

Combinatorial-Probabilistic Trade-Off: P-Values of Community Properties Test in the Stochastic Block Models

Shuting Shen

SHS145@G.HARVARD.EDU

*Department of Biostatistics
Harvard School of Public Health
Boston, MA 02115, USA*

Junwei Lu

JUNWEILU@HSPH.HARVARD.EDU

*Department of Biostatistics
Harvard School of Public Health
Boston, MA 02115, USA*

Editor:

Abstract

In this paper, we propose an inferential framework testing the general community combinatorial properties of the stochastic block model. We aim to test the hypothesis on whether a certain community property is satisfied and provide p-values to assess the statistical uncertainty. For instance, we test whether a given set of nodes belong to the same community. We present a general framework applicable to all symmetric community properties. To ease the challenges caused by the combinatorial nature of communities properties, we develop a novel shadowing bootstrap testing method. Utilizing the symmetry, our method can find a shadowing representative of the true assignment and the number of assignments to be tested in the alternative can be largely reduced. In theory, we introduce a combinatorial distance between two community classes and show a combinatorial-probabilistic trade-off phenomenon. Our test is honest as long as the product of the combinatorial distance between two communities and the probabilistic distance between two assignment probabilities is sufficiently large. Besides, we show that such trade-off also exists in the information-theoretic lower bound of the community property test. We also implement numerical experiments on both the synthetic data and the protein interaction application to show the validity of our method.

Keywords: Combinatorial inference, stochastic block models, community properties, minimax lower bound.

1. Introduction

Clustering is an important feature for network studies, which refers to the presence of node communities in the underlying graph. Community partitions the nodes into subgroups, within which a higher level of connectivity is perceived. The broad spectrum of applications for inferring the network community include the fields of sociology (Wasserman and Faust, 1994), biology (Barabási and Oltvai, 2004), physics (Newman, 2003) and internet (Albert et al., 1999). Stochastic block model (SBM) (Holland et al., 1983) is one of the most widely studied statistical model to depict the community structures in networks. It is a random

graph model which divides the nodes into disjoint communities and assigns the probability of connection between two nodes according to their community memberships.

The main goal of our paper is to test the hypotheses on the general structural properties in SBM with the p-values to evaluate the statistical uncertainty of the tests. One of the central problem in the previous study of the stochastic block model is the community detection. However, most of existing research focused on estimating the community labeling without p-values assessing the uncertainty in inference (Choi et al., 2012; Airola et al., 2013; Mossel et al., 2012, 2018; Massoulié, 2014; Hajek et al., 2016; Abbe et al., 2016; Mossel et al., 2016). Some fundamental limits of the community recovery have also been established in the previous studies. For example, Abbe et al. (2016) showed the optimal phase transition for the exact recovery of the community assignments using the maximum likelihood. The semi-definite relaxation methods (Abbe et al., 2016; Hajek et al., 2016; Agarwal et al., 2017; Bandeira, 2018) and the spectral methods (Abbe and Sandon, 2015; Yun and Proutiere, 2014; Gao et al., 2017; Abbe et al., 2020) are also shown to be optimal in exact recovery. Besides the exact recovery, Zhang and Zhou (2016) quantified the statistical rate of the community estimation via the mis-match ratio and showed the minimax rate of the mis-match ratio for community detection.

The consistency of the community estimation has two major limits: 1) it does not provide the p-values to evaluate the uncertainty of the estimation, which are essential in many scientific applications, and 2) it requires the recovery of community assignments for all nodes, while in many scientific applications we are interested in the community properties of a specific subset of nodes. For instance, Tabouy et al. (2020) studied the ESR1 protein-protein interaction network in breast cancer and aimed to test if a given set of cancer-related proteins belongs to the same community. Another example is the application in human brain connectome: Faskowitz et al. (2018) studied whether two specific areas of brains belong to the same cluster. This reduces to the statistical hypothesis that if two sets of cerebral nodes belong to the same community. We can formulate the above applications as the following examples of statistical hypotheses.

Example 1.1 (Same community test for m nodes). We want to test whether m given nodes are in the same cluster or not. Without loss of generality, we have the hypothesis:

H_0 : Nodes $1, \dots, m$ belong to the same community,

H_1 : There exists two nodes $1 \leq j \neq k \leq m$ belonging to two different communities.

Example 1.2 (Group community test). Like the applications in human brain connectome (Faskowitz et al., 2018), we have two groups of nodes and within each group, we know in prior that they belong to the same community. We aim to further test whether these two groups of nodes belong to the same cluster. We denote one node set as $S_m = \{1, \dots, m\}$ and the other node set as $S_{m'} = \{m + 1, \dots, m + m'\}$. The group community hypothesis is

H_0 : Nodes in $S_m \cup S_{m'}$ belong to the same community,

H_1 : Nodes in S_m belong to community a , but nodes in $S_{m'}$ belong to community $b \neq a$.

Example 1.3 (Equal-sized communities test). Given an SBM of n nodes and K communities, we aim to test whether each community has the same size. Namely, we aim to test the

hypothesis:

H_0 : Each community has the size n/K ,

H_1 : At least one of the communities size is not equal to n/K .

In order to conduct hypothesis tests including the above examples, we develop a general inference community property test. We consider the SBM with n nodes and K communities. Denote the community assignment of the nodes as $z = (z(1), \dots, z(n)) \in \{1, \dots, K\}^n$ such that $z(j) = k$ implies that the j -th node belongs to the k -th community. In order to specify the true assignment, we assume z is deterministic. The homogeneous SBM assumes that the edges of the random graph are independent Bernoulli random variables, i.e., the probability of the nodes i and j being connected is p if $z(i) = z(j)$ and the probability of the nodes i and j being connected is q if $z(i) \neq z(j)$. Let $\mathcal{C}_0, \mathcal{C}_1 \subset \{1, \dots, K\}^n$ be two disjoint communities assignment families. We are interested in the *general community property test*:

$$H_0 : z \in \mathcal{C}_0 \text{ versus } H_1 : z \in \mathcal{C}_1. \quad (1.1)$$

The concrete examples of \mathcal{C}_0 and \mathcal{C}_1 are listed in Examples 1.1 - 1.3. We characterize the hardness of differentiating the null hypothesis from the alternative by two kinds of “distances”: the probabilistic distance: how close between p and q and the combinatorial distance: how close between \mathcal{C}_0 and \mathcal{C}_1 . The existing literature in the study of the community detection only focused on the probability distance, e.g., $\sqrt{p} - \sqrt{q}$ (Abbe and Sandon, 2015) or the Rényi divergence (Zhang and Zhou, 2016)

$$I(p, q) = -2 \log (\sqrt{pq} + \sqrt{(1-p)(1-q)}). \quad (1.2)$$

In comparison to these results, our paper introduce a novel combinatorial distance between \mathcal{C}_0 and \mathcal{C}_1 denoted as $d(\mathcal{C}_0, \mathcal{C}_1)$ measuring the number of misalignments between two families (we refer the exact definition to Definition 2.4). The main result of our paper is that, for a wide range of SBM models, we can propose a general testing method which is honest and powerful when

$$\text{Combinatorial-Probabilistic Trade-Off: } I(p, q)d(\mathcal{C}_0, \mathcal{C}_1) = \Omega(n^\epsilon) \quad (1.3)$$

for some arbitrarily small $\epsilon > 0$. On the other hand, we show the minimax lower bound of the test in the sense that H_0 and H_1 in (1.1) cannot be differentiated when $I(p, q)d(\mathcal{C}_0, \mathcal{C}_1) \leq c \log n$ for some constant $c > 0$.¹ The multiplication between $I(p, q)$ and $d(\mathcal{C}_0, \mathcal{C}_1)$ reveals the trade-off between the probabilistic distance and the combinatorial distance in the general community property test. Our paper makes the following specific contributions to achieve such trade-off.

- **Methodology.** We propose a likelihood ratio test for the community property test in (1.1). We also provide p-values for uncertainty assessment. We show that our test is generally honest and powerful as long as the tested community properties \mathcal{C}_0 and \mathcal{C}_1 are symmetric under community assignment permutation transforms, which covers all examples

1. We refer to Theorem 3.2 and Theorem 4.1 for the rigorous arguments about the upper and lower bounds.

above. Comparing to the likelihood ratio test on the community numbers (Wang and Bickel, 2017), our likelihood ratio test could be evaluated over a much larger family of community assignments such that the limiting distribution of our test statistic is no longer always normal. Therefore, the method of Wang and Bickel (2017) is no longer applicable and we need to develop a new multiplier bootstrap method to estimate the p-value of our statistic. To achieve this, there are two major challenges. Firstly, the possible assignments in the alternative space \mathcal{C}_1 are so large such that the multiplier bootstrap statistic in Chernozhukov et al. (2013) cannot be applied directly. To overcome this, instead of considering the entire alternative class \mathcal{C}_1 , our testing method shows that it suffices to consider the boundary of \mathcal{C}_1 , which will significantly reduce the computation complexity. Secondly, the naive multiplier bootstrap requires to know the true assignment. We propose a “shadowing bootstrap” method by utilizing the symmetry of \mathcal{C}_0 and \mathcal{C}_1 . Instead of using the true community assignment, we use a “shadowing assignment” in the bootstrap which remains to be valid due to the symmetry of the community properties.

- **Theory.** We show the validity and power of the proposed test when the probability and combinatorial distances satisfies the general relationship in (1.3). We also prove the minimax lower bound of the general community property test and show that the proposed test is nearly optimal. Due to the generality of the property test, the existing theoretical results on the community detection, e.g., Abbe et al. (2016) and Zhang and Zhou (2016), can not be directly applied. To derive the general lower bound, we take a set of the hardest assignments in the alternative \mathcal{C}_1 which are closest to \mathcal{C}_0 . These hardest assignments are dependent among each other and we control their dependency via comparing to the distribution of high dimensional Gaussian vectors. To the best of our knowledge, it is the first time for our paper to derive a minimax lower bound for general community properties.

1.1 Related Papers

There are several existing papers discussing the inference of the community properties of the stochastic block model. Bickel and Sarkar (2016) designed a recursive bipartitioning algorithm based on the test statistic derived from the principal eigenvalue of the normalized adjacency matrix to automatically determine the number of clusters k . Similarly, Lei (2016) developed a goodness-of-fit test based on the largest singular value of the residual matrix for estimating the number of communities. Also interested in inferring the number of the communities, Wang and Bickel (2017) employed a likelihood ratio statistic and showed the asymptotic normality of the proposed statistic. Based on the algorithm proposed by Mossel et al. (2016), an asymptotically powerful test can also be conducted for Example 1.1 when m is fixed. In comparison, our method allows m to increase up to the order of n . Compared to the above works that only focus on an exact community property, the hypotheses testing problem in our paper is more general. Under the framework of the mixed membership model, Fan et al. (2019) studied the one-sample test of the weight vector of the mixed membership via a singular value decomposition based method. It covers Example 1.1 for $m = 2$ in the stochastic block model, however, it cannot be directly applied to other examples mentioned above. In comparison, our method covers wider range of scenarios. Moreover, we also study the general lower bound of the community test which reveals a novel bridge between

the probability distance and the combinatorial distance between the null and alternative community families.

Besides inferring the number of communities, Gao et al. (2018) proposed a community detection algorithm in degree-corrected block models that can be reduced to the hypothesis tests on the membership of a given node when the cluster labels of other nodes are the truth. Gao et al. (2018) integrated the hypothesis tests as a technical procedure in community detection without providing any uncertainty assessment. Rich literature can also be found in the testing of underlying random graph models (Bubeck et al., 2016; Karwa et al., 2016; Gao and Lafferty, 2017; Ghoshdastidar et al., 2020, 2017; Tang et al., 2017a,b; Shumovskaia and Panov, 2018). Compared to those works where no uncertainty assessment is provided, our methods provides p-values to quantify the random errors.

1.2 Organization of the Paper

The rest of the paper is organized as follows. Section 2 provides the definitions and background knowledge that will be useful for inference on SBM models. At the end of Section 2, we introduce the inference method for community properties test with *symmetric* structures, where we mainly focus on SBM with even cluster sizes, and we provide concrete case studies to illustrate the method procedure. Theoretical results for the methods are developed in Section 3. In Section 4, we focus on the lower bound of the community property test, and in Section 5 we generalize our method to SBM with uneven cluster sizes. Finally in Section 6 and Section 7 we conduct numerical analysis both on synthetic data and real-world protein interaction data to evaluate the performance of our method.

Notations

We denote $|\cdot|$ to be the cardinality of a set. For two positive sequences x_n and y_n , we say $x_n = \Omega(y_n)$ if there exists a positive constant C not depending on n such that $x_n \geq Cy_n$ for all n sufficiently large. We say $x_n \lesssim y_n$ or $x_n = O(y_n)$ if $x_n \leq Cy_n$ for $C > 0$ not depending on n . We say $x_n \asymp y_n$ if $x_n \lesssim y_n$ and $y_n \lesssim x_n$. If $\lim_{n \rightarrow \infty} x_n/y_n = 0$ then we say $x_n = o(y_n)$.

2. Community Properties of the Stochastic Block Model

In our paper, we consider the fixed assignment stochastic block model. Recall that p and q are respectively the within-community and between-community probabilities. Denote $[n] = \{1, \dots, n\}$ for any integer n . For the stochastic block model with n nodes and K communities, the fixed assignment SBM assumes the community assignment $z = (z(1), \dots, z(n)) \in [K]^n$ is the prefixed parameter of the model. For simplicity, we start with considering the situation that the community sizes of the assignment are even. We denote the even assignment class $\mathcal{K}^n := \{z \in [K]^n : |\{i : z(i) = k\}| = n/K, \forall k \in [K]\}$. We will generalize our analysis to the uneven case in Section 5. We denote the fixed assignment stochastic block model as $\mathcal{M}(n, K, p, q, z)$. In our paper, we assume the number of communities K is bounded. Let $\mathbf{A} \in \{0, 1\}^{n \times n}$ be the symmetric adjacency matrix of the random graph generated from the SBM. We say $\mathbf{A} \sim \mathcal{M}(n, K, p, q, z)$ if the upper triangular entries of \mathbf{A} are independent Bernoulli random variable and $\mathbb{P}(\mathbf{A}_{ij} = \mathbf{A}_{ji} = 1) = p$ if $z(i) = z(j)$ and $\mathbb{P}(\mathbf{A}_{ij} = \mathbf{A}_{ji} = 1) = q$

if $z(i) \neq z(j)$ for any $i \neq j \in [n]$. In the following part of the paper, we will study the community property test with an observation of the adjacency matrix $\mathbf{A} \sim \mathcal{M}(n, K, p, q, z)$.

2.1 Symmetric Community Properties

In this section, we aim to define the community property and the distance between two community families. In general, we say a community property is a subset of $[K]^n$. However, such definition is too general and may include some ill-posed examples. For example, if we can transfer one assignment to another under certain permutation of the community labels, they are essentially the same assignment and should belong to the same community property. This motivates us to give the following definition of equivalent assignments.

Definition 2.1 (Equivalent community assignments). Let S_K be the symmetric group containing all bijections from $[K]$ to itself. We say two assignments z and $z' \in [K]^n$ are equivalent, denoted as $z \simeq z'$, if there exists a permutation $\sigma \in S_K$ such that $\sigma(z) = z'$. Here $\sigma(z)$ means implementing the permutation σ to each entry of the vector z . More generally, given a node set $\mathcal{N} \subseteq [n]$, we denote $z_{\mathcal{N}}$ as the sub-vector of z with entries in \mathcal{N} . We say $z_{\mathcal{N}} \simeq z'_{\mathcal{N}}$ if there exists a permutation $\sigma \in S_K$ such that $\sigma(z_{\mathcal{N}}) = z'_{\mathcal{N}}$.

With the concept of equivalent assignments, we can give the definition of symmetric community properties as follows.

Definition 2.2 (Symmetric community properties). We say a community property \mathcal{C}_0 is symmetric, if there exist a node set $\mathcal{N} \subseteq [n]$ and a specified assignment $\tilde{z} \in \mathcal{K}^n$, such that $\mathcal{C}_0 = \{z \in \mathcal{K}^n : z_{\mathcal{N}} \simeq \tilde{z}_{\mathcal{N}}\}$. We say some $\mathcal{C}_1 \subseteq \mathcal{K}^n \setminus \mathcal{C}_0$ is an alternative property of \mathcal{C}_0 if \mathcal{C}_1 is closed under permutations on the support \mathcal{N} , i.e., for any $z \in \mathcal{C}_1 \subseteq \mathcal{K}^n \setminus \mathcal{C}_0$, if some z' satisfies $z'_{\mathcal{N}} \simeq z_{\mathcal{N}}$, then $z' \in \mathcal{C}_1$ as well.

Intuitively, the node set \mathcal{N} and the assignment \tilde{z} in Definition 2.2 are the representative node set and assignment generating all possible assignments in the community property via permutation. The community property \mathcal{C}_0 is “symmetric” in the sense that all its assignments are equivalent on the support of node set \mathcal{N} . Therefore, we impose the following assumption on testing symmetric properties.

