambda^{\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m} A_{2,\pi_{i_1}\dots\pi_{i_m}}} I[\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m} A_{2,\pi_{i_1}\dots\pi_{i_m}} \leq \zeta(n_1)]\right] \\
\leq\ & \mathbb{E}\left[\lambda^{t\sum_{\{i_1,i_2,\dots i_m\}\subset I} A_{1,i_1 i_2\dots i_m} A_{2,\pi_{i_1}\dots\pi_{i_m}} + (1-t)\zeta(n_1)}\right] \\
=\ & \lambda^{\zeta(n_1)} \frac{\left(1 + (\lambda^t - 1)p^2 s^2\right)^{\binom{n_1}{m}}}{\lambda^{t\zeta(n_1)}}.
\end{aligned}$$

Let $g(y) = \frac{\left(1+(y-1)p^2s^2\right)^{\binom{n_1}{m}}}{y^{\zeta(n_1)}}$. It is easy to verify that $g(y)$ attains minimum value at $y_0 = \frac{\zeta(n_1)(1-p^2s^2)}{p^2s^2\left(\binom{n_1}{m}-\zeta(n_1)\right)} \in [1,\lambda]$. Let $h(x) = -x\log x - (1-x)\log(1-x)$. Then

$$\begin{aligned}
\prod_{\{i_1,\dots,i_m\}\subset I} \mathbb{E}(L_{i_1 i_2\dots i_m} I[E_I]) &\leq \exp\left(-2(1+o(1))\binom{n_1}{m} ps^2(1-p)\right) \lambda^{\zeta(n_1)} \frac{\left(1+(y_0-1)p^2s^2\right)^{\binom{n_1}{m}}}{y_0^{\zeta(n_1)}} \\
&= \exp\left(-\binom{n_1}{m} ps^2(2-p) + \zeta(n_1)(\log s^2 + o(1)) + \binom{n_1}{m} h\left(\frac{\zeta(n_1)}{\binom{n_1}{m}}\right)\right) \\
&= \exp\left(-\binom{n_1}{m} ps^2(2-p) + \zeta(n_1)\log\frac{e\binom{n_1}{m}s^2}{\zeta(n_1)} + o(\zeta(n_1))\right).
\end{aligned}$$

Then the second moment of the likelihood ratio is bounded by

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right] \leq (1+o(1))\mathbb{E}\left[\prod_{O\in\mathcal{O}}(1+\rho^{2|O|})I[n_1\leq n\alpha_p^{\frac{1}{m-1}}]\right]$$
$$+\mathbb{E}\left[\prod_{O\notin\mathcal{O}_1}(1+\rho^{2|O|})\exp\left(-\binom{n_1}{m}ps^2(2-p)+\zeta(n_1)\log\frac{e\binom{n_1}{m}s^2}{\zeta(n_1)}+o(\zeta(n_1))\right)I[n_1\geq n\alpha_p^{\frac{1}{m-1}}]\right].$$

Recall that $\rho=(1+o(1))s$ and $\log\frac{1}{p}=o(\log n)$. Then

$$\rho^4 n^m=\frac{\log^2 n}{n^{m-2+o(1)}\log^2\frac{1}{p}}=o(1).$$

By the proof of Lemma 2.3, we have

$$\prod_{O\notin\mathcal{O}_1}(1+\rho^{2|O|})=(1+\rho^2)^{n_2\binom{n_1}{m-2}(1+o(1))}(1+\rho^4)^{n^m}=(1+o(1))(1+\rho^2)^{n_2\binom{n_1}{m-2}}\leq(1+o(1))e^{\rho^2 n_2\binom{n_1}{m-2}}.$$

Besides,

$$\prod_{O\in\mathcal{O}_1}(1+\rho^{2|O|})=(1+\rho^2)^{\binom{n_1}{m}}\leq e^{\rho^2\binom{n_1}{m}}.$$