Assumption 2.1 (Symmetric community property test). In the hypothesis test $H_0 : z \in \mathcal{C}_0$ v.s. $H_1 : z \in \mathcal{C}_1$, we assume $\mathcal{C}_0, \mathcal{C}_1 \subseteq \mathcal{K}^n$ and \mathcal{C}_0 is symmetric and \mathcal{C}_1 is an alternative property of \mathcal{C}_0 .

By Definition 2.2, $\mathcal{C}_1 = \mathcal{C}_0^c$ is an alternative property of \mathcal{C}_0 . Meanwhile, \mathcal{C}_1 satisfying the assumption above could be a strict subset of \mathcal{C}_0^c , which allows more examples in practice. In fact, we can show that Examples 1.1 and 1.2 given in the introduction satisfies Assumption 2.1. We will give concrete forms of the representative node set \mathcal{N} and assignment \tilde{z} for these two examples below. Before going to the detailed discussion, we also want to remark that the assumption that $\mathcal{C}_0, \mathcal{C}_1 \subseteq \mathcal{K}^n$, i.e., the community sizes are even, is only for the simplicity of our statement. We will discuss the uneven cases of Examples 1.1 and 1.2 as well as Example 1.3 in Section 5.

• Example 1.1: Same community test for m nodes. For the null hypothesis that nodes $1, \dots, m$ belong to the same community, we can define

$$\mathcal{C}_0 = \{z \in \mathcal{K}^n : z(1) = \dots = z(m)\} \text{ and } \mathcal{C}_1 = \mathcal{K}^n \setminus \mathcal{C}_0. \quad (2.1)$$

Consider $\mathcal{N} = [m]$ and \tilde{z} is any assignment satisfying $\tilde{z}_{\mathcal{N}} = (1, \dots, 1) \in [K]^m$. As \tilde{z} represents an assignment whose first m nodes belong to one community, we can check that $\mathcal{C}_0 = \{z : z_{\mathcal{N}} \simeq \tilde{z}_{\mathcal{N}}\}$ and thus is symmetric.

• Example 1.2: Same community test for groups. Recall that the null hypothesis is that nodes $1, \dots, m, m+1, \dots, m+m'$ belong to the same community. Therefore, the null property is similar to Example 1.1. Following the same argument of the previous example, \mathcal{C}_0 is symmetric by choosing $\mathcal{N} = [m+m']$. On the other hand, the alternative hypothesis is different from the previous example. In fact, we have

$$\begin{aligned} \mathcal{C}_0 &= \{z \in \mathcal{K}^n : z(1) = \dots = z(m) = z(m+1) = \dots = z(m+m')\} \\ \mathcal{C}_1 &= \{z \in \mathcal{K}^n : z(1) = \dots = z(m) \neq z(m+1) = \dots = z(m+m')\} \end{aligned} \quad (2.2)$$

Notice \mathcal{C}_1 is a strict subset of $\mathcal{K}^n \setminus \mathcal{C}_0$. We can check \mathcal{C}_1 is an alternative property of \mathcal{C}_0 by Definition 2.2.

2.2 Combinatorial Distance Between Community Properties

The major difference between the two examples above is their alternative properties. With similar null properties, \mathcal{C}_1 in (2.1) is the complement of \mathcal{C}_0 , while in (2.2), \mathcal{C}_1 is a strict subset. From this perspective, the distance between the null and alternative hypotheses in Example 1.1 is smaller than Example 1.2. In other words, Example 1.2 is easier to test in comparison to Example 1.1. Therefore, in order to depict the relationship between \mathcal{C}_0 and \mathcal{C}_1 , we will propose a metric of distance in terms of the number of misaligned edges. We first define the set of misaligned edges between two assignments.

Definition 2.3. For any two assignments $z_0 \in \mathcal{C}_0$ and $z_1 \in \mathcal{C}_1$, we define the two sets of misaligned edges as

$$\begin{aligned} \mathcal{E}_1(z_0, z_1) &= \{(i, j) : i < j, i, j \in [n], z_0(i) = z_0(j), z_1(i) \neq z_1(j)\} \text{ and} \\ \mathcal{E}_2(z_0, z_1) &= \{(i, j) : i < j, i, j \in [n], z_0(i) \neq z_0(j), z_1(i) = z_1(j)\}, \end{aligned}$$

where $\mathcal{E}_1(z_0, z_1)$ contains the edges whose corresponding nodes are assigned to the same community in z_0 but to two different communities by z_1 , and $\mathcal{E}_2(z_0, z_1)$ is the opposite. See Figure 1 for an illustration. We denote $n_i(z_0, z_1) = |\mathcal{E}_i(z_0, z_1)|$, for $i = 1, 2$ as the cardinality of the two edge sets.

With the definition of misaligned edges, we are ready to propose the metric of assignment distance defined as follows.

Definition 2.4 (Community property distance). We define the distance between two assignments z_0 and z_1 as $d(z_0, z_1) = n_1(z_0, z_1) \vee n_2(z_0, z_1)$. Correspondingly, we also define

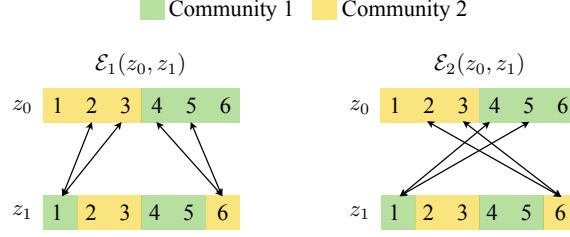


Figure 1: Example of misaligned edges in Definition 2.3.

$d(z_0, \mathcal{C}_1) = \inf_{z_1 \in \mathcal{C}_1} d(z_0, z_1)$ and the distance between two community properties

$$d(\mathcal{C}_0, \mathcal{C}_1) = \inf_{z_0 \in \mathcal{C}_0, z_1 \in \mathcal{C}_1} d(z_0, z_1).$$

By its definition, the distance $d(\mathcal{C}_0, \mathcal{C}_1)$ is the minimal number of misaligned edges between \mathcal{C}_0 and \mathcal{C}_1 . For instance, consider $\mathcal{C}_0 = \{z \in \mathcal{K}^n : z(1) = z(2)\}$ and $\mathcal{C}_1 = \{z \in \mathcal{K}^n : z(1) \neq z(2)\}$ with $K = 2$. On the one side, since the alignment of nodes 1 and 2 are different, we have $d(\mathcal{C}_0, \mathcal{C}_1) \geq |\mathcal{E}_1(z_0, z_1)| \geq n - 2$ given any $z_0 \in \mathcal{C}_0$ and $z_1 \in \mathcal{C}_1$. On the other hand, we can easily find two concrete examples of z_0, z_1 , illustrated in Figure 1, such that $d(z_0, z_1) = n - 2$. We refer the computation of $d(\mathcal{C}_0, \mathcal{C}_1)$ for more general examples to Section 2.4.1.

2.3 Likelihood-Ratio Test for Community Properties

Our method starts with defining a likelihood-ratio test statistic. We denote the observed adjacency matrix from the true model as $\mathbf{A} \sim \mathcal{M}(n, K, p, q, z^*)$, where z^* is the true assignment. The likelihood function of the stochastic block model is

$$f(\mathbf{A}; z, p, q) = \prod_{i < j} p^{\mathbb{1}(z(i)=z(j))\mathbf{A}_{ij}} (1-p)^{\mathbb{1}(z(i)=z(j))(1-\mathbf{A}_{ij})} q^{\mathbb{1}(z(i) \neq z(j))\mathbf{A}_{ij}} (1-q)^{\mathbb{1}(z(i) \neq z(j))(1-\mathbf{A}_{ij})}.$$

We then denote the log-likelihood ratio statistic as

$$\text{LRT} = \log \frac{\sup_{z \in \mathcal{C}_1} f(\mathbf{A}; z, p, q)}{\sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} f(\mathbf{A}; z, p, q)}. \quad (2.3)$$

In order to conduct the property test, we aim to study the limiting distribution of the likelihood ratio statistic. In specific, when $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$, where ρ_n is the signal strength, we are able to decompose the LRT as follows

$$\begin{aligned} \text{LRT} &= \sup_{z \in \mathcal{C}_1} \log f(\mathbf{A}; z, p, q) - \log f(\mathbf{A}; z^*, p, q) + o(1) \\ &= \sup_{z \in \mathcal{C}_1} g(p, q) \left(\sum_{(i,j) \in \mathcal{E}_2(z^*, z)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z)} \mathbf{A}_{ij} \right) + o(1), \end{aligned} \quad (2.4)$$

where $g(p, q) = \log p(1-q)/(q(1-p))$. The first equality above is due to the consistency of the maximum likelihood estimator and the second equality is derived via controlling the

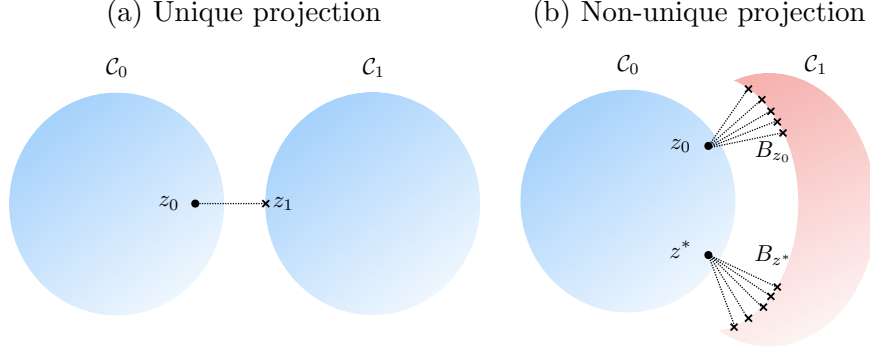


Figure 2: The boundary of \mathcal{C}_1 given z_0 . On the left panel, the projection is unique and $B_{z_0} = z_1$. On the right panel, the projection is not unique. z_0 is the shadowing assignment of z^* . B_{z^*} and B_{z_0} have similar structures.

remainder term. We defer the proof details to Appendix B.2. We observe that the leading term in (2.4) is the difference of edges in two edge sets: $\mathcal{E}_2(z^*, z)$ and $\mathcal{E}_1(z^*, z)$. By Definition 2.4, the property distance $d(\mathcal{C}_0, \mathcal{C}_1)$ is larger when the two edge sets are larger, which makes the leading term larger as well. This implies why $d(\mathcal{C}_0, \mathcal{C}_1)$ characterizes the difficulty of the test.

Remark 2.1. The likelihood ratio statistic in (2.3) is similar to the one proposed in Wang and Bickel (2017). They considered the hypothesis on a specific community property: the number of communities.

$$\mathcal{C}_0 = \{z | z \in [K-1]^n\} \text{ and } \mathcal{C}_1 = \{z | z \in [K]^n\} \setminus \mathcal{C}_0 \quad (2.5)$$

They show that the suprema $\sup_{z \in \mathcal{C}_1} f(\mathbf{A}; z, p, q)$ used in (2.3) is unique, as illustrated in Figure 2(a). This makes the LRT in Wang and Bickel (2017) asymptotically normal for $\mathcal{C}_0, \mathcal{C}_1$ in (2.5). However, this is not always true for the general community properties. For some properties, there will be an exponential number of candidate assignments maximizing the likelihood in (2.3), as illustrated in Figure 2(b). Thus the LRT is no longer asymptotically normal for the general case. Therefore, despite the similar formality of the likelihood ratio statistic comparing to the one in Wang and Bickel (2017), our testing procedure will be different from their method.

Since we want to characterize the suprema in the LRT, this motivates us to define the boundary of the alternative properties as follows.

Definition 2.5 (Boundary of communities class). For a given null assignment $z_0 \in \mathcal{C}_0$, we define the boundary of \mathcal{C}_1 as:

$$B_{z_0} = \{z \in \mathcal{C}_1 : d(z_0, z) = d(z_0, \mathcal{C}_1)\}$$

By the definition above, B_{z_0} is the projection of z_0 onto \mathcal{C}_1 using the distance in Definition 2.4. As we explained in Remark 2.1, the maximizer of the likelihood in \mathcal{C}_1 might not be

unique. Our analysis shows that the likelihood maximizer is asymptotically equivalent to the boundary B_{z^*} .² The later scenario will be more challenging. In the next section, we will study the asymptotic property of the LRT under such case by studying the structure of B_{z_0} .

2.4 Shadowing Bootstrap for the Property Test

In this section, we propose a bootstrap method to estimate the asymptotic p-value for the likelihood ratio statistic in (2.3).

There are two major challenges to estimate the p-value of LRT. First, the suprema of the likelihood is not unique and therefore the limiting distribution of LRT is not necessarily normal. By (2.4), we can in turn study the limiting distribution of the leading term

$$L := \sup_{z \in \mathcal{C}_1} g(p, q) \left(\sum_{(i,j) \in \mathcal{E}_2(z^*, z)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z)} \mathbf{A}_{ij} \right), \quad (2.6)$$

which is a maximum of a sequence of empirical processes indexed by z . Chernozhukov et al. (2013) studied the limiting distribution of the maximal of high dimensional empirical process and proposed to estimate its quantile by multiplier bootstrap. However, their method restricts the scaling condition that the dimension d of the empirical process and the sample size n satisfies $\log d/n^{1/5} = o(1)$, whereas in (2.6), the dimension $d = |\mathcal{C}_1|$ could be of the order K^n and violates the scaling condition. To handle such problem, our key observation is that the supreme over the alternative \mathcal{C}_1 can be represented by the supreme over its boundary B_{z^*} . In particular, we show that the leading term L is asymptotically the same as the following statistic:

$$L_0 := \sup_{z \in B_{z^*}} g(p, q) \left(\sum_{(i,j) \in \mathcal{E}_2(z^*, z)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z)} \mathbf{A}_{ij} \right). \quad (2.7)$$

We refer to Section B.2 in the Appendix for the detailed proof. The cardinality of B_{z^*} is much smaller than the one of \mathcal{C}_1 . In Table 1, we can see that for example, $|B_{z^*}|$ is of the order polynomial to n and therefore satisfies the scaling condition of high dimensional multiplier bootstrap.

Although B_{z^*} is much smaller than \mathcal{C}_1 , we cannot construct B_{z^*} in practice as z^* is unknown. This leads to the second challenge: how to find B_{z^*} in practice? Our key insight to solve the second challenge is to utilize the symmetry property in Definition 2.2. This insight relies on the following lemma characterizing the covariance of two processes.

Lemma 2.2 (Shadowing symmetry). For a given $z \in \mathcal{C}_0$, we list the assignments in the boundary B_z as $z_1, z_2, \dots, z_{|B_z|}$. Define a $|B_z|$ -dimensional vector \mathbf{L}_z as

$$(\mathbf{L}_z)_k = g(p, q) \left(\sum_{(i,j) \in \mathcal{E}_2(z, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z, z_k)} \mathbf{A}_{ij} \right), \text{ for } k = 1, 2, \dots, |B_z|.$$

2. See Lemma B.2 in the Appendix for the rigorous argument.

Suppose Assumption 2.1 holds. For any $z_0, z'_0 \in \mathcal{C}_0$, we have $|B_{z_0}| = |B_{z'_0}|$ and $\text{Cov}(\mathbf{L}_{z_0})$ equals to $\text{Cov}(\mathbf{L}_{z'_0})$ up to permutation, i.e., there existing a permutation $\tau \in S_{|B_{z_0}|}$ such that $\text{Cov}(\mathbf{L}_{z_0})_{kl} = \text{Cov}(\mathbf{L}_{z'_0})_{\tau(k)\tau(l)}$ for all $k, l = 1, \dots, |B_{z_0}|$.

We refer to Section D.1 in the Appendix for proof of Lemma 2.2. We call the above lemma as the shadowing symmetry lemma, because it implies that the covariance of \mathbf{L}_{z_0} is same up to permutation to any other “shadowing assignment” $z'_0 \in \mathcal{C}_0$. Therefore, we can avoid directly constructing B_{z^*} . Instead, we can choose any $z \in \mathcal{C}_0$ as a “shadowing assignment” and consider the shadowing statistic

$$L_0(z_0) := \sup_{z \in B_{z_0}} g(p, q) \left(\sum_{(i,j) \in \mathcal{E}_2(z_0, z)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z_0, z)} \mathbf{A}_{ij} \right), \quad (2.8)$$

illustrated in Figure 2(b). Applying Lemma 2.2, the following proposition shows that the quantile of $L_0(z_0)$ is asymptotically same as the quantile of L_0 .

Proposition 2.3. Suppose Assumption 2.1 holds, $\log |B_{z^*}| = O(\log n)$ and $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$. For any $z_0 \in \mathcal{C}_0$, we have

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |\mathbb{P}(L_0 < t) - \mathbb{P}(L_0(z_0) < t)| = 0.$$

We defer the proof to Appendix B.1. Now we are ready to present the shadowing bootstrap procedure. Based on the previous discussion, we aim to estimate the p-value based on $L_0(z_0)$. To achieve this, we take an arbitrary $z_0 \in \mathcal{C}_0$, and generate one realization of the adjacency matrix $\hat{\mathbf{A}} \sim \mathcal{M}(n, K, \hat{p}, \hat{q}, z_0)$. Here \hat{p} and \hat{q} are the maximum likelihood estimator

$$(\hat{p}, \hat{q}) = \underset{(p,q)}{\text{argmax}} \sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} f(\mathbf{A}; z, p, q) \quad (2.9)$$

The likelihood ratio statistic is

$$\widehat{\text{LRT}} = \log \sup_{z \in \mathcal{C}_1} f(\mathbf{A}; z, \hat{p}, \hat{q}) - \log \sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} f(\mathbf{A}; z, \hat{p}, \hat{q}). \quad (2.10)$$

The next step is to find the assignments in the boundary B_{z_0} . We can construct B_{z_0} by Definition 2.5 in general. We refer to Section 2.4.1 on how to construct B_{z_0} for the concrete examples. To estimate the p-value, we apply the Gaussian multiplier bootstrap. Let $\{e_{ij}\}_{1 \leq i < j \leq n}$ be independent standard Gaussian random variables and define

$$W_n = \sup_{z \in B_{z_0}} \sum_{1 \leq i < j \leq n} (\hat{\mathbf{A}}_{ij} - \mathbb{E}_{\hat{p}, \hat{q}}(\hat{\mathbf{A}}_{ij})) (\mathbb{1}[(i, j) \in \mathcal{E}_2(z_0, z)] - \mathbb{1}[(i, j) \in \mathcal{E}_1(z_0, z)]) e_{ij}, \quad (2.11)$$

where $\mathbb{E}_{\hat{p}, \hat{q}}(\hat{\mathbf{A}}_{ij}) = \hat{p}$ if $z_0(i) = z_0(j)$ and $\mathbb{E}_{\hat{p}, \hat{q}}(\hat{\mathbf{A}}_{ij}) = \hat{q}$ otherwise.

Then we estimate the p-value via the multiplier bootstrap

$$p_W = \mathbb{P}(g(\hat{p}, \hat{q})W_n + g(\hat{p}, \hat{q})\hat{\mu}_0 \geq \widehat{\text{LRT}} | \hat{\mathbf{A}}, \mathbf{A}), \quad (2.12)$$

where $\hat{\mu}_0 = d(\mathcal{C}_0, \mathcal{C}_1)(\hat{q} - \hat{p})$ is the estimator of the mean of the process in (2.8). Finally, we reject the null $H_0 : z^* \in \mathcal{C}_0$ if $p_W \leq \alpha$ and do not reject H_0 otherwise.

2.4.1 CASE STUDY OF THE BOUNDARY

In this part, we provide concrete algorithm to construct the boundary B_{z_0} for Examples 1.1 and 1.2. We can find B_{z_0} via computationally efficient algorithms for both examples. Meanwhile, we will also calculate $d(\mathcal{C}_0, \mathcal{C}_1)$ needed in (2.12).

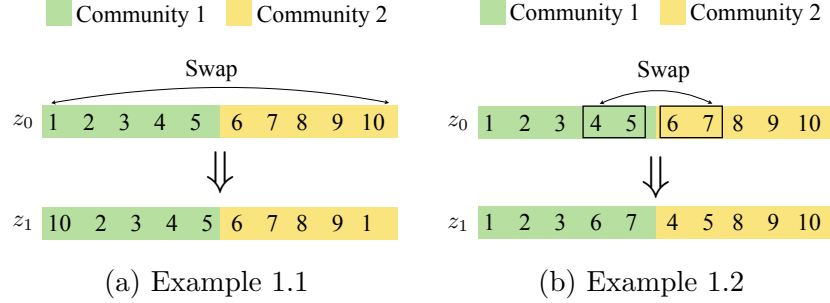


Figure 3: Procedure to construct the assignment in the boundary B_{z_0} . Panel (a) is to test whether the first 3 nodes belong to the same cluster, and Panel (b) is to test whether node sets $\{1, 2, 3\}$ and $\{4, 5\}$ belong to the same cluster.

- Example 1.1: Same community test for m nodes. Recall that \mathcal{C}_0 and \mathcal{C}_1 are defined in (2.1). For any $z_0 \in \mathcal{C}_0$, to find assignments in B_{z_0} , we aim to find assignments whose distance to z_0 is $d(z_0, \mathcal{C}_1)$. The simplest way is to exchange the community assignment of one node $s \in [m]$ with another node s' from a different community (see z_1 in Figure 3(a) for an example when $n = 10$, $m = 3$, and $K = 2$). It is easy to check all such assignments belong to B_{z_0} . On the other hand, any other operation will incur more node-wise misclassification and the edge-wise misalignment will be much larger. To find $d(\mathcal{C}_0, \mathcal{C}_1)$, we start with evaluating $n_1(z_0, z_1)$ for some $z_1 \in B_{z_0}$. The edges whose connection probability is changed from p to q will be the edges between node s and the rest of the nodes in its original community and between node s' and the rest of the nodes in its original community. Therefore, we have $n_1(z_0, z_1) = 2(n/K - 1)$. Similarly, $n_2(z_0, z_1) = 2(n/K - 1)$ and thus $d(\mathcal{C}_0, \mathcal{C}_1) = 2(n/K - 1)$.

In summary, B_{z_0} is composed of all the assignments which can be obtained from reassigning the label of one of z_0 's node in $[m]$ to a different community. The distance between two classes is $d(\mathcal{C}_0, \mathcal{C}_1) = 2(n/K - 1)$.

- Example 1.2: Same community test for groups. Recall that \mathcal{C}_0 and \mathcal{C}_1 are defined in (2.2). In this example, without loss of generality we can assume that $m' \leq m$. Then to project an arbitrary $z_0 \in \mathcal{C}_0$ onto \mathcal{C}_1 , we will exchange the cluster assignment of the set $\mathcal{S}_{m'} = \{m+1, \dots, m+m'\}$ with another set $\mathcal{S}'_{m'}$ from a different cluster of cardinality m' to obtain the smallest number of edges that are misaligned. See Figure 3(b). Correspondingly, for $z_1 \in B_{z_0}$, we have $d(z_0, z_1) = n_1(z_0, z_1) = n_2(z_0, z_1) = 2m \wedge m'(n/K - m \wedge m')$, and thus $d(\mathcal{C}_0, \mathcal{C}_1) = 2m \wedge m'(n/K - m \wedge m')$.

In summary, B_{z_0} is composed of all the assignments which can be obtained from reassigning the label of nodes $m + 1, \dots, m + m'$ in z_0 . The distance between two classes is $d(\mathcal{C}_0, \mathcal{C}_1) = 2m \wedge m'(n/K - m \wedge m')$.

3. Validity of Community Property Test

In this section, we show the theoretical results that our testing method is honest and powerful. Before presenting our theorems, we first give the following assumption for the alternative class \mathcal{C}_1 .

Assumption 3.1 (Scattering of \mathcal{C}_1). For any $z_0 \in \mathcal{C}_0$, we have $|B_{z_0}| = O(n^{c_0})$ for some constant $c_0 > 0$.

Remark 3.1. We call this assumption as the *scattering assumption* as it ensures that the assignments in \mathcal{C}_1 are uniformly scattered in \mathcal{C}_1 and there are not too many assignments concentrating on the boundary. In specific, we assume the cardinality of the boundary B_{z_0} is at most polynomial to n . In Section 2.4.1, we construct B_{z_0} for Examples 1.1 and 1.2 and they both satisfy this assumption. We refer to Proposition 3.3 or Table 1 for the specific rates of $|B_{z_0}|$ under each example.

Recall that p_W in (2.12) is our estimate of p-value of the property test. We reparameterize p, q as $p = \rho_n \lambda_1$ and $q = \rho_n \lambda_2$, where we assume λ_1 and λ_2 are constants independent to n . The following main theorem shows that our test is honest and powerful for general symmetric community properties.

Theorem 3.2. Suppose Assumptions 2.1 and 3.1 hold, $d(\mathcal{C}_0, \mathcal{C}_1) = o(n^{c_1})$ for some constant $c_1 < 2$, and $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$. We have

$$\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(p_W \leq \alpha) = \alpha \text{ and } \lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\text{reject } H_0) = \alpha.$$

Moreover, if $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \Omega(n^\varepsilon)$ for some arbitrarily small constant $\varepsilon > 0$, we have

$$\lim_{n \rightarrow \infty} \inf_{z^* \in \mathcal{C}_1} \mathbb{P}(\text{reject } H_0) = 1.$$

Remark 3.2. We defer the proof of theorem to Appendix B.2. One may note that under the regime of Theorem 3.2, exact recovery of the true assignment can be achieved. However, previous perfect recovery methods cannot provide p-values or uncertainty assessment (Yun and Proutiere, 2014; Abbe et al., 2016; Abbe and Sandon, 2015; Hajek et al., 2016; Abbe et al., 2020; Gao et al., 2017; Agarwal et al., 2017; Bandeira, 2018). On the other hand, our approach can provide the p-values for the general community properties which are more challenging comparing to estimating the community labeling. Taking Example 1.1 for instance, if one is only interested in deciding whether the true label belongs to \mathcal{C}_0 and \mathcal{C}_1 without uncertainty assessment, we can impose a weaker scaling assumption $\rho_n = \log n/n$ and $1 < \sqrt{\lambda_1} - \sqrt{\lambda_2} < \sqrt{2}$, under which we are able to differentiate \mathcal{C}_0 and \mathcal{C}_1 in the Example 1.1 via the singular value decomposition (Abbe et al., 2020).

The scaling condition $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \Omega(n^\varepsilon)$ in the theorem demonstrates the *combinatorial-probabilistic* trade-off in the community property test. In order to differentiate two community

properties \mathcal{C}_0 versus \mathcal{C}_1 , we know that both the combinatorial distance between \mathcal{C}_0 and \mathcal{C}_1 and the distance between two assignment probabilities p and q should be large enough. Theorem 3.2 implies that the combinatorial distance can be measured by $d(\mathcal{C}_0, \mathcal{C}_1)$ and the probabilistic distance can be measured by the Rényi divergence $I(p, q)$. Our test is powerful if the product of two distance increases faster than n^ε for some arbitrarily small constant $\varepsilon > 0$.

We now apply Theorem 3.2 to Examples 1.1 and 1.2 by checking Assumption 3.1. Due to the discussion in Section 2.4.1, for Example 1.1, B_{z_0} is composed of all the assignments in \mathcal{C}_1 that can be obtained by swapping one node in $[m]$ with another node from a different cluster (see Figure 3(a) for illustration). Therefore, $|B_{z_0}| = m(n/K)(K-1) = O(mn)$. For Example 1.2, without loss of generality, we assume $m' \leq m$. Therefore, B_{z_0} is composed of all the assignments which can be obtained from reassigning the label of nodes $m+1, \dots, m+m'$ in z_0 (see Figure 3(b) for illustration). Therefore, $|B_{z_0}| = (K-1) \binom{n/K}{m \wedge m'} = O(K(n/K)^{m \wedge m'})$. The $\binom{n/K}{m \wedge m'}$ term is for choosing the set $\mathcal{S}'_{m'}$ from a different community. Combining the discussion on $d(\mathcal{C}_0, \mathcal{C}_1)$ in Section 2.4.1, we summarize the results in the following proposition.

Proposition 3.3. For Examples 1.1, we have $|B_{z_0}| = O(mn)$ and thus it satisfies Assumption 3.1. We also have $d(\mathcal{C}_0, \mathcal{C}_1) = 2(n/K - 1) = O(n/K)$. For Examples 1.2, we have $|B_{z_0}| = O(K(n/K)^{m \wedge m'})$ which satisfies Assumption 3.1. We also have $d(\mathcal{C}_0, \mathcal{C}_1) = 2(m \wedge m')(n/K - m \wedge m')$.

Plugging these results to the general Theorem 3.2, we have the following corollary.

Corollary 3.4 (Examples 1.1 and 1.2). Suppose $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$. For \mathcal{C}_0 and \mathcal{C}_1 in (2.1) for any $m \leq n/K$ or (2.2) for $m \wedge m' = O(1)$, our test for the hypothesis $H_0 : z^* \in \mathcal{C}_0$ versus $H_1 : z^* \in \mathcal{C}_1$ is honest, i.e.,

$$\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\text{reject } H_0) = \alpha.$$

Moreover, if $I(p, q)n/K = \Omega(n^\varepsilon)$ for some small positive constant ε , we have

$$\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_1} \mathbb{P}(\text{reject } H_0) = 1.$$

4. Information-Theoretic Lower Bound

We have shown that our shadowing bootstrap method is honest and powerful when the product of the combinatorial distance and probabilistic distance satisfies $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \Omega(n^\epsilon)$ for some small $\epsilon > 0$. In this section, we will discuss the information-theoretic lower bound of community property test. We will give the lower bound of the minimax risk of all possible test ψ for $H_0 : z \in \mathcal{C}_0$ v.s. $H_1 : z \in \mathcal{C}_1$, defined as

$$r(\mathcal{C}_0, \mathcal{C}_1) = \inf_{\psi} \left\{ \sup_{z \in \mathcal{C}_0} \mathbb{P}_z(\psi = 1) + \sup_{z \in \mathcal{C}_1} \mathbb{P}_z(\psi = 0) \right\}.$$

We will show that the combinatorial-probabilistic trade-off phenomenon appears in the lower bound as well, thus it essentially characterizes the hardness of the community property test.

4.1 Packing Number of Communities

In order to establish the lower bound, we first introduce the concept of packing number of community class \mathcal{C}_1 . Similar to the minimax theory of the hypothesis testing for continuous parameters or the graph properties (Yang and Barron, 1999; Neykov et al., 2019), we find that the packing number is also essential in the lower bound of community properties test.

A key element in the definition of the packing number is the metric assigned to the community class \mathcal{C}_1 . Recall the community property distance $d(\mathcal{C}_0, \mathcal{C}_1)$ in Definition 2.4. It counts the misaligned edges $\mathcal{E}_1(z_0, z_1)$ and $\mathcal{E}_2(z_0, z_1)$ in Definition 2.3 for all $z_0 \in \mathcal{C}_0$ and $z_1 \in \mathcal{C}_1$. Our first insight is that the more misaligned edges there are, the easier it is to differentiate \mathcal{C}_1 from \mathcal{C}_0 . This motivates us to consider the misaligned edge set $\mathcal{E}_{1,2}(z_0, z_1) = \mathcal{E}_1(z_0, z_1) \cup \mathcal{E}_2(z_0, z_1)$ and use its cardinality as a “metric” in the following definition of packing number. Our second insight is that how hard it is to differentiate \mathcal{C}_1 from \mathcal{C}_0 does not depends on the complexity of the entire set \mathcal{C}_1 but the boundary set B_{z_0} in Definition 2.5. Our shadowing bootstrap statistic in (2.11) implies that B_{z_0} is representative to \mathcal{C}_1 . Therefore, we give the following definition of packing number of B_{z_0} to characterize the hardness of test.

Definition 4.1 (ε -packing of B_{z_0}). For any $z_0 \in \mathcal{C}_0$, we say $\{z_1, z_2, \dots, z_N\} \subseteq B_{z_0}$ is an ε -packing of B_{z_0} , if for any $z_j \neq z_k$ we have $|\mathcal{E}_{1,2}(z_0, z_j) \cap \mathcal{E}_{1,2}(z_0, z_k)| \leq \varepsilon$. The ε -packing number of B_{z_0} , denoted as $N(B_{z_0}, \varepsilon)$, is the maximum cardinality of any ε -packing of B_{z_0} .

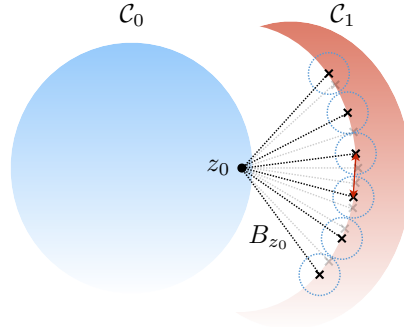


Figure 4: The ε -packing set of B_{z_0} .

We illustrate the packing set of B_{z_0} in Figure 4. By Definition 2.5, B_{z_0} collects the alignments in \mathcal{C}_1 which are closest to $z_0 \in \mathcal{C}_0$. Therefore, these alignments are the hardest cases to test. The following theorem shows the lower bound of the community property test can be characterized by the packing number of these hardest cases.

Theorem 4.1. Suppose $\mathcal{C}_0, \mathcal{C}_1 \subseteq \mathcal{K}^n$, $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$ and $p \leq 1 - \delta$ for some constant $\delta > 0$. If there exists a $z_0 \in \mathcal{C}_0$ such that $\log N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}) = O(\log n)$ and

$$\limsup_{n \rightarrow \infty} \frac{d(z_0, \mathcal{C}_1) I(p, q)}{\log N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)})} < 1, \quad (4.1)$$

then $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

Remark 4.1. We defer the proof of the theorem to Appendix C.1. The combinatorial-probabilistic trade-off in the lower bound is characterized by (4.1). We cannot differentiate \mathcal{C}_1 from \mathcal{C}_0 if

$$d(z_0, \mathcal{C}_1)I(p, q) < \log N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}),$$

for sufficiently large n . The packing entropy $\log N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)})$ is the lower bound of the signal strength. Our lower bound shows that the packing entropy of community class plays a similar role as the packing entropy in parametric hypothesis test (Yang and Barron, 1999) and in graph property test (Neykov et al., 2019). We derive the rate of packing entropy for Examples 1.1 and 1.2 in Proposition 4.3. In general, the packing entropy is $O(\log n)$. Comparing to the upper bound $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \Omega(n^\varepsilon)$ for some arbitrarily small constant $\varepsilon > 0$ in Theorem 3.2, there is a gap to $O(\log n)$ in the lower bound. We conjecture that this gap exists as both our upper and lower bounds are for general community property test. We will find a finer analysis in future research.