Then we get

$$\mathbb{E}\left[\left(\frac{Q(A_1,A_2)}{P(A_1,A_2)}\right)^2\right]\leq(1+o(1))\mathbb{E}\left[e^{\rho^2\binom{n_1}{m}+\rho^2 n_2\binom{n_1}{m-2}}I[n_1\leq n\alpha_p^{\frac{1}{m-1}}]\right]$$
$$+(1+o(1))\mathbb{E}\left[\exp\left(\rho^2 n_2\binom{n_1}{m-2}-\binom{n_1}{m}ps^2(2-p)+\zeta(n_1)\log\frac{e\binom{n_1}{m}s^2}{\zeta(n_1)}+o(\zeta(n_1))\right)I[n_1\geq n\alpha_p^{\frac{1}{m-1}}]\right]$$
$$=(e)+(f).$$

Next, we are going to show $(e)=1+o(1)$ and $(f)=o(1)$.

We show $(e)=1+o(1)$ first. Similar to the proof of Lemma 2.3, we have

$$\begin{aligned}
&\mathbb{E}\left[e^{\rho^2\binom{n_1}{m}+\rho^2 n_2\binom{n_1}{m-2}}I[n_1\leq n\alpha_p^{\frac{1}{m-1}}]\right]\\
=\;&\mathbb{E}\left[e^{\rho^2\binom{n_1}{m}+\rho^2 n_2\binom{n_1}{m-2}}I[n_1\leq\sqrt{n}]\right]+\mathbb{E}\left[e^{\rho^2\binom{n_1}{m}+\rho^2 n_2\binom{n_1}{m-2}}I[\sqrt{n}<n_1\leq n\alpha_p^{\frac{1}{m-1}}]\right]\\
\leq\;&1+o(1)+e^{\frac{3}{2}}\mathbb{E}\left[e^{\rho^2\binom{Z_1}{m}}I[\sqrt{n}<Z_1\leq n\alpha_p^{\frac{1}{m-1}}]\right]\\
\leq\;&1+o(1)+e^{-1}\sum_{k=\sqrt{n}}^{n\alpha_p^{\frac{1}{m-1}}}e^{\rho^2\binom{k}{m}-k\log k-k}\\
=\;&1+o(1)+e^{-1}\sum_{k=\sqrt{n}}^{n\alpha_p^{\frac{1}{m-1}}}e^{k\left(\frac{(1-\epsilon)\log n}{\alpha_p}\frac{k^{m-1}}{n^{m-1}}-\log k\right)-k}
\end{aligned}$$

Let $f(k)=\frac{(1-\epsilon)\log n}{\alpha_p}\frac{k^{m-1}}{n^{m-1}}-\log k$. It is easy to see

$$f(k)\leq\max\{f(\sqrt{n}),f(n\alpha_p^{\frac{1}{m-1}})\}=-\min\{0.5,\epsilon\}\log n.$$

Hence $(e)\leq 1+o(1)$.

Next, we prove $(f) = o(1)$. To this end, for a large positive constant C, define

$$\beta_1 = \left(\frac{\log^2 \frac{m!\binom{n}{m}ps^2}{n}}{\frac{m!\binom{n}{m}ps^2}{n}} \right)^{\frac{1}{m-1}}, \quad \beta_2 = \left(\frac{\log \frac{m!\binom{n}{m}ps^2}{n}}{C\frac{m!\binom{n}{m}ps^2}{n}} \right)^{\frac{1}{m-1}}.$$

Then n_1 falls in one of the three intervals $[\beta_1 n, n]$, $[\beta_2 n, \beta_1 n]$ and $[n\alpha_p^{\frac{1}{m-1}}, \beta_2 n]$.