The following theorem gives an alternative lower bound result relaxing the scaling conditions in Theorem 4.1.

Theorem 4.2. Suppose $\mathcal{C}_0, \mathcal{C}_1 \subseteq \mathcal{K}^n$, $0 < q < p \leq 1 - \delta$ for some constant $\delta > 0$ and $\lim_{n \rightarrow \infty} d(\mathcal{C}_0, \mathcal{C}_1)p = \infty$. If one of the following conditions:

- (1) $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) \leq c$ for some sufficiently small constant c ;
- (2) $\lim_{n \rightarrow \infty} d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \infty$, but there exists a $z_0 \in \mathcal{C}_0$ such that

$$\limsup_{n \rightarrow \infty} \frac{d(z_0, \mathcal{C}_1)I(p, q)}{\log N(B_{z_0}, 0)} < 1,$$

is satisfied, then $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

We defer the proof of the theorem to Appendix C.1. Notice that the scaling condition on $d(z_0, \mathcal{C}_1)p$ is different from the one on $d(z_0, \mathcal{C}_1)I(p, q)$ in the lower bound. $I(p, q)$ measures the difference between p and q , whereas the condition $d(z_0, \mathcal{C}_1)p = \Omega(\log n)$ is to guarantee that the edge connection probability cannot be too small. Theorems 4.1 and 4.2 both show the lower bound with the combinatorial-probabilistic trade-off. Theorem 4.1 has a sharper lower bound on $d(z_0, \mathcal{C}_1)I(p, q)$ under a stronger scaling condition. In comparison, Theorem 4.2 has a less sharp lower bound with weaker scaling conditions. When $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q)$ is bounded, we cannot differentiate two hypotheses. When $d(\mathcal{C}_0, \mathcal{C}_1)$ goes to infinity, Theorem 4.2 condition (2) shows the lower bound $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) < \log N(B_{z_0}, 0)$. If we have stronger scaling conditions in Theorem 4.1, we get a sharper lower bound $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) < \log N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)})$.

4.2 Case Study of Lower Bound

In this section we apply the general theorems for the lower bound to Examples 1.1 and 1.2. By (4.1) in Theorem 4.1, a key quantity for the lower bound is the packing number $N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)})$. The following proposition gives concrete results for the two examples.

	$d(\mathcal{C}_0, \mathcal{C}_1)$	$ B_{z_0} $	$N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)})$
Example 1.1	$2(n/K - 1)$	$O(mn)$	m
Example 1.2	$2m \wedge m'(n/K - m \wedge m')$	$O(K(n/K)^{m \wedge m'})$	1

Table 1: Important values for even cases of Example 1.1 and Example 1.2

Proposition 4.3. The packing number for Example 1.1 is $N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}) = m$, and the packing number for Example 1.2 is $N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}) = N(B_{z_0}, 0) = 1$.

The proof of the proposition is deferred to Appendix A.1. Table 1 summarizes important quantities for Examples 1.1 and 1.2.

Recall that $\lambda_1 = p/\rho_n$ and $\lambda_2 = q/\rho_n$. We present the lower bound of two examples below. Applying Theorem 4.1 and Proposition 4.3, we have the following lower bound of same community test in Example 1.1.

Corollary 4.4. For \mathcal{C}_0 and \mathcal{C}_1 defined in (2.1), if $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$, $p < 1 - \delta$ for some constant $\delta > 0$ and

$$\limsup_{n \rightarrow \infty} 2nI(p, q)/(K \log m) < 1,$$

we have $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

Remark 4.2. Similar results on nodes classification can be found in Mossel et al. (2012) and Mossel et al. (2016). However, in our example, the number of nodes m is allowed to increase with n , whereas in Mossel et al. (2012) m was fixed.

Applying Theorem 4.2 and Proposition 4.3, we have the following lower bound of same community test for groups in Example 1.2.

Corollary 4.5. For \mathcal{C}_0 and \mathcal{C}_1 defined in (2.2), if $np \rightarrow \infty$, $0 < q < p \leq 1 - \delta$ for some constant $\delta > 0$ and

$$\limsup_{n \rightarrow \infty} nI(p, q) < c,$$

for some sufficiently small constant $c > 0$, we have $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

5. General Framework for Uneven Community Sizes

In this section, we generalize our theory to the community property tests when the community sizes in \mathcal{C}_0 and \mathcal{C}_1 are not necessarily even, e.g., Example 1.3. For any $z \in \mathcal{C}_0 \cup \mathcal{C}_1$, denote the community size $n_k(z) = |\{z(i) = k \mid i \in [n]\}|$ for $k \in [K]$. Let

$$c_K = \max_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} \max_{1 \leq k \leq K} |n_k(z) - n/K|. \quad (5.1)$$

When the community sizes are even, we have $c_K = 0$. In this section, we consider the cases when c_K could be larger than zero. We will show that the shadowing bootstrap method in Section 2.4 can be applied to test the uneven community property as well. The information-theoretic lower bound is also similar to the one in Section 4.

5.1 General Symmetric Community Properties

For the uneven community class, we still need some symmetry property for the assignments in \mathcal{C}_0 and \mathcal{C}_1 . When community sizes are even, Definition 2.2 depicts the symmetry via the representative node set \mathcal{N} and the representative assignment \tilde{z} . However, for many community properties of interest, e.g., the community size test in Example 1.3, we cannot find such \mathcal{N} and \tilde{z} . In Example 1.3, we are interested in testing the community size and thus there is no representative nodes. See Figure 5 for illustration.

Therefore, we define the following generalized symmetric community property pair.

Definition 5.1 (Generalized symmetric community property pair). We say two disjoint community properties \mathcal{C}_0 and \mathcal{C}_1 is a *generalized symmetric property pair* if for any $z, z' \in \mathcal{C}_0$, there exist permutations $\sigma \in S_K$ and $\tau \in S_n$ such that

- (1) $\tau \circ \sigma(z) := (\sigma(z(\tau(1))), \dots, \sigma(z(\tau(n)))) = z'$ and
- (2) \mathcal{C}_1 is also closed under such transform $\tau \circ \sigma$, i.e., for any $z'' \in \mathcal{C}_1$, $\tau \circ \sigma(z'') \in \mathcal{C}_1$.

Definition 5.1 generalizes the concept of symmetric community property in Definition 2.2 via introducing the permutation transform. We can check that Examples 1.1 and 1.2 are still symmetric by Definition 5.1. See Figure 5(a) for an example of choosing σ and τ . On the other hand, the community sizes properties

$$\mathcal{C}_0 = \{z \in [K]^n : \text{all community sizes} = n/K\} \text{ and } \mathcal{C}_1 = \mathcal{C}_0^c, \quad (5.2)$$

are also symmetric by Definition 5.1 but not Definition 2.2. See Figure 5(b) for illustration. In fact, the following proposition shows that Definition 2.2 is a special case of Definition 5.1.

Proposition 5.1. If $\mathcal{C}_0, \mathcal{C}_1 \subseteq \mathcal{K}^n$ satisfy Assumption 2.1, then \mathcal{C}_0 and \mathcal{C}_1 is a generalized symmetric property pair. Moreover, the property pairs in (2.1), (2.2) and (5.2) are generalized symmetric property pairs.

We defer the proof of the proposition to Appendix A.2. In Figure 5, we show how to choose concrete permutation transforms σ and τ for Examples 1.1 and 1.3.

5.2 Shadowing Bootstrap for General Case

We now generalize the testing method proposed in Section 2.4 to the uneven case. A key step is to generalize the boundary B_{z_0} in Definition 2.5. Recall that for the even case, our insight is that the statistic L in (2.6) taking the supremum over \mathcal{C}_1 is asymptotically equal to the L_0 in (2.7) taking the supremum over B_{z^*} , which is much smaller than \mathcal{C}_1 . Similar insight applies to the uneven case using the following generalized definition of boundary.

Definition 5.2. For a given $z_0 \in \mathcal{C}_0$, we define the boundary centered at z_0 with radius r as

$$B_{z_0}(r) = \{z \in \mathcal{C}_1 | d(z_0, z) \leq r\}.$$

We illustrate the two types of boundary in Figure 6. From Figure 6(a), we can see that $B_{z_0} = B_{z_0}(d(\mathcal{C}_0, \mathcal{C}_1))$. Therefore, Definition 5.2 is a generalization of Definition 2.5. For the

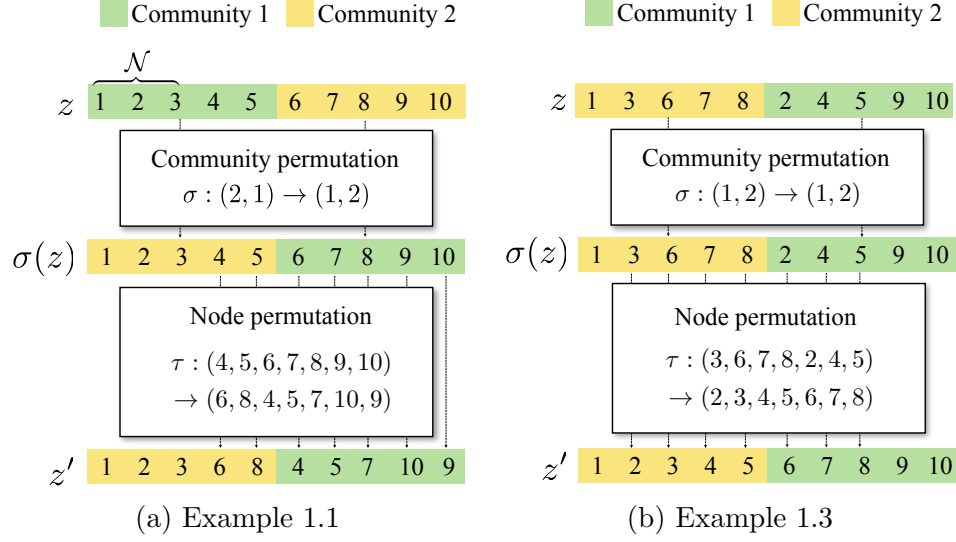
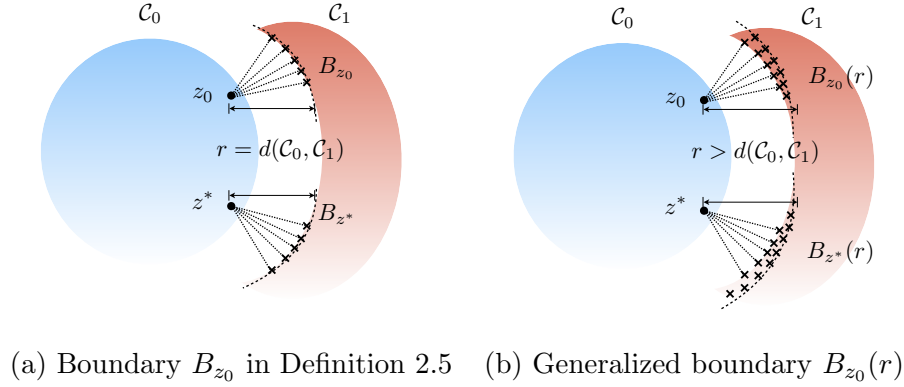


Figure 5: Permutation of null assignments in Example 1.1 and Example 1.3


 Figure 6: The boundary B_{z_0} defined previously for even cases is in essence a ball centered at z_0 with radius $r = d(C_0, C_1)$

uneven case, L_0 is no longer asymptotically equal to L . We need to enlarge B_{z_0} to $B_{z_0}(r)$ for some $r > d(\mathcal{C}_0, \mathcal{C}_1)$ and modify the statistic L_0 in (2.7) by taking the supremum over $B_{z^*}(r)$.

In fact, we can still use the shadowing bootstrap method in Section 2.4 to the uneven case. All procedures are exactly same as Section 2.4 except that we only need to replace the bootstrap statistic W_n in (2.11) by

$$W_n = \sup_{z \in B_{z_0}(r)} \sum_{1 \leq i < j \leq n} (\hat{\mathbf{A}}_{ij} - \mathbb{E}_{\hat{p}, \hat{q}}(\hat{\mathbf{A}}_{ij})) (\mathbb{1}[(i, j) \in \mathcal{E}_2(z_0, z)] - \mathbb{1}[(i, j) \in \mathcal{E}_1(z_0, z)]) e_{ij}, \quad (5.3)$$

where r is a tuning parameter to be specified in the following theorem.

Theorem 5.2. Suppose \mathcal{C}_0 and \mathcal{C}_1 are generalized symmetric community property pair and $c_K = O(1)$. Suppose $d(\mathcal{C}_0, \mathcal{C}_1) = o(n^{c_1})$ for some constant $c_1 < 2$, and $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$. We choose the radius r in (5.3) as $r \geq r_K := d(\mathcal{C}_0, \mathcal{C}_1) + c_K^2 pK / (2(p-q))$ and $r = d(\mathcal{C}_0, \mathcal{C}_1) + O(1)$. If for any $z_0 \in \mathcal{C}_0$, we have $|B_{z_0}(r)| = O(n^{c_0})$ for some positive constant c_0 , then

$$\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(p_W \leq \alpha) = \alpha \text{ and } \lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\text{reject } H_0) = \alpha.$$

Moreover, if $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \Omega(n^\varepsilon)$ for some arbitrarily small constant $\varepsilon > 0$, we have

$$\lim_{n \rightarrow \infty} \inf_{z^* \in \mathcal{C}_1} \mathbb{P}(\text{reject } H_0) = 1.$$

We defer the proof of theorem to Appendix B.2. The scaling assumptions in Theorem 5.2 are similar to Theorem 3.2. The condition $|B_{z_0}(r)| = O(n^{c_0})$ for some $c_0 > 0$ is similar to Assumption 3.1. We need c_K in (5.1) to be bounded to prevent a specific community from being too large. By the theorem, we need to choose $r \geq r_K := d(\mathcal{C}_0, \mathcal{C}_1) + c_K^2 pK / (2(p-q))$, while p, q, c_K are unknown. In practice, we suggest to choose the radius as $r = d(\mathcal{C}_0, \mathcal{C}_1) + C\hat{p}K / (\hat{p} - \hat{q})$ for some sufficiently large C . In fact, for many concrete examples, even though r_K is unknown, we can directly construct $B_{z_0}(r_K)$. The following proposition shows how to construct $B_{z_0}(r_K)$ for Examples 1.1-1.3. Moreover, it shows the conditions on $d(\mathcal{C}_0, \mathcal{C}_1)$ and $|B_{z_0}(r_K)|$ in Theorem 5.2 are true for all these examples.

Proposition 5.3. For any $z_0 \in \mathcal{C}_0$, $B_{z_0}(r_K)$ can be constructed as follows.

- (1) Example 1.1: $B_{z_0}(r_K)$ is composed of all the assignments obtained from reassigning one node of any $z_0 \in \mathcal{C}_0$ in $[m]$ to a different community. See Figure 7(a) for an illustration. Moreover, we have $d(\mathcal{C}_0, \mathcal{C}_1) = n/K$ and $|B_{z_0}(r_K)| = m(K-1)$.
- (2) Example 1.2: Suppose $m \wedge m' \leq c_K$, $B_{z_0}(r_K)$ is composed of all the assignments obtained from reassigning nodes $m+1, \dots, m+m'$ in any $z_0 \in \mathcal{C}_0$ collectively to a different community. Moreover, we have $d(\mathcal{C}_0, \mathcal{C}_1) = n(m \wedge m')/K$ and $|B_{z_0}(r_K)| = K-1$. Suppose $m \wedge m' > c_K$, $B_{z_0}(r_K)$ is composed of all the assignments obtained from exchanging label of nodes $m+1, \dots, m+m'$ collectively with another m' nodes from a different community for any $z_0 \in \mathcal{C}_0$. See Figure 7(b) for an illustration. Moreover, we have $d(\mathcal{C}_0, \mathcal{C}_1) = 2m \wedge m'(n/K - m \wedge m')$ and $|B_{z_0}(r_K)| = O(K(n/K)^{m \wedge m'})$.

- (3) Example 1.3: For an arbitrary $z_0 \in \mathcal{C}_0$, $B_{z_0}(r_K)$ can be constructed by reassigning any node of z_0 to a different community. See Figure 7(c) for an illustration. Moreover, we have $d(\mathcal{C}_0, \mathcal{C}_1) = n/K$ and $|B_{z_0}(r_K)| = n(K-1)$.

We defer the proof to Appendix A.3. The construction of $B_{z_0}(r_K)$ is visualized in Figure 7. We also summarize the results in Table 2.

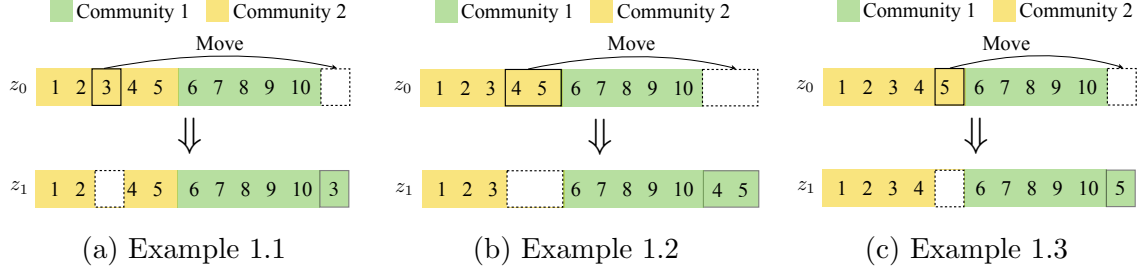


Figure 7: Construction of B_{z_0} in Proposition 5.3: (a) \mathcal{C}_0 is that nodes $\{1, 2, 3\}$ belong to the same community; (b) \mathcal{C}_0 is that the nodes set $\{1, 2, 3\}$ and $\{4, 5\}$ belong to the same community; (c) \mathcal{C}_0 is that community 1 and community 2 have equal size of 5.

We therefore have the following corollary of Theorem 5.2.

Corollary 5.4 (Examples 1.1 -1.3). Suppose $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$ and $c_K = O(1)$. We assume that $m \wedge m' = O(1)$ in Example 1.2. For Examples 1.1 -1.3, with $B_{z_0}(r_K)$ constructed in Proposition 5.3 our test for the hypothesis $H_0 : z^* \in \mathcal{C}_0$ versus $H_1 : z^* \in \mathcal{C}_1$ is honest, i.e.,

$$\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\text{reject } H_0) = \alpha.$$

Moreover, if $I(p, q)n/K = \Omega(n^\varepsilon)$ for some small positive constant ε , we have

$$\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_1} \mathbb{P}(\text{reject } H_0) = 1.$$

5.3 General Lower Bound

We can also generalize the information-theoretic lower bound in Theorem 4.1 and Theorem 4.2 to the uneven case. Similar to the even case, we need to define packing number of $B_{z_0}(r)$, which follows the same definition of $N(B_{z_0}, \varepsilon)$ in Definition 4.1. We then have the lower bound of the general case as follows.

Theorem 5.5. Suppose $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$, $p \leq 1 - \delta$ for some constant $\delta > 0$ and $c_K = O(1)$. If there exists a $z_0 \in \mathcal{C}_0$ and some $r = d(z_0, \mathcal{C}_1) + O(1)$ such that $\log N(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)}) = O(\log n)$, and

$$\limsup_{n \rightarrow \infty} \frac{d(z_0, \mathcal{C}_1)I(p, q)}{\log N(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)})} < 1, \quad (5.4)$$

then $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

	$d(z_0, \mathcal{C}_1)$	$ B_{z_0}(r_K) $	$N(B_{z_0}(r_K), \sqrt{d(z_0, \mathcal{C}_1)})$
Example 1.1	n/K	$m(K-1)$	m
Example 1.2 $m \wedge m' \leq c_K$	$n(m \wedge m')/K$	$K-1$	1
Example 1.2 $m \wedge m' > c_K$	$2m \wedge m'(n/K - m \wedge m')$	$O(K(n/K)^{m \wedge m'})$	1
Example 1.3	n/K	$n(K-1)$	n

Table 2: Important values for general cases of Examples 1.1-1.3.

Remark 5.1. If we choose $r = d(z_0, \mathcal{C}_1)$, as $B_{z_0} = B_{z_0}(d(z_0, \mathcal{C}_1))$, (5.4) reduces to (4.1). The relaxed assumption on $r = d(z_0, \mathcal{C}_1) + O(1)$ can give us a better lower bound.

We can also generalize Theorem 4.2 to the following theorem.

Theorem 5.6. Suppose $0 < q < p \leq 1 - \delta$ for some constant $\delta > 0$ and $\lim_{n \rightarrow \infty} d(\mathcal{C}_0, \mathcal{C}_1)p = \infty$. If one of the following conditions:

- (1) $d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) \leq c$ for some sufficiently small constant c ;
- (2) $\lim_{n \rightarrow \infty} d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \infty$, but there exists a $z_0 \in \mathcal{C}_0$ and some $r = d(z_0, \mathcal{C}_1) + O(1)$ such that $\limsup_{n \rightarrow \infty} d(z_0, \mathcal{C}_1)I(p, q)/\log N(B_{z_0}(r), 0) < 1$,

is satisfied, then $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

We defer the proof of the above two theorems to Appendix C.1.

To apply the general lower bound theorem to Examples 1.1-1.3, we need the following proposition on the packing number.

Proposition 5.7. We have the packing number $N(B_{z_0}(r_K), \sqrt{d(z_0, \mathcal{C}_1)})$ for three examples as follows:

- Example 1.1: $N(B_{z_0}(r_K), \sqrt{d(z_0, \mathcal{C}_1)}) = m$;
- Example 1.2: $N(B_{z_0}(r_K), \sqrt{d(z_0, \mathcal{C}_1)}) = 1$;
- Example 1.3: $N(B_{z_0}(r_K), \sqrt{d(z_0, \mathcal{C}_1)}) = n$.

We defer the proof to Appendix A.4. The results is also summarized in Table 2.

Since $r_K = d(\mathcal{C}_0, \mathcal{C}_1) + c_K^2 pK/(2(p-q))$, where $c_K^2 pK/(2(p-q)) = O(1)$ and $d(\mathcal{C}_0, \mathcal{C}_1) = d(z_0, \mathcal{C}_1)$ by the symmetry of $\mathcal{C}_0, \mathcal{C}_1$, we have that $r_K = d(z_0, \mathcal{C}_1) + O(1)$. Applying Theorem 5.5 and Proposition 5.7, we have the following lower bound of same community test in Example 1.1.

Corollary 5.8. For \mathcal{C}_0 and \mathcal{C}_1 defined in Example 1.1, if $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$, $p < 1 - \delta$ for some constant $\delta > 0$, $c_K = O(1)$ and

$$\limsup_{n \rightarrow \infty} nI(p, q)/(K \log m) < 1,$$

we have $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

Applying Theorem 5.6 and Proposition 5.7, we have the following lower bound of same community test for groups in Example 1.2.

Corollary 5.9. For \mathcal{C}_0 and \mathcal{C}_1 defined in (2.2), if $np \rightarrow \infty$, $0 < q < p < 1 - \delta$ for some $\delta > 0$ and

$$\limsup_{n \rightarrow \infty} nI(p, q) < c,$$

for some sufficiently small constant $c > 0$, we have $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

For Example 1.3, applying Theorem 5.5 and Proposition 5.7, we have the following result.

Corollary 5.10. For \mathcal{C}_0 and \mathcal{C}_1 defined in (5.2), if $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$, $p < 1 - \delta$ for some constant $\delta > 0$ and

$$\limsup_{n \rightarrow \infty} nI(p, q)/(K \log n) < 1,$$

we have $\liminf_{n \rightarrow \infty} r(\mathcal{C}_0, \mathcal{C}_1) \geq 1/2$.

6. Numerical Results on Synthetic Data

We conduct the shadowing bootstrap on Examples 1.1 and 1.3. We test both hypotheses at the significance level $\alpha = 0.05$. We consider the number of nodes $n = 200, 600, 1000$ and the number of clusters $K = 2$. The connection probabilities are set to be $p = (1 + \Delta)\rho_n$ and $q = (1 - \Delta)\rho_n$, where $\rho_n = (n/K)^{-0.3}$ and Δ is the parameter that controls the difference between p and q . We choose Δ varying from 0 to 0.8. The maximum likelihood estimator is calculated via using singular value decomposition estimator as an initialization to boost the computation. For Example 1.1, we set $m = \lceil (n/K)^\delta / 2 \rceil$ for $\delta = 0.3, 0.5, 0.7$ to explore the influence of m on type-I and II errors.

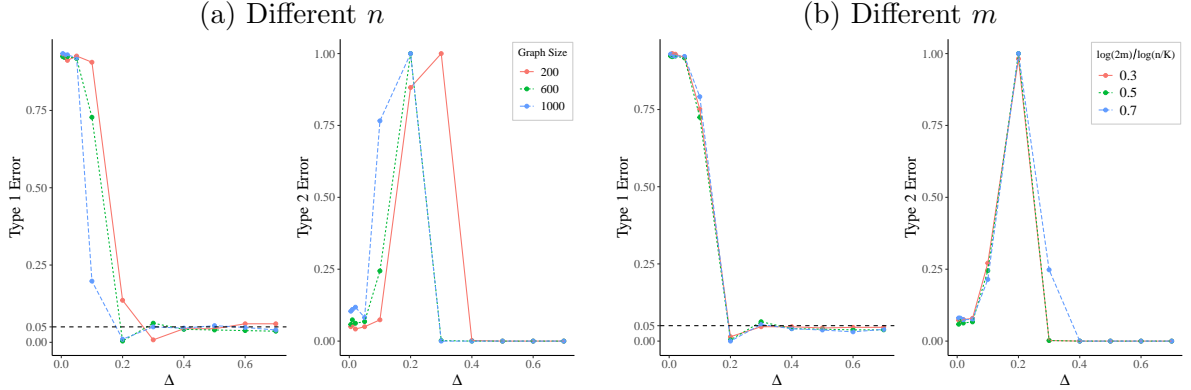


Figure 8: Type-I and type-II errors for Example 1.1 as Δ varies in $[0, 0.8]$. Panel (a) shows the results for different graph sizes $n = 200, 600, 1000$ with m set as $\lceil (n/K)^{0.5}/2 \rceil$. Panel (b) shows the results $m = \lceil (n/K)^\delta/2 \rceil$ with $\delta = 0.3, 0.5, 0.7$ and the graph size $n = 600$.

For both Examples 1.1 and 1.3, the histograms of the shadowing bootstrap statistic in (2.11) by choosing different shadowing assignment $z_0 \in \mathcal{C}_0$ are illustrated in Figure 10. We

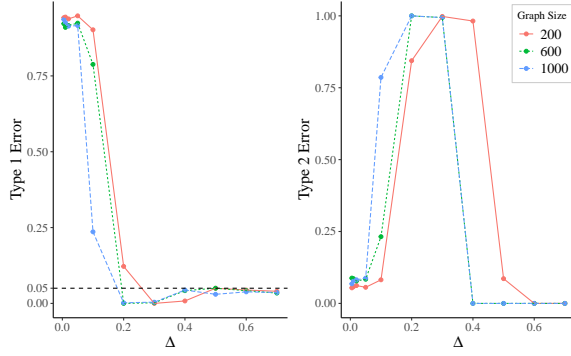


Figure 9: Type-I and type-II errors for Example 1.3 as Δ varies in $[0, 0.8]$ and the graph size $n = 200, 600, 1000$.

can see that the quantiles are almost the same for different z_0 's. This validates the rationale of our shadowing bootstrap method: we can estimate the p-value based on L_0 in (2.7) by replacing the unknown truth z^* with some shadowing assignment $z_0 \in \mathcal{C}_0$. In Figure 8, we show how the type-I and type-II errors vary with the signal strength Δ , the graph size n and the number of tested nodes m for Example 1.1. Both the type-I and type-II errors are estimated via 500 repetitions. As Δ increases, the type-I error converges to the significance level 0.05, which shows that our method is honest. Type-II error is small when Δ is around zero as the test will always reject the null when the signal strength is too small, while it increases drastically as type-I error drops to 0. When Δ is large enough, the type-II error converges to 0, showing that our test is powerful. In Figure 8(b), we can see that the type-I and type-II errors for different m 's converge similarly as Δ increases. This is consistent with Corollary 3.4 as the scaling condition is irrelevant to m . Simulation results for Example 1.3 are shown in Figure 9. The type 1 and type 2 error rates vary similarly as in Example 1.1.

We also compare our method with SIMPLE proposed in Fan et al. (2019). SIMPLE was designed to conduct a two sample test on whether two nodes belong to the same community. This is the case of Example 1.1 in our paper with $m = 2$. We also compare SIMPLE with our method under Example 1.1 with $m > 2$ by combining SIMPLE with Bonferroni correction. We consider $n = 600$, $K = 2$ and $m = 2, 3, 28$, where $m = 28$ is chosen by setting $\delta = 0.7$ in the formula $\lceil (n/K)^\delta / 2 \rceil$. The probabilities p and q are chosen in the same way as the previous examples. The results of the comparison are shown in Table 3. We applied the Bonferroni Correction to the SIMPLE method when doing the multiple comparison. From the results, our method outperforms SIMPLE in both the single-pair testing and the multiplicity testing. For the multiplicity testing results, Bonferroni Correction would result in a more conservative type-I error, yet the reported size from SIMPLE is still larger than the desired size of 0.05, indicating a lack of accuracy of the method under our setting of parameters. In comparison, the performance of our method is good and stable under all settings.

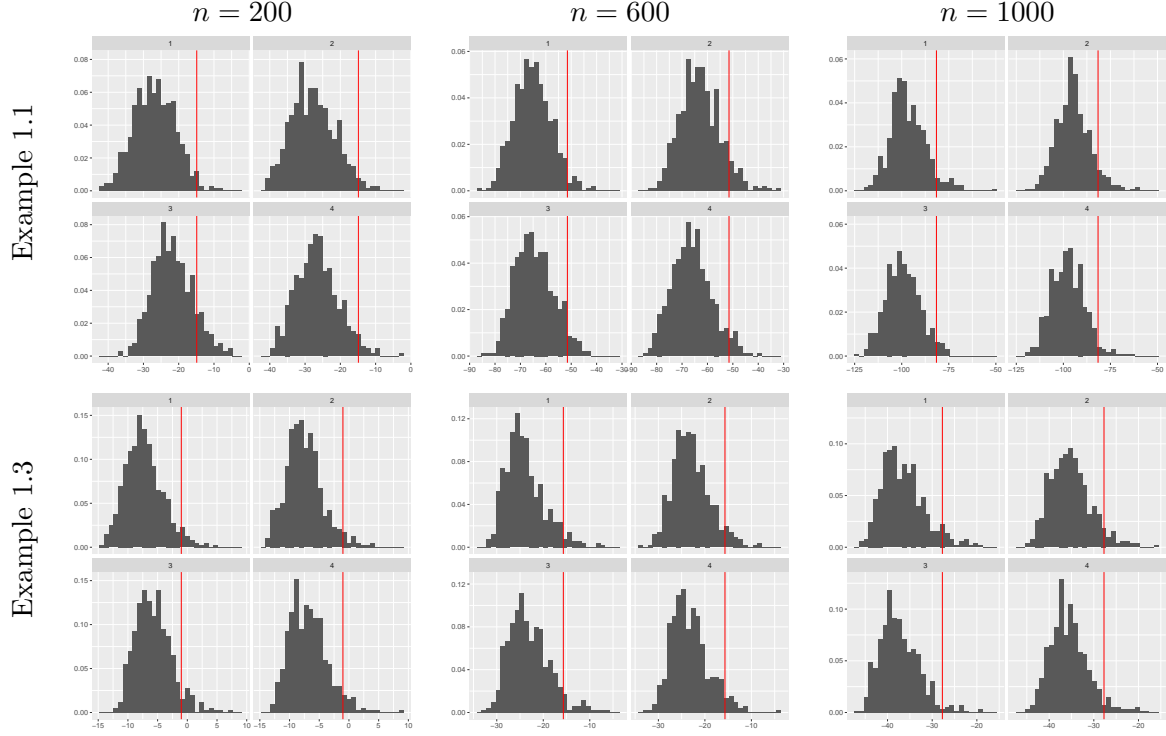


Figure 10: The histograms of the transformed shadowing bootstrap statistic $g(\hat{p}, \hat{q})W_n + g(\hat{p}, \hat{q})\hat{\mu}_0$ for Examples 1.1 and 1.3 with 500 bootstrap samples. We vary $n = 200, 600, 1000$ and for each n we choose four different $z_0 \in \mathcal{C}_0$. For Example 1.1, we choose $\Delta = 0.5$ and $m = \lceil (n/K)^{0.5}/2 \rceil$. For Example 1.3, we choose $\Delta = 0.5$. Each panel shows the distribution of the statistic W_n under different true z_0 . According to the histogram, the distribution remains consistent for different true assignments in \mathcal{C}_0 .