If $n_1 \in [\beta_1 n, n]$, then $\gamma = o(1)$ and hence $\zeta(n_1) = (1+o(1))\binom{n_1}{m}ps^2$. In this case,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\rho^2 n_2\binom{n_1}{m-2} - \binom{n_1}{m}ps^2(2-p) + \zeta(n_1)\log\frac{e\binom{n_1}{m}s^2}{\zeta(n_1)} + o(\zeta(n_1))\right) I[n_1 \geq n\beta_1]\right] \\
= \; &\mathbb{E}\left[\exp\left(\rho^2 n_2\binom{n_1}{m-2} - \binom{n_1}{m}ps^2(2-p) + \binom{n_1}{m}ps^2\log\frac{e}{p}\right) I[n_1 \geq n\beta_1]\right] \\
\leq \; &e^{1.5}\mathbb{E}\left[\exp\left(\binom{Z_1}{m}s^2\alpha_p\right) I[Z_1 \geq n\beta_1]\right] \\
\leq \; &e^{0.5}\sum_{k=\beta_1 n}^{n} e^{k\left(\frac{(1-\epsilon)\log n}{n^{m-1}}k^{m-1} - \log k\right) - k} = o(1).
\end{aligned}$$

If $n_1 \in [\beta_2 n, \beta_1 n]$, then $\gamma = \frac{C}{m!(m-1)}(1+o(1))$ and hence $\zeta(n_1) = \Theta\left(\binom{n_1}{m}ps^2\right)$. In this case, for some constant C_1,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\rho^2 n_2\binom{n_1}{m-2} - \binom{n_1}{m}ps^2(2-p) + \zeta(n_1)\log\frac{e\binom{n_1}{m}s^2}{\zeta(n_1)} + o(\zeta(n_1))\right) I[\beta_2 n \leq n_1 \leq n\beta_1]\right] \\
&\leq \mathbb{E}\left[\exp\left(\rho^2 n_2\binom{n_1}{m-2} + C_1\binom{n_1}{m}ps^2\log\frac{e}{p}\right) I[\beta_2 n \leq n_1 \leq n\beta_1]\right] \\
&\leq e^{1.5}\mathbb{E}\left[\exp\left(C_1\binom{Z_1}{m}ps^2\log\frac{1}{p}\right) I[\beta_2 n \leq Z_1 \leq n\beta_1]\right] \\
&\leq e^{0.5}\sum_{k=\beta_2 n}^{\beta_1 n} e^{k\left(C_1\frac{(1-\epsilon)\log n}{n^{m-1}}k^{m-1} - \log k\right) - k} = o(1).
\end{aligned}$$

If $n_1 \in [n\alpha_p^{\frac{1}{m-1}}, \beta_2 n]$, then $\gamma \geq \frac{Cm!}{m-1}(1+o(1))$. For sufficiently large C, we have $\zeta(n_1) = O\left(\frac{n_1\log\frac{2en}{n_1}}{\log\gamma}\right)$. Then

$$\zeta(n_1)\log\frac{e\binom{n_1}{m}s^2}{\zeta(n_1)} + o(\zeta(n_1)) \leq n_1 R_n,$$

where

$$R_n = C_2\frac{\log\frac{n}{n_1}}{\log\gamma}\log\frac{n_1^{m-1}s^2\log\gamma}{\log\frac{n}{n_1}},$$

for a constant C_2. Suppose $R_n = o(\log n)$, then

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\rho^2 n_2\binom{n_1}{m-2} - \binom{n_1}{m}ps^2(2-p) + \zeta(n_1)\log\frac{e\binom{n_1}{m}s^2}{\zeta(n_1)} + o(\zeta(n_1))\right) I[\alpha_p^{\frac{1}{m-1}} n \le n_1 \le n\beta_2]\right] \\
&\le \mathbb{E}\left[\exp\left(\rho^2 n_2\binom{n_1}{m-2} - \binom{n_1}{m}ps^2(2-p) + n_1 R_n\right) I[\alpha_p^{\frac{1}{m-1}} n \le n_1 \le n\beta_2]\right] \\
&\le e^{1.5}\mathbb{E}\left[\exp\left(-\binom{Z_1}{m}ps^2 + Z_1 o(\log n)\right) I[\alpha_p^{\frac{1}{m-1}} n \le n_1 \le n\beta_2]\right] \\
&\le e^{0.5}\sum_{k=\alpha_p^{\frac{1}{m-1}} n}^{\beta_2 n} e^{-k\left(\frac{(1-\epsilon)\log n}{n^{m-1}(\log\frac{1}{p}-1+p)}k^{m-1} + \log k - o(\log n)\right) - k} = o(1).
\end{aligned}$$

Here in the last equality we used the fact that $\log\frac{1}{p} - 1 + p \ge c > 0$ for some constant c, since p is bounded away from one.