$m = 2$	Δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Our Method	Size	0.770	0.008	0.056	0.040	0.060	0.048	0.056
	Power	0.768	0.018	0.998	1.000	1.000	1.000	1.000
SIMPLE	Size	0.270	0.116	0.102	0.086	0.082	0.076	0.064
	Power	0.312	0.756	0.990	1.000	1.000	1.000	1.000
$m = 3$	Δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Our Method	Size	0.752	0.014	0.048	0.048	0.044	0.046	0.046
	Power	0.728	0.018	0.998	1.000	1.000	1.000	1.000
SIMPLE	Size	0.344	0.118	0.070	0.074	0.070	0.084	0.070
	Power	0.426	0.816	0.992	1.000	1.000	1.000	1.000
$m = 28$	Δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Our Method	Size	0.792	0.000	0.054	0.040	0.036	0.030	0.038
	Power	0.784	0.000	0.752	1.000	1.000	1.000	1.000
SIMPLE	Size	0.938	0.232	0.126	0.088	0.072	0.06	0.044
	Power	0.950	0.998	1.000	1.000	1.000	1.000	1.000

Table 3: Comparison of our method with SIMPLE for Example 1.1 with $m = 2, 3, 28$.

7. Real Application to Protein-Protein Network

We apply our method to study the Protein-Protein Interaction (PPI) networks. We aim to test whether different protein functional groups belong to the same community in PPI (Tabouy et al., 2020). The protein network is extracted from the STRING platform³ (Szklarczyk et al., 2015). For proteins i and j , STRING dataset assigns an interaction score $v_{ij} \in [0, 1]$ between the two proteins. We construct the adjacency matrix \mathbf{A} of PPI as follows:

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{if } v_{ij} > 0.95, \\ 0 & \text{if } v_{ij} \leq 0.95. \end{cases}$$

We select the proteins from the two major orthologous groups: prokaryotic clusters (COGs) and eukaryotic clusters (KOGs) (Koonin et al., 2004; Galperin et al., 2017) and remove those proteins in both COGs and KOGs. We select 222 proteins in total. We further divide COGs and KOGs into the following 5 subgroups according to the Clusters of Orthologous Groups of proteins database⁴:

- G_1 : KOG-Information storage and processing;
- G_2 : KOG-Cellular processes and signaling;
- G_3 : COG-Cellular processes and signaling;
- G_4 : COG-Information storage and processing;
- G_5 : COG-Metabolism.

3. <https://string-db.org>.

4. <http://www.ncbi.nlm.nih.gov/COG/>

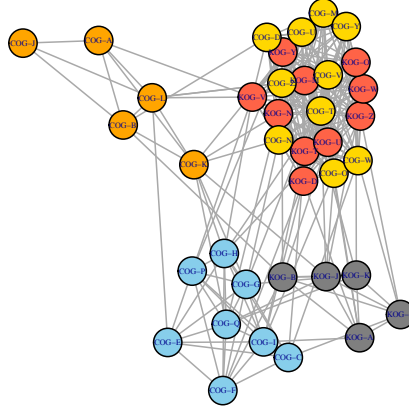


Figure 11: The functional network of protein interaction network. Each node represent a functional subgroup in KOGs or COGs (e.g. “A”-RNA processing and modification⁵). The colors denote different functional categories. Gray nodes belongs to G_1 : KOG-Information storage and processing. Red nodes belong to G_2 : KOG-Cellular processes and signaling. Yellow nodes belong to G_3 : COG-Cellular processes and signaling. Orange nodes belong to G_4 : COG-Information storage and processing. Blue nodes belong to G_5 : COG-Metabolism.

The protein-protein interaction network is visualized in Figure 11. We aim to test whether any two subgroups above belong to the same community in PPI. We can formulate the hypotheses as follows:

$H_{0,ij}$: Subgroup G_i and G_j belong to the same community,

$H_{1,ij}$: Subgroup G_i belong to community a , but G_j belong to community $b \neq a$.

for $1 \leq i \neq j \leq 5$. These hypotheses are the same community test for groups in Example 1.2. We apply the shadowing bootstrap method to test $H_{0,ij}$ for all $1 \leq i \neq j \leq 5$. We summarize the p -values of these hypothesis tests in Table 4.

Subgroups	G_1	G_2	G_3	G_4	G_5
G_1	-	$< 10^{-6}$	$< 10^{-6}$	0.027806	0.013166
G_2	$< 10^{-6}$	-	0.165860	0.000032	0.000178
G_3	$< 10^{-6}$	0.165860	-	0.003800	0.029974
G_4	0.027806	0.000032	0.003800	-	0.023648
G_5	0.013166	0.000178	0.029974	0.023648	-

Table 4: P-values of the test $H_{0,ij}$ on whether two subgroups G_i and G_j are in the same community for $1 \leq i \neq j \leq 5$.

5. <https://ftp.ncbi.nih.gov/pub/COG/COG2014/static/lists/homeCOGs.html>

According to the results, the p -value for G_2 and G_3 does not reach the significant level and we fail to reject the null hypothesis that they are in the same community. This is consistent with the fact that both G_2 and G_3 belong to the cellular processing and signaling functional category. On the other hand, both G_1 and G_4 belong to the information storage and processing functional category, but the p -value between them is significant. This is consistent with the KOGs and COGs are heterogeneous in cellular processing (Koonin et al., 2004). We also notice that p -values between G_1 and G_2 , G_1 and G_3 , G_2 and G_4 are highly significant, as the information storage and processing and the cellular processes and signaling are different functional processes (Ur Rehman et al., 2017; Brun et al., 2003; Pal and Eisenberg, 2005).

Acknowledgments

We are thankful to the support from NSF1916211, NIH1R35CA220523-01A1, and NIH5U01CA209414-02.

Appendix A. Proofs of Community Properties

In this section we mainly focus on the proofs concerning community properties, including the generalization of symmetric community property pairs from even to uneven cluster sizes, the size of the ball $B_{z_0}(r_K)$ in three examples, and the packing number of the ball in each case.

A.1 Proof of Proposition 4.3

Example 1.1: In this case, for a given $z_0 \in \mathcal{C}_0$, we have derived the form of B_{z_0} . For any $z_i, z_j \in \mathcal{P}(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)})$, we know from Section 2.4 that they are transformed from z_0 by swapping one of the first m nodes with another node from a different cluster. The node among the first m to be swapped $s \in [m]$ cannot be the same for the two assignments, otherwise $|\mathcal{E}_{1,2}(z_0, z_i) \cap \mathcal{E}_{1,2}(z_0, z_j)| \geq |\mathcal{E}_1(z_0, z_i) \cap \mathcal{E}_1(z_0, z_j)| = n/K - 1 \gg \sqrt{d(z_0, \mathcal{C}_1)}$. Thus each $z \in \mathcal{P}(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)})$ corresponds to a different swapped node among the first m nodes, and we have $N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}) \leq m$. On the other hand, for the given assignment z_0 , we can construct the following set $\{z_k\}_{k=1}^m$: we take a set of nodes $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$ from a cluster different from the cluster to which the first m nodes of z_0 belong. Then for each k , we swap the cluster assignment of node k with node s_k , $k = 1, \dots, m$, and obtain the corresponding alternative assignment z_k . Then for any two alternative assignments z_i and z_j obtained this way, we have $|\mathcal{E}_{1,2}(z_0, z_i) \cap \mathcal{E}_{1,2}(z_0, z_j)| \leq 4$. Thus $N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}) = m$.

Example 1.2: For a given $z_0 \in \mathcal{C}_0$ and the corresponding boundary B_{z_0} , it can be perceived that $N(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}) = N(B_{z_0}, 0) = 1$, because any $z \in B_{z_0}$ involves swapping the set \mathcal{S}_2 so that $\forall z_i, z_j \in B_{z_0}, |\mathcal{E}_{1,2}(z_0, z_i) \cap \mathcal{E}_{1,2}(z_0, z_j)| \geq m \wedge m'(n/K - m \wedge m')$.

A.2 Proof of Proposition 5.1

To prove that Definition 2.2 is a special case of Definition 5.1 when the community size is even, it suffices for us to construct a concrete community label permutation σ and node label permutation τ satisfying Definition 5.1 based on \mathcal{N} and \tilde{z} . Here we use Figure 5 to illustrate the construction. Given any $z, z' \in \mathcal{C}_0$, we first construct σ . Since $z_{\mathcal{N}} \simeq z'_{\mathcal{N}} \simeq \tilde{z}_{\mathcal{N}}$, by Definition 2.2, there must exist a $\sigma \in S_K$ mapping z to z' on the support \mathcal{N} , i.e., $\sigma(z_{\mathcal{N}}) = z'_{\mathcal{N}}$. For example, in Figure 5, we construct a σ swapping communities 1 and 2. After matching the community labels, we now construct τ in order to transform $\sigma(z)$ to z' . Since the community size is even and $\sigma(z_{\mathcal{N}}) = z'_{\mathcal{N}}$, $\sigma(z)$ and z' have equal cluster sizes on the support of \mathcal{N}^c . Therefore, there exists $\tau \in S_n$ such that $\tau(\sigma(z)_{\mathcal{N}^c}) = z'_{\mathcal{N}^c}$ and $\tau(\sigma(z)_{\mathcal{N}}) = \sigma(z)_{\mathcal{N}} = z'_{\mathcal{N}}$. We can see the example of τ in Figure 5. Using σ and τ constructed above, we can check that $\tau \circ \sigma(z) = z'$. We now check the last condition in Definition 5.1. For any $z'' \in \mathcal{C}_1$, since τ is invariant on \mathcal{N} , we have $\tau \circ \sigma(z''_{\mathcal{N}}) = \sigma(z''_{\mathcal{N}}) \simeq z''_{\mathcal{N}}$. By Definition 2.2, the alternative community \mathcal{C}_1 is closed under permutation on the support of \mathcal{N} , we have $\tau \circ \sigma(z'') \in \mathcal{C}_1$. Therefore, we check that Definition 2.2 is a special case of Definition 5.1.

Since the property pairs in (2.1) and (2.2) are symmetric property pairs, they are also generalized symmetric property pairs following the preceding arguments. As for the property pair in (5.2), we can see from Figure 5(b) that for any two assignments $z, z' \in \mathcal{C}_0$, since they have equal community sizes, we can take σ to be the identity map and there exists $\tau \in S_n$

such that $\tau(z) = z'$. Then for any $z'' \in \mathcal{C}_1$, since τ does not change the community sizes, we know that $\tau(z'')$ still have uneven community sizes and $\tau(z'') \in \mathcal{C}_1$. Therefore, by Definition 5.1, the property pair in (5.2) is a generalized symmetric property pair.

A.3 Proof of Proposition 5.3

Example 1.1: To construct $B_{z_0}(r_K)$, we need to find all the assignments in \mathcal{C}_1 whose distance from z_0 is no larger than $d(\mathcal{C}_0, \mathcal{C}_1)$ by an extra constant term. To construct assignments in \mathcal{C}_1 closest to z_0 , we would pick one node in $[m]$ and reassign it to a different community (see Figure 7 (a)). Assignments constructed in such ways will satisfy $d(z_0, z_1) = d(\mathcal{C}_0, \mathcal{C}_1) = n/K$. If we make community changes to any other nodes on the basis of such construction, then $d(z_0, z_1)$ would increase by at least $n/K - 2$, which exceeds the constant level. Thus $B_{z_0}(r_K)$ consists of all assignments constructed by moving one node of z_0 in $[m]$ to a different cluster. Since we can pick m nodes in total and reassign them to $K - 1$ different clusters, $|B_{z_0}(r_K)| = (K - 1)m = O(m)$.

Example 1.2: For an arbitrary $z_0 \in \mathcal{C}_0$, without loss of generality, we assume that $m' \leq m$. Then when $m' \leq c_K$, to construct assignments in \mathcal{C}_1 that are closest to z_0 , we need to reassign nodes $m + 1, \dots, m + m'$ collectively to a different community (see Figure 7 (b)). Such constructed assignments have distance $d(z_0, z_1) = d(\mathcal{C}_0, \mathcal{C}_1) = m'n/K$. Similar to the previous example, any community changes to other nodes on the basis of such construction would result in increase of $d(z_0, z_1)$ by at least $n/K - m' - 1$. Therefore, $B_{z_0}(r_K)$ consists of those assignments in \mathcal{C}_1 constructed by reassigning nodes $m + 1, \dots, m + m'$. Since there are $K - 1$ other clusters to reassign in total, we have $|B_{z_0}(r_K)| = K - 1 = O(1)$. On the other hand, when $m' > c_K$, then we cannot reassign nodes $m + 1, \dots, m + m'$ collectively without exchanging with other nodes, otherwise the community size bound will be violated. Then $d(\mathcal{C}_0, \mathcal{C}_1)$ and $B_{z_0}(r_K)$ is exactly the same as the even case and the claim follows.

Example 1.3: As for the ball $B_{z_0}(r_K)$ for an arbitrary $z_0 \in \mathcal{C}_0$, to transform z_0 into an assignment $z_1 \in \mathcal{C}_1$, the simplest way is to reassign an arbitrary node to a different community, and $d(z_0, z_1) = d(\mathcal{C}_0, \mathcal{C}_1) = n/K \asymp n$. Further community changes will result in increasing in $d(z_0, z_1)$ that exceeds the constant level. Since we can obtain such z_1 by reassigning any one of the n nodes into the other $K - 1$ clusters, we have $|B_{z_0}(r_K)| = n(K - 1) = O(n)$.

A.4 Proof of Proposition 5.7

The arguments for Example 1.1 and Example 1.2 are almost the same as in the even cases and are hence omitted.

Example 1.3: For a given $z_0 \in \mathcal{C}_0$, from previous discussion we can see that the ball $B_{z_0}(r)$ with $r = d(z_0, \mathcal{C}_1) + O(1)$ is composed of all the assignments that differ from z_0 by one mis-aligned node. For any $z_i, z_j \in \mathcal{P}(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)})$, the misaligned node s cannot be the same, otherwise $|\mathcal{E}_{1,2}(z_0, z_i) \cap \mathcal{E}_{1,2}(z_0, z_j)| \geq n/K \gg \sqrt{d(z_0, \mathcal{C}_1)}$. Thus we have $N(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)}) \leq n$. Also since the set $\{z_k\}_{k=1}^n$ where each z_k is obtained by reassigning the node k into another cluster obviously satisfies the condition that $|\mathcal{E}_1(z_0, z_i) \cap \mathcal{E}_1(z_0, z_j)| + |\mathcal{E}_2(z_0, z_i) \cap \mathcal{E}_2(z_0, z_j)| \leq 1$, we have that $N(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)}) = n$.

Appendix B. Proof of Inference Results

In this section, we provide the proofs of the theorems on inference results. We will first prove Proposition 2.3 which implies that the p-value based on the maximal leading term L_0 can be estimated without knowing the true assignment, then we prove the main Theorem 3.2 using Proposition 2.3 along with other lemmas. The proof of the technical lemmas will be deferred to Section D.

In the following part of our paper, we use $c, C, c_1, c_2, C_1, C_2, \dots$ to represent generic constants and their values may vary in different places.

B.1 Proof of Proposition 2.3

To prove Proposition 2.3, we need the following generalized version of Lemma 2.2 stated previously

Lemma B.1 (Shadowing symmetry lemma). For a given $z \in \mathcal{C}_0$ and a given radius $r > 0$, we list the assignments in the ball $B_z(r)$ as $z_1, z_2, \dots, z_{|B_z(r)|}$. Define a $|B_z(r)|$ -dimensional vector \mathbf{L}_z as

$$(\mathbf{L}_z)_k = g(p, q) \left(\sum_{(i,j) \in \mathcal{E}_2(z, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z, z_k)} \mathbf{A}_{ij} \right), \text{ for } k = 1, 2, \dots, |B_z(r)|.$$

Suppose \mathcal{C}_0 and \mathcal{C}_1 satisfy definition 5.1, then for any $z_0, z'_0 \in \mathcal{C}_0$, we have $|B_{z_0}(r)| = |B_{z'_0}(r)|$ and $\text{Cov}(\mathbf{L}_{z_0})$ equals to $\text{Cov}(\mathbf{L}_{z'_0})$ up to permutation, i.e., there existing a permutation $\tau \in S_{|B_{z_0}(r)|}$ such that $\text{Cov}(\mathbf{L}_{z_0})_{kl} = \text{Cov}(\mathbf{L}_{z'_0})_{\tau(k)\tau(l)}$ for all $k, l = 1, \dots, |B_{z_0}(r)|$.