Below we prove $R_n = o(\log n)$. Note that $\log\frac{1}{\alpha_p} = o(\log n)$ and hence

$$R_n = C_2\frac{\log\frac{n}{n_1}}{\log\gamma}\log\frac{n_1^{m-1}s^2}{\log\frac{n}{n_1}} + C_2\frac{\log\frac{n}{n_1}}{\log\gamma}\log\log\gamma = C_2\frac{\log\frac{n}{n_1}}{\log\gamma}\log\frac{n_1^{m-1}s^2}{\log\frac{n}{n_1}} + o(\log n).$$

Then it suffices to show

$$\frac{\log\frac{n}{n_1}}{\log\gamma}\log\frac{n_1^{m-1}s^2}{\log\frac{n}{n_1}} = o(\log n). \tag{3.11}$$

Let $x = \frac{n}{n_1} \in [\frac{1}{\beta_2}, \alpha_p^{-\frac{1}{m-1}}]$. Then $\gamma = \Theta\left(\frac{x^{m-1}\log x}{n^{m-1}ps^2}\right)$. To prove (3.11), we only need to prove

$$\max_{x\in[\frac{1}{\beta_2},\alpha_p^{-\frac{1}{m-1}}]}\frac{\log x}{\log\frac{x^{m-1}\log x}{n^{m-1}ps^2}}\log\frac{n^{m-1}s^2}{x^{m-1}\log n} = o(\log n). \tag{3.12}$$

Let $\delta = \left(\log\frac{\log n}{\log\frac{1}{\alpha_p}}\right)^{-1}$. Then $\delta = o(1)$. To prove (3.12), it suffices to show $\psi(x) \le 0$, with $\psi(x)$ given by

$$\psi(x) = \log(x)\log\frac{n^{m-1}s^2}{x^{m-1}\log n} - \delta\log(n)\log\frac{x^{m-1}\log x}{n^{m-1}ps^2}.$$

Straightforward calculation yields

$$\psi'(x) = \frac{\log(n^{m-1}s^2) - 2(m-1)\log x - \log\log x - 1 - (m-1)\delta\log n - \frac{\delta\log n}{\log x}}{x}.$$

It is easy to see that

$$\frac{\log\frac{1}{\alpha_p}}{\delta\log n} = \frac{\log\frac{\log n}{\log\frac{1}{\alpha_p}}}{\frac{\log n}{\log\frac{1}{\alpha_p}}} = o(1).$$

Hence

$$\log(n^{m-1}s^2) = \log\log n + \log\frac{1}{\alpha_p} = o(\delta\log n).$$

This implies $\psi'(x) \le 0$ for $x \in [\frac{1}{\beta_2}, \alpha_p^{-\frac{1}{m-1}}]$. Then $\psi(x) \le \psi(\frac{1}{\beta_2})$. Since $\frac{1}{\beta_2^{m-1}}\log\frac{1}{\beta_2} = \frac{C}{m-1}(1 + o(1))n^{m-1}ps^2$, then

$$\begin{aligned}\psi\left(\frac{1}{\beta_2}\right) &= \log\left(\frac{1}{\beta_2}\right)\log\frac{n^{m-1}s^2}{\frac{1}{\beta_2^{m-1}}\log n} - \delta\log(n)\log\frac{\frac{1}{\beta_2^{m-1}}\log\frac{1}{\beta_2}}{n^{m-1}ps^2}\\ &= \log\left(\frac{1}{\beta_2}\right)\log\frac{1}{Cp} - \log(C)\delta\log n,\end{aligned}$$