We defer the proof of Lemma B.1 to Section D.1. Now we are ready to prove Proposition 2.3. In fact, the boundary in the definition of L_0 can be generalized to the ball $B_z(r)$ with $r \geq r_K := d(\mathcal{C}_0, \mathcal{C}_1) + c_K^2 pK / (2(p - q))$ and $r = d(\mathcal{C}_0, \mathcal{C}_1) + O(1)$. For the true assignment $z^* \in \mathcal{C}_0$, we have that

$$\begin{aligned} L_0 &= \sup_{z_k \in B_{z^*}(r)} \left\{ g(p, q) \left(\sum_{(i,j) \in \mathcal{E}_2(z^*, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z_k)} \mathbf{A}_{ij} \right) \right\} \\ &= g(p, q) \sup_{z_k \in B_{z^*}(r)} \left\{ \sum_{i < j} \left\{ (\mathbf{A}_{ij} - \mathbb{E}(\mathbf{A}_{ij})) (\mathbb{1}[(i, j) \in \mathcal{E}_2(z^*, z_k)] - \mathbb{1}[(i, j) \in \mathcal{E}_1(z^*, z_k)]) \right\} \right\} \\ &\quad + g(p, q) \mu_0 + \delta_n \\ &= g(p, q) \sigma_0 \sup_{k \in [|B_{z^*}(r)|]} \left\{ \frac{1}{\sigma_0} \sum_{i < j} (\mathbf{X}_{ij})_k \right\} + g(p, q) \mu_0 + \delta_n. \end{aligned}$$

where the vector $\mathbf{X}_{ij} \in \mathbb{R}^{|B_{z^*}(r)|}$ and $(\mathbf{X}_{ij})_k = (\mathbf{A}_{ij} - \mathbb{E}(\mathbf{A}_{ij})) (\mathbb{1}[(i, j) \in \mathcal{E}_2(z^*, z_k)] - \mathbb{1}[(i, j) \in \mathcal{E}_1(z^*, z_k)])$, $\delta_n = O(\rho_n)$, and $\sigma_0 = \sqrt{d(\mathcal{C}_0, \mathcal{C}_1)(p(1-p) + q(1-q))}$, $\mu_0 = d(\mathcal{C}_0, \mathcal{C}_1)(q - p)$. We can see that for different (i, j) , the vector \mathbf{X}_{ij} are independent of each other. For a fixed $k \in [|B_{z^*}(r)|]$, when $(i, j) \notin \mathcal{E}_{1,2}(z^*, z_k)$, $(\mathbf{X}_{ij})_k = 0$. When $(i, j) \in \mathcal{E}_{1,2}(z^*, z_k)$, under the regime $1/\rho_n = o(n^{1-c_2})$ for some positive c_2 , there exists $B_n = 1/\sqrt{\rho_n} = o(n^{(1-c_2)/2})$ such that $|(\mathbf{X}_{ij})_k / \sqrt{\rho_n}| \leq B_n$ and $B_n^2 (\log 2d(\mathcal{C}_0, \mathcal{C}_1) |B_{z_0}(r)|)^7 / n \leq n^{-c_2/2}$, where $d(\mathcal{C}_0, \mathcal{C}_1) =$

$o(n^2)$. Therefore, following a very similar proof as Theorem 2.2 and Corollary 2.1 in Chernozhukov et al. (2013), we have

$$g(p, q) \sup_{k \in [|B_{z^*}(r)|]} \left\{ \sum_{i < j} (\mathbf{X}_{ij})_k / \sigma_0 \right\} \xrightarrow{d} \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k,$$

where $\tilde{Z} \sim N(0, \Sigma_{z^*} / \sigma_0^2)$, and $\Sigma_{z^*} = \text{Cov}(\mathbf{L}_{z^*})$. Therefore, we have that

$$\begin{aligned} & \sup_{t \in \mathbb{R}} \left| \mathbb{P}(L_0 \leq t) - \mathbb{P}(\sigma_0 \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k + g(p, q)\mu_0 \leq t) \right| \\ & \leq \sup_{t \in \mathbb{R}} \left| \mathbb{P}(L_0 \leq t) - \mathbb{P}(\sigma_0 \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k + g(p, q)\mu_0 + \delta_n \leq t) \right| \\ & \quad + \sup_{t \in \mathbb{R}} \left| \mathbb{P}(\sigma_0 \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k + g(p, q)\mu_0 + \delta_n \leq t) - \mathbb{P}(\sigma_0 \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k + g(p, q)\mu_0 \leq t) \right| \\ & \leq o(1) + \sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(\left| \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k - (t - g(p, q)\mu_0) / \sigma_0 \right| \leq \delta_n / \sigma_0 \right) \right|. \end{aligned}$$

We know that $\min_{k \in [|B_{z^*}(r)|]} \text{Var}(\tilde{Z}_k) = \Omega(g(p, q)^2) = \Omega(1)$, $\log |B_{z^*}(r)| = O(\log n)$ and $\delta_n / \sigma_0 = O(n^{-1/2})$. Then by Lemma 2.1 in Chernozhukov et al. (2013), we have

$$\begin{aligned} & \sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(\left| \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k - (t - g(p, q)\mu_0) / \sigma_0 \right| \leq \delta_n / \sigma_0 \right) \right| \\ & \lesssim \frac{\delta_n}{\sigma_0} \left\{ \sqrt{2 \log |B_{z^*}(r)|} + \sqrt{\min_{k \in [|B_{z^*}(r)|]} \text{Var}(\tilde{Z}_k) \sigma_0 / \delta_n} \right\} \leq n^{-1/4}. \end{aligned}$$

And thus we have

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(L_0 \leq t) - \mathbb{P}(\sigma_0 \sup_{k \in [|B_{z^*}(r)|]} \tilde{Z}_k + g(p, q)\mu_0 \leq t) \right| = o(1).$$

Following the same procedure with z^* replaced by z_0 , we also have

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(L_0(z_0) \leq t) - \mathbb{P}(\sigma_0 \sup_{k \in [|B_{z_0}(r)|]} (\tilde{Z}')_k + g(p, q)\mu_0 \leq t) \right| = o(1),$$

where $\tilde{Z}' \sim N(0, \Sigma_{z_0} / \sigma_0^2)$, and $\Sigma_{z_0} = \text{Cov}(\mathbf{L}_{z_0})$. By Lemma B.1 we know that Σ_{z^*} and Σ_{z_0} are equal up to permutation. Therefore, the claim follows. We may also notice that the validity of the proof does not depend on the values of p, q as long as the regime is $1/\rho_n = o(n^{1-c_2})$ for some constant $c_2 > 0$, and thus the statement is also true for $\hat{L}_0 := \sup_{z_k \in B_{z^*}(r)} \left\{ g(\hat{p}, \hat{q}) \left(\sum_{(i,j) \in \mathcal{E}_2(z^*, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z_k)} \mathbf{A}_{ij} \right) \right\}$ with plugged-in estimators \hat{p}, \hat{q} .

B.2 Proof of Theorem 3.2

In fact, Proposition 5.1 shows that the symmetric community property pairs defined in Section 2 are general symmetric property pairs under the general framework, and Theorem 5.2 is a generalization of Theorem 3.2 under uneven cluster sizes. Thus we can just prove the more general Theorem 5.2 and the proof will also apply to Theorem 3.2.

The proof of the main theorem requires the help of Proposition 2.3 and the following lemma that shows why the maximizer in the alternative assignment space can be restricted to the ball centered at the true assignment $z_0 \in \mathcal{C}_0$.

Lemma B.2. We denote z^* as the true assignment, and $B_{z^*}(r_K)$ is the ball centered at z^* with radius $r_K = d(z^*, \mathcal{C}_1) + \frac{pK}{2(p-q)}c_K^2$, $c_K = O(1)$. Under the same conditions of Theorem 5.2, when $z^* \in \mathcal{C}_0$

$$\sup_{z \in \mathcal{C}_1} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) = \sup_{z \in B_{z^*}(r_K)} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) + O_P(\rho_n); \quad (\text{B.1})$$

Moreover, for any true assignment z^* , we have

$$\sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) = \log f(\mathbf{A}; z^*, \hat{p}, \hat{q}) + O_P(\rho_n). \quad (\text{B.2})$$

With help of this lemma, instead of taking the supremum over the entire assignment space \mathcal{C}_1 , we are able to restrict the maximizer to a much smaller set $B_{z^*}(r_K)$ so that the Central Limit Theorem can be applied. Recall that the boundary B_{z^*} defined in Section 2.3 is in essence a ball with radius $d(z^*, \mathcal{C}_1)$. We defer the proof of Lemma B.2 to Appendix D.2.

Now we are ready to present the proof of Theorem 5.2:

We first define the α quantile of the $\widehat{\text{LRT}}$ statistic. Let $C_W(\alpha)$ be the $1 - \alpha$ quantile of W_n conditioning on $\widehat{\mathbf{A}}$ and \mathbf{A} , i.e., $\mathbb{P}(W_n \leq C_W(\alpha) | \widehat{\mathbf{A}}, \mathbf{A}) = 1 - \alpha$. We then estimate the quantile of LRT by

$$q_\alpha = g(\hat{p}, \hat{q})C_W(\alpha) + g(\hat{p}, \hat{q})\hat{\mu}_0, \quad (\text{B.3})$$

then it can be seen that the two events $\{\widehat{\text{LRT}} \geq q_\alpha\}$ and $\{p_W \leq \alpha\}$ are equivalent. Therefore, it suffices to show that $\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\widehat{\text{LRT}} \leq q_\alpha) = \alpha$. The proof is mainly composed of three parts. The first part is to briefly illustrate the derivation of L_0 as the leading term of the log-likelihood ratio, the second part is to control the error caused by plugging in the estimators of connection probabilities \hat{p}, \hat{q} , and the third part is to illustrate the multiplier bootstrap as a valid approximation of the LRT quantile.

B.2.1 DERIVATION OF THE LEADING TERM FOR LRT

For a given true assignment $z^* \in \mathcal{C}_0$, by Lemma B.2 we have:

$$\begin{aligned}\widehat{\text{LRT}} &= \log \frac{\sup_{z \in \mathcal{C}_1} f(\mathbf{A}; z, \widehat{p}, \widehat{q})}{\sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} f(\mathbf{A}; z, \widehat{p}, \widehat{q})} \\ &= \sup_{z \in \mathcal{C}_1} \log f(\mathbf{A}; z, \widehat{p}, \widehat{q}) - \log f(\mathbf{A}; z^*, \widehat{p}, \widehat{q}) + O_P(\rho_n) \\ &= \sup_{z_k \in B_{z^*}(r)} \left(\log f(\mathbf{A}; z_k, \widehat{p}, \widehat{q}) - \log f(\mathbf{A}; z^*, \widehat{p}, \widehat{q}) \right) + O_P(\rho_n),\end{aligned}$$

where $r \geq r_K := d(\mathcal{C}_0, \mathcal{C}_1) + c_K^2 pK / (2(p - q))$ and $r = d(\mathcal{C}_0, \mathcal{C}_1) + O(1)$. In practice, due to the consistency of \widehat{p}, \widehat{q} , when we choose the radius $r = d(\mathcal{C}_0, \mathcal{C}_1) + C\widehat{p}K / (\widehat{p} - \widehat{q})$ for some sufficiently large C , we can make sure that the conditions on the radius is satisfied with probability $1 - o(1)$. Thus we can see that the $\widehat{\text{LRT}}$ is essentially the supremum of the log-likelihood difference between the true assignment z^* and the alternative assignments in the ball $B_{z^*}(r)$. We further expand the log-likelihood terms and can write

$$\begin{aligned}\widehat{\text{LRT}} &= \sup_{z_k \in B_{z^*}(r)} \left\{ g(\widehat{p}, \widehat{q}) \left(\sum_{(i,j) \in \mathcal{E}_2(z^*, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z_k)} \mathbf{A}_{ij} \right) + \log \left(\frac{1 - \widehat{q}}{1 - \widehat{p}} \right) \left(n_1(z^*, z_k) - n_2(z^*, z_k) \right) \right\} \\ &\quad + O_P(\rho_n) \\ &= \widehat{L}_0 + \delta_n.\end{aligned}$$

where $\delta_n = \sup_{z_k \in B_{z^*}(r)} \left\{ \log \left((1 - \widehat{q}) / (1 - \widehat{p}) \right) \left(n_1(z^*, z_k) - n_2(z^*, z_k) \right) \right\} + O_P(\rho_n) = O_P(\rho_n)$, and $\widehat{L}_0 = g(\widehat{p}, \widehat{q}) \sup_{z_k \in B_{z^*}(r)} \left(\sum_{(i,j) \in \mathcal{E}_2(z^*, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z_k)} \mathbf{A}_{ij} \right)$. From Proposition 2.3 we have that $\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |\mathbb{P}(\widehat{L}_0 < t) - \mathbb{P}(\widehat{L}_0(z_0) < t)| = 0$ for any $z_0 \in \mathcal{C}_0$. Therefore, it suffices for us to prove that $\mathbb{P}(\widehat{\text{LRT}} \geq q_\alpha) = \alpha + o(1)$ for one given true assignment $z_0 \in \mathcal{C}_0$. Now we are ready to prove the validity of multiplier bootstrap for estimating the quantile based on the leading term.

B.2.2 BOUNDING OF ERROR CAUSED BY PLUGGING IN \widehat{p}, \widehat{q}

From previous section we know that

$$\widehat{L}_0(z_0) = g(\widehat{p}, \widehat{q}) \sigma_0 \sup_{k \in [|B_{z_0}(r)|]} \left\{ \frac{1}{\sigma_0} \sum_{i < j} (\mathbf{X}_{ij})_k \right\} + g(\widehat{p}, \widehat{q}) \mu_0 + O_P(\rho_n),$$

where $(\mathbf{X}_{ij})_k = (\mathbf{A}_{ij} - \mathbb{E}(\mathbf{A}_{ij})) (\mathbb{1}[(i, j) \in \mathcal{E}_2(z_0, z_k)] - \mathbb{1}[(i, j) \in \mathcal{E}_1(z_0, z_k)])$. For any $z_0 \in \mathcal{C}_0$, we give the following notations:

$$T_0 = \sup_{k \in [|B_{z_0}(r)|]} \left\{ \frac{1}{\sigma_0} \sum_{i < j} (\mathbf{X}_{ij})_k \right\}, \quad \Xi_0 = \sup_{k \in [|B_{z_0}(r)|]} \left\{ \frac{1}{\sigma_0} \sum_{i < j} \{\xi_{ij}\}_k \right\}, \quad \Xi'_0 = \sup_{k \in [|B_{z_0}(r)|]} \left\{ \frac{1}{\widehat{\sigma}_0} \sum_{i < j} \{\widehat{\xi}_{ij}\}_k \right\},$$

and denote

$$\widetilde{W}_n = W_n/\widehat{\sigma}_0 = \sup_{k \in [|B_{z_0}(r)|]} \left\{ \frac{1}{\widehat{\sigma}_0} \sum_{i < j} (\widehat{\mathbf{X}}_{ij})_k e_{ij} \right\},$$

where $(\widehat{\mathbf{X}}_{ij})_k = (\widehat{\mathbf{A}}_{ij} - \mathbb{E}_{\widehat{p}, \widehat{q}}(\widehat{\mathbf{A}}_{ij}))(\mathbb{1}[(i, j) \in \mathcal{E}_2(z_0, z_k)] - \mathbb{1}[(i, j) \in \mathcal{E}_1(z_0, z_k)])$ and the adjacency matrix $\widehat{\mathbf{A}}$ is generated by \widehat{p}, \widehat{q} , and $\widehat{\sigma}_0 = \sqrt{d(\mathcal{C}_0, \mathcal{C}_1)(\widehat{p}(1 - \widehat{p}) + \widehat{q}(1 - \widehat{q}))}$. ξ_{ij} and $\widehat{\xi}_{ij}$ are the independent mean zero Gaussian vectors with covariance matrix equal to that of \mathbf{X}_{ij} and $\widehat{\mathbf{X}}_{ij}$ respectively ($\{\xi_{ij}\}_k = 0$ if $(i, j) \notin \mathcal{E}_{1,2}(z_0, z_k)$, and the same for $\{\widehat{\xi}_{ij}\}_k$). $\{e_{ij}\}_{i < j}$ are i.i.d standard Gaussian. By Corollary 2.1 in Chernozhukov et al. (2013), we have

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(T_0 \leq t) - \mathbb{P}(\Xi_0 \leq t)| = o(1);$$

Also, by Lemma 3.2 and Corollary 3.1 of Chernozhukov et al. (2013) we have

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(\widetilde{W}_n \leq t | \widehat{\mathbf{X}}_{ij}) - \mathbb{P}(\Xi'_0 \leq t) \right| = o_P(1).$$

We let Σ^{Ξ_0} and $\Sigma^{\Xi'_0}$ be the covariance matrix of the vectors $\left\{ \sum_{i < j} \{\xi_{ij}\}_k / \sigma_0 \right\}_k$ and $\left\{ \sum_{i < j} \{\widehat{\xi}_{ij}\}_k / \widehat{\sigma}_0 \right\}_k$ respectively. Thus for $k, l \in [|B_{z_0}(r)|]$ we have:

$$\begin{aligned} \Sigma_{k,l}^{\Xi_0} &= \frac{1}{\sigma_0^2} \text{Cov}\left(\sum_{ij} \{\xi_{ij}\}_k, \sum_{ij} \{\xi_{ij}\}_l\right) \\ &= \frac{1}{\sigma_0^2} \text{Cov}\left(\sum_{i < j} (\mathbf{X}_{ij})_k, \sum_{i < j} (\mathbf{X}_{ij})_l\right) \\ &= \frac{|\mathcal{E}_2(z_0, z_k) \cap \mathcal{E}_2(z_0, z_l)|q(1 - q) + |\mathcal{E}_1(z_0, z_k) \cap \mathcal{E}_1(z_0, z_l)|p(1 - p)}{d(\mathcal{C}_0, \mathcal{C}_1)(p(1 - p) + q(1 - q))}. \end{aligned}$$