which is negative if p is bounded away from zero. Assume $p = o(1)$. Then

$$\frac{1}{\delta}\log\left(\frac{1}{\beta_2}\right) = O\left(\log\frac{\log n}{\log\frac{1}{\alpha_p}}\log\frac{\frac{\log n}{\log\frac{1}{\alpha_p}}}{\log\frac{\log n}{\log\frac{1}{\alpha_p}}}\right) = O\left(\log^2\frac{\log n}{\log\frac{1}{\alpha_p}}\right) = o\left(\frac{\log n}{\log\frac{1}{\alpha_p}}\right),$$

which implies $\psi\left(\frac{1}{\beta_2}\right) \leq 0$ for large n. Then the proof is complete. □

4. Additional Lemmas

In this section, several lemmas are given. Firstly, we present the Hanson-Wright inequality (Wu et al. (2023)) below.

Lemma 4.1 (Hanson-Wright). *Let $X, Y \in \mathbb{R}^d$ be standard Gaussion random variables such that (X_i, Y_i), $i = 1, 2, \ldots, d$ are independent and have correlation coefficient ρ. Then with probability at $1 - 2\delta$,*

$$|X^T Y - d\rho| \leq C\left(d\sqrt{\log\frac{1}{\delta}} + \log\frac{1}{\delta}\right),$$

for a constant $C > 0$.

The following lemma presents the Chernoff bound for binomial distribution (Wu et al. (2023)).

Lemma 4.2 (Chernoff bound). *Let $X \sim Bin(n, p)$ and $\mu = np$. Then for any $\delta > 0$,*

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{-\mu(1+\delta)\log(1+\delta)-\delta},$$

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta}{2}\mu}.$$

Particularly, for $\tau = \mu\exp\left(1 + W(\frac{t}{e\mu} - \frac{1}{\mu})\right)$ with $W(x)$ be the solution to the equation $f(x)e^{f(x)} = x$, then

$$\mathbb{P}(X \leq \tau) \leq e^{-t}.$$

The following lemma presents a fact about random permutation (Arratia and Tavaré (1992)).

Lemma 4.3. *Let n_t be the number of t-cycles in a random permutation $\sigma \in P_n$. Let $Z_t \sim Poisson(\frac{1}{t})$ be independent Poisson random variables.. Then*

$$\mathbb{E}(g(n_1, n_2, \ldots, n_L)) \leq e^{1+\frac{1}{2}+\cdots+\frac{1}{L}}\mathbb{E}(g(Z_1, Z_2, \ldots, Z_L)),$$

for any nonnegative function g.

5. Conclusion and Future Problems

In this paper, we study the problem of testing hypergraph correlation. We derive the sharp statistical testing limit for both Gaussian-Wigner uniform hypergraphs and dense Erdös-Rényi uniform hypergraphs. Below the limit, it is impossible to distinguish the alternative hypothesis from the null hypothesis. Above the limit, we construct tests that can achieve asymptotic power one.

We conclude the paper with several possible future topics: 1) Derive sharp detection limit for sparse Erdös-Rényi uniform hypergraphs. The main difficulty lies in the proof of the impossibility. The key step of the proof in Wu et al. (2023) is to analyze the pseudoforest structure of graphs. The pseudoforest structure of hypergraph is much more complex than the graph case. 2) Determine the sharp detection limit for heterogeneous hypergraphs or graphs. It is well-known that real networks are usually heterogeneous, that is, the degrees of nodes are not the same Ke et al. (2020); Jin et al. (2021); Gao et al. (2018). It is interesting to study how the degree heterogeneity changes the limit. 3) Investigate how side information affects the detection limit. Real networks usually have node covariates or other side information. It is shown that incorporating side information can improve the limit of community detection Mossel and Xu (2016); Deshpande et al. (2018); Weng and Feng (2022); Zhao et al. (2021). It is interesting to study whether similar result holds in testing graph or hypergraph correlation.

References

Arratia, R. and Tavaré, S. The cycle structure of random permutations. *Ann. Probab.*, **20** (3), 1567–1591 (1992). MR1175278.

Barak, B., Chou, C.-N., Lei, Z., et al. (Nearly) Efficient Algorithms for the Graph Matching Problem on Correlated Random Graphs. In Wallach, H. et al., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc. (2019). ISBN 9781713807933.

Berg, A., Berg, T., and Malik, J. Shape matching and object recognition using low distortion correspondences. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pp. 26–33 (2005). DOI: 10.1109/CVPR.2005.320.

Cho, M. and Lee, K. M. Progressive graph matching: Making a move of graphs via probabilistic voting. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 398–405 (2012). DOI: 10.1109/CVPR.2012.6247701.

Cour, T., Srinivasan, P., and Shi, J. Balanced Graph Matching. In Schölkopf, B. et al., editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press (2006).

Deshpande, Y., Sen, S., Montanari, A., and Mossel, E. Contextual Stochastic Block Models. In Bengio, S. et al., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc. (2018).

Ding, J., Ma, Z., Wu, Y., and Xu, J. Efficient random graph matching via degree profiles. *Probab. Theory Related Fields*, **179** (1-2), 29–115 (2021). MR4221654.

Duchenne, O., Bach, F., Kweon, I.-S., and Ponce, J. A Tensor-Based Algorithm for High-Order Graph Matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, **33** (12), 2383–2395 (2011). DOI: 10.1109/TPAMI.2011.110.

Estrada, E. and Rodriguez-Velazquez, J. A. Complex networks as hypergraphs. *ArXiv Mathematics e-prints* (2005). arXiv: physics/0505137.

Fan, Z., Mao, C., Wu, Y., and Xu, J. Spectral graph matching and regularized quadratic relaxations I Algorithm and Gaussian analysis. *Found. Comput. Math.*, **23** (5), 1511–1565 (2023a). MR4649430.

Fan, Z., Mao, C., Wu, Y., and Xu, J. Spectral graph matching and regularized quadratic relaxations II. *Found. Comput. Math.*, **23** (5), 1567–1617 (2023b). MR4649431.

Ganassali, L., Lelarge, M., and Massoulié, L. Spectral alignment of correlated Gaussian matrices. *Adv. in Appl. Probab.*, **54** (1), 279–310 (2022). MR4397868.

Ganassali, L. and Massoulié, L. From tree matching to sparse graph alignment. In Abernethy, J. and Agarwal, S., editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 1633–1665. PMLR (2020).

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. Community detection in degree-corrected block models. *Ann. Statist.*, **46** (5), 2153–2185 (2018). MR3845014.

Ghoshal, G., Zlatić, V., Caldarelli, G., and Newman, M. E. J. Random hypergraphs and their applications. *Phys. Rev. E (3)*, **79** (6), 066118, 10 (2009). MR2551286.

Ghoshdastidar, D. and Dukkipati, A. Consistency of Spectral Partitioning of Uniform Hypergraphs under Planted Partition Model. In Ghahramani, Z. et al., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc. (2014).

Hou, J., Yuan, H., and Pelillo, M. Game-theoretic hypergraph matching with density enhancement. *Pattern Recognition*, **133**, 109035 (2023). DOI: 10.1016/j.patcog.2022.109035.

Jin, J., Ke, Z. T., and Luo, S. Optimal adaptivity of signed-polygon statistics for network testing. *Ann. Statist.*, **49** (6), 3408–3433 (2021). MR4352535.

Ke, Z. T., Shi, F., and Xia, D. Community detection for hypergraph networks via regularized tensor power iteration. *ArXiv Mathematics e-prints* (2020). arXiv: 1909.06503.

Korula, N. and Lattanzi, S. An efficient reconciliation algorithm for social networks. *Proceedings of the VLDB Endowment*, **7** (5), 377–388 (2014). See also arXiv: 1307.1690.

Lee, J., Cho, M., and Lee, K. M. Hyper-graph matching via reweighted random walks. In *CVPR 2011*, pp. 1633–1640 (2011). DOI: 10.1109/CVPR.2011.5995387.

Liao, X., Xu, Y., and Ling, H. Hypergraph Neural Networks for Hypergraph Matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1266–1275 (2021).

Mao, C., Wu, Y., Xu, J., and Yu, S. H. Testing network correlation efficiently via counting trees. *ArXiv Mathematics e-prints* (2021). arXiv: 2110.11816.

Mossel, E. and Xu, J. Local algorithms for block models with side information [extended abstract]. In *ITCS'16—Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 71–80. ACM, New York (2016). MR3629812.

Mossel, E. and Xu, J. Seeded graph matching via large neighborhood statistics. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1005–1014. SIAM, Philadelphia, PA (2019). MR3909531.

Newman, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E*, **64**, 016131 (2001). DOI: 10.1103/PhysRevE.64.016131.

Nguyen, Q., Tudisco, F., Gautier, A., and Hein, M. An Efficient Multilinear Optimization Framework for Hypergraph Matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (6), 1054–1075 (2017). DOI: 10.1109/TPAMI.2016.2574706.

Ouvrard, X., Goff, J.-M. L., and Marchand-Maillet, S. Networks of collaborations: hypergraph-modeling and visualisation. *ArXiv Mathematics e-prints* (2017). arXiv: 1707.00115.

Park, S., Park, S.-K., and Hebert, M. Fast and Scalable Approximate Spectral Matching for Higher Order Graph Matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, **36** (3), 479–492 (2014). DOI: 10.1109/TPAMI.2013.157.

Pedarsani, P. and Grossglauser, M. On the privacy of anonymized networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pp. 1235–1243. Association for Computing Machinery, New York, NY, USA (2011). ISBN 9781450308137. DOI: 10.1145/2020408.2020596.

Ramasco, J. J., Dorogovtsev, S. N., and Pastor-Satorras, R. Self-organization of collaboration networks. *Phys. Rev. E*, **70**, 036106 (2004). DOI: 10.1103/PhysRevE.70.036106.

Vogelstein, J. T., Conroy, J. M., Lyzinski, V., et al. Fast Approximate Quadratic Programming for Graph Matching. *PLOS ONE*, **10** (4), 1–17 (2015). DOI: 10.1371/journal.pone.0121002.

Wang, R., Yan, J., and Yang, X. Neural Graph Matching Network: Learning Lawler's Quadratic Assignment Problem With Extension to Hypergraph and Multiple-Graph Matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (9), 5261–5279 (2022). DOI: 10.1109/TPAMI.2021.3078053.

Weng, H. and Feng, Y. Community detection with nodal information: likelihood and its variational approximation. *Stat*, **11**, Paper No. e428, 17 (2022). MR4394988.

Wu, Y., Xu, J., and Yu, S. H. Testing correlation of unlabeled random graphs. *Ann. Appl. Probab.*, **33** (4), 2519–2558 (2023). MR4612649.

Yu, L., Xu, J., and Lin, X. Graph matching with partially-correct seeds. *J. Mach. Learn. Res.*, **22**, Paper No. [280], 54 (2021). MR4353059.

Yuan, M., Liu, R., Feng, Y., and Shang, Z. Testing community structure for hypergraphs. *Ann. Statist.*, **50** (1), 147–169 (2022). MR4382012.

Zass, R. and Shashua, A. Probabilistic graph and hypergraph matching. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008). DOI: 10.1109/CVPR.2008.4587500.

Zhao, X., Zhao, W., and Yuan, M. Information Limits for Community Detection in Hypergraph with Label Information. *Symmetry*, **13** (11) (2021). DOI: 10.3390/sym13112060.