Accordingly,

$$\begin{aligned} \Sigma_{k,l}^{\Xi'_0} &= \frac{1}{\widehat{\sigma}_0^2} \text{Cov}\left(\sum_{i < j} (\widehat{\mathbf{X}}_{ij})_k, \sum_{i < j} (\widehat{\mathbf{X}}_{ij})_l\right) \\ &= \frac{|\mathcal{E}_2(z_0, z_k) \cap \mathcal{E}_2(z_0, z_l)|\widehat{q}(1 - \widehat{q}) + |\mathcal{E}_1(z_0, z_k) \cap \mathcal{E}_1(z_0, z_l)|\widehat{p}(1 - \widehat{p})}{d(\mathcal{C}_0, \mathcal{C}_1)(\widehat{p}(1 - \widehat{p}) + \widehat{q}(1 - \widehat{q}))} \\ &= \frac{|\mathcal{E}_2(z_0, z_k) \cap \mathcal{E}_2(z_0, z_l)|q(1 - q) + |\mathcal{E}_1(z_0, z_k) \cap \mathcal{E}_1(z_0, z_l)|p(1 - p) + O_P(d(\mathcal{C}_0, \mathcal{C}_1)\sqrt{\rho_n}/n)}{d(\mathcal{C}_0, \mathcal{C}_1)(p(1 - p) + q(1 - q)) + O_P(d(\mathcal{C}_0, \mathcal{C}_1)\sqrt{\rho_n}/n)}. \end{aligned}$$

Then we have

$$\Delta_0 = \max_{k,l} |\Sigma_{k,l}^{\Xi_0} - \Sigma_{k,l}^{\Xi'_0}| \leq \left| \frac{O_P(d(\mathcal{C}_0, \mathcal{C}_1)\sqrt{\rho_n}/n)}{\widehat{\sigma}_0^2} \right| + \left| \Sigma_{k,l}^{\Xi_0} \frac{O_P(d(\mathcal{C}_0, \mathcal{C}_1)\sqrt{\rho_n}/n)}{\widehat{\sigma}_0^2} \right| = O_P\left(\frac{1}{\sqrt{n^2 \rho_n}}\right).$$

Thus by Lemma 3.1 in Chernozhukov et al. (2013), there exists a constant C such that

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(\Xi_0 \leq t) - \mathbb{P}(\Xi'_0 \leq t)| \leq C \Delta_0^{1/3} (1 \vee \log(|B_{z_0}(r)|/\Delta_0))^{2/3} = o_P(n^{-1/6-c_2/12}).$$

and thus

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(\Xi_0 \leq t) - \mathbb{P}(\Xi'_0 \leq t)| = o_P(1),$$

and in turn we have

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(T_0 \leq t) - \mathbb{P}(\widetilde{W}_n \leq t | \widehat{\mathbf{X}}_{ij})| = o_P(1).$$

B.2.3 VALIDITY OF MULTIPLIER BOOTSTRAP IN ESTIMATING LRT QUANTILE

Now recall that $C_{\widetilde{W}_n}(\alpha)$ is the α quantile of \widetilde{W}_n conditional on $\widehat{\mathbf{X}}_{ij}$, and we would like to control the order of $C_{\widetilde{W}_n}(\alpha)$ in order to bound the error in estimating the quantile of $\widehat{\text{LRT}}$. Give a constant $t > \sqrt{2c_0}$, we have

$$\begin{aligned} \mathbb{P}(\widetilde{W}_n \geq t\sqrt{\log n} | \widehat{\mathbf{X}}_{ij}) &= \mathbb{P}(\Xi'_0 \geq t\sqrt{\log n}) + o_P(1) \\ &\leq \sum_{k \in [|B_{z_0}(r)|]} \mathbb{P}\left(\left\{\frac{1}{\widehat{\sigma}_0} \sum_{i < j} \{\widehat{\xi}_{ij}\}_k\right\} \geq t\sqrt{\log n}\right) + o_P(1) \\ &\lesssim |B_{z_0}(r)| e^{-\frac{t^2}{2} \log n} + o_P(1) = O_P\left(n^{c_0-t^2/2}\right) + o_P(1) = o_P(1). \end{aligned}$$

Thus we know that $C_{\widetilde{W}_n}(\alpha) = O_P(\sqrt{\log n})$. We know that $q_\alpha = g(\widehat{p}, \widehat{q})\widehat{\sigma}_0 C_{\widetilde{W}_n}(\alpha) + g(\widehat{p}, \widehat{q})\widehat{\mu}_0$, $\widehat{\text{LRT}} = \widehat{L}_0 + \delta_n$ and also $\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |\mathbb{P}(\widehat{L}_0 < t) - \mathbb{P}(\widehat{L}_0(z_0) < t)| = 0$. Therefore,

$$\begin{aligned} \mathbb{P}(\widehat{\text{LRT}} \geq q_\alpha) &= \mathbb{P}(\widehat{L}_0 + \delta_n \geq q_\alpha) = \mathbb{P}(\widehat{L}_0(z_0) + \delta_n \geq q_\alpha) + o(1) \\ &= \mathbb{P}(g(\widehat{p}, \widehat{q})\sigma_0 T_0 + g(\widehat{p}, \widehat{q})\mu_0 + \delta_n \geq g(\widehat{p}, \widehat{q})\widehat{\sigma}_0 C_{\widetilde{W}_n}(\alpha) + g(\widehat{p}, \widehat{q})\widehat{\mu}_0) + o(1) \\ &= \mathbb{P}\left(T_0 \geq \frac{\widehat{\sigma}_0}{\sigma_0} C_{\widetilde{W}_n}(\alpha) + \frac{\widehat{\mu}_0 - \mu_0}{\sigma_0} - \frac{\delta_n}{g(\widehat{p}, \widehat{q})\sigma_0}\right) + o(1). \end{aligned}$$

We have that $|\widehat{\sigma}_0 - \sigma_0| = O_P(\sqrt{d(\mathcal{C}_0, \mathcal{C}_1)}/n)$ and $|\widehat{\mu}_0 - \mu_0| = O_P(d(\mathcal{C}_0, \mathcal{C}_1)\sqrt{\rho_n}/n)$. Therefore,

$$\begin{aligned} \mathbb{P}(\widehat{\text{LRT}} \geq q_\alpha) &= \mathbb{P}\left(T_0 \geq C_{\widetilde{W}_n}(\alpha) + \frac{C_1 \sqrt{d(\mathcal{C}_0, \mathcal{C}_1)}}{\sigma_0 n} C_{\widetilde{W}_n}(\alpha) + \frac{C_2 d(\mathcal{C}_0, \mathcal{C}_1) \sqrt{\rho_n}}{\sigma_0 n} - \frac{\delta_n}{g(\widehat{p}, \widehat{q})\sigma_0}\right) + o(1) \\ &= \mathbb{P}\left(T_0 \geq C_{\widetilde{W}_n}(\alpha) + C_1 \sqrt{\log n}/(n^2 \rho_n) + C_2 \sqrt{d(\mathcal{C}_0, \mathcal{C}_1)}/n + C_3 \sqrt{\frac{\rho_n}{d(\mathcal{C}_0, \mathcal{C}_1)}}\right) + o(1) \\ &= \mathbb{P}(T_0 \geq C_{\widetilde{W}_n}(\alpha) + \Delta_n) + o(1), \end{aligned}$$

where $\Delta_n = o_P(n^{-c})$ for some positive constant $c > 0$. Now from previous results we have

$$\begin{aligned} |\mathbb{P}(\widehat{\text{LRT}} \geq q_\alpha) - \alpha| &\leq |\mathbb{P}(T_0 \geq C_{\widetilde{W}_n}(\alpha) + \Delta_n) - \mathbb{P}(\widetilde{W}_n \geq C_{\widetilde{W}_n}(\alpha) + \Delta_n)| \\ &\quad + |\mathbb{P}(\widetilde{W}_n \geq C_{\widetilde{W}_n}(\alpha) + \Delta_n) - \mathbb{P}(\widetilde{W}_n \geq C_{\widetilde{W}_n}(\alpha))| + o(1) \\ &\leq \mathbb{P}(|\widetilde{W}_n - C_{\widetilde{W}_n}(\alpha)| \leq \Delta_n) + o_P(1). \end{aligned}$$

Now we study the distribution of \widetilde{W}_n : if we denote $Y_k = \frac{1}{\widehat{\sigma}_0} \sum_{i < j} (\widehat{\mathbf{X}}_{ij})_k e_{ij}$, then $Y_k | \widehat{\mathbf{X}} \sim N(0, \sigma_k^2)$, where $\sigma_k^2 = \sum_{i < j} (\widehat{\mathbf{X}}_{ij})_k^2 / \widehat{\sigma}_0^2$, and $\sup_k |\mathbb{E}(\sigma_k^2) - 1| \leq |\widehat{\sigma}_0^2 / \sigma_0^2 - 1| + o_P(1) = o_P(1)$. Also, $|(\widehat{\mathbf{X}}_{ij})_k^2| < 1$. Under the event $\mathcal{A} = \{\widehat{p} = o(1)\} \cap \{\widehat{q} = o(1)\}$ with $\mathbb{P}(\mathcal{A}) = 1 - o(1)$, by Bernstein's inequality, we have

$$\mathbb{P}_{\widehat{\mathbf{X}}}(|\sigma_k^2 - \mathbb{E}_{\widehat{\mathbf{X}}}(\sigma_k^2)| > 1/2) \leq 2 \exp\left(-\frac{\frac{1+1}{8}\widehat{\sigma}_0^4}{(\frac{1}{6} + 1)\widehat{\sigma}_0^2}\right) = 2 \exp\left(-\frac{3\widehat{\sigma}_0^2}{14}\right),$$

where $\mathbb{P}_{\widehat{\mathbf{X}}}$ and $\mathbb{E}_{\widehat{\mathbf{X}}}$ denotes probability and expectation with \widehat{p} and \widehat{q} fixed and consider only the randomness of $\widehat{\mathbf{X}}$. Also

$$\begin{aligned} \mathbb{P}_{\widehat{\mathbf{X}}}(\min_k \sigma_k^2 < 1/2) &\leq \sum_k \mathbb{P}_{\widehat{\mathbf{X}}}(|\sigma_k^2 - \mathbb{E}_{\widehat{\mathbf{X}}}(\sigma_k^2)| > 1/2) \\ &= 2|B_{z_0}(r)| \exp\left(-\frac{3\widehat{\sigma}_0^2}{14}\right) = o_P(1), \end{aligned}$$

where the last $o_P(1)$ term is due to the fact that $\widehat{\sigma}_0^2 = \Omega_P(n\rho_n) = \Omega_P(n^{c_2})$ and $|B_{z_0}(r)| = O(n^{c_0})$. Then by Lemma 2.1 in Chernozhukov et al. (2013), we have

$$\begin{aligned} \mathbb{P}(|\widetilde{W}_n - C_{\widetilde{W}_n}(\alpha)| \leq \Delta_n) &= \mathbb{P}(|\max_k Y_k - C_{\widetilde{W}_n}(\alpha)| \leq \Delta_n) \leq \sup_{z \in \mathbb{R}} \mathbb{P}(|\max_k Y_k - z| \leq \Delta_n) \\ &= O_P\left(\Delta_n \left\{ \sqrt{2 \log |B_{z_0}(r)|} + \sqrt{\log(\min_k \sigma_k^2 / \Delta_n)} \right\}\right) = o_P(1), \end{aligned}$$

and thus $\lim_{n \rightarrow \infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\widehat{\text{LRT}} \geq q_\alpha) = \alpha$.

As for the Type I error, from the preceding proof we see that $\mathbb{P}(\text{LRT} \geq q_\alpha) = \alpha + o_P(1)$, and the convergence of the $o_P(1)$ term is independent of $z^* \in \mathcal{C}_0$ due to the symmetry of \mathcal{C}_0 . Therefore, we have

$$\sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\text{reject } H_0) = \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\text{LRT} \geq q_\alpha) = \alpha + o_P(1),$$

and hence the claim follows. As for the Type II error, when the true assignment is $z^* \in \mathcal{C}_1$, by (B.2) in Lemma B.2, we have

$$\begin{aligned} \text{LRT} &= \log \frac{\sup_{z \in \mathcal{C}_1} f(\mathbf{A}; z, \hat{p}, \hat{q})}{\sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} f(\mathbf{A}; z, \hat{p}, \hat{q})} \\ &= \log \frac{\sup_{z \in \mathcal{C}_1} f(\mathbf{A}; z, \hat{p}, \hat{q})}{f(\mathbf{A}; z^*, \hat{p}, \hat{q})} + \log \frac{f(\mathbf{A}; z^*, \hat{p}, \hat{q})}{\sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} f(\mathbf{A}; z, \hat{p}, \hat{q})} = O_P(\rho_n). \end{aligned}$$

And since $\hat{\sigma}_0 \asymp \sqrt{d(\mathcal{C}_0, \mathcal{C}_1)\hat{p}}\hat{\mu}_0 \asymp -d(\mathcal{C}_0, \mathcal{C}_1)\hat{p} = -\Omega_P(n^{c_2})$ and $C_{\widehat{W}_n}(\alpha) = O_P(\sqrt{\log n})$, we have $q_\alpha = g(\hat{p}, \hat{q})\hat{\sigma}_0 C_{\widehat{W}_n}(\alpha) + g(\hat{p}, \hat{q})\hat{\mu}_0 \rightarrow -\infty$. Since the convergence is independent of z^* , we have for any true assignment $z_1 \in \mathcal{C}_1$,

$$\inf_{z^* \in \mathcal{C}_1} \mathbb{P}(\text{reject } H_0) = 1 - \sup_{z^* \in \mathcal{C}_1} \mathbb{P}(\text{LRT} \leq q_\alpha) = 1 - o_P(1).$$

Appendix C. Proof of Theorems for the Lower Bound

In this section, we will prove the theorems for the lower bound. Similar as the upper bound, since Theorem 5.5 and Theorem 5.6 are general versions of Theorem 4.1 and Theorem 4.2, we will only prove the general versions and the proof can be applied to Theorem 4.1 and Theorem 4.2, too. Also, the proof of Theorem 5.5 is actually based on the proof of Theorem 5.6 under a stronger regime. Therefore, we will prove the two theorems together: we will first prove Theorem 5.6 under more general conditions, and then we will apply the proof of Theorem 5.6 to the proof of Theorem 5.5 under stronger conditions.

C.1 Proof of Theorem 5.5 and Theorem 5.6

The proof proceeds in the following order: we first prove the results under the two conditions of Theorem 5.6, namesly the proof of Theorem 5.6 (1) and the proof of Theorem 5.6 (2), then we provide the proof of Theorem 5.5.

C.1.1 PROOF OF THEOREM 5.6 (1)

As for the minimax rate, we have:

$$\begin{aligned} r(\mathcal{C}_0, \mathcal{C}_1) &= \min_{\psi} \left\{ \sup_{z \in \mathcal{C}_0} \mathbb{P}_z(\psi = 1) + \sup_{z \in \mathcal{C}_1} \mathbb{P}_z(\psi = 0) \right\} \\ &\geq \min_{\psi} \left\{ \mathbb{P}_{z_0}(\psi = 1) + \mathbb{P}_{z_1}(\psi = 0) \right\}. \end{aligned}$$

where z_0 and z_1 are fixed cluster assignments in \mathcal{C}_0 and \mathcal{C}_1 respectively. For a given adjacency matrix \mathbf{A} , we know that ψ is a function of \mathbf{A} , and the only information of \mathbf{A} relevant to classification of the true assignment is $\{\mathbf{A}_{ij}, (i, j) \in \mathcal{E}_1(z_0, z_1) \cup \mathcal{E}_2(z_0, z_1)\}$. Larger size of $\mathcal{E}_1(z_0, z_1)$ and $\mathcal{E}_2(z_0, z_1)$ will provide more information and lead to smaller type I and type II error. Thus, the worst case is when the size of $\mathcal{E}_1(z_0, z_1)$ and $\mathcal{E}_2(z_0, z_1)$ obtains the infimum, i.e., $d(z_0, z_1) = n_1(z_0, z_1) \vee n_2(z_0, z_1) = d(\mathcal{C}_0, \mathcal{C}_1)$.

To obtain $\inf_{\psi} \left\{ \sup_{z \in \mathcal{C}_0} \mathbb{P}_z(\psi = 1) + \sup_{z \in \mathcal{C}_1} \mathbb{P}_z(\psi = 0) \right\}$, the optimal method $\tilde{\psi}$ must be

the mode of the posterior distribution. For the convenience of notations, we denote $L(z, \mathbf{A})$ as $f(\mathbf{A}; z, p, q)$, and n_i as $n_i(z_0, z_1)$, $i = 1, 2$ for short:

$$\begin{aligned} L(z_0, \mathbf{A}) &\propto p^{\sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} (1-p)^{n_1 - \sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} q^{\sum_{(i,j) \in \mathcal{E}_2(z_0, z_1)} \mathbf{A}_{ij}} (1-q)^{n_2 - \sum_{(i,j) \in \mathcal{E}_2(z_0, z_1)} \mathbf{A}_{ij}} \\ L(z_1, \mathbf{A}) &\propto p^{\sum_{(i,j) \in \mathcal{E}_2(z_0, z_1)} \mathbf{A}_{ij}} (1-p)^{n_2 - \sum_{(i,j) \in \mathcal{E}_2(z_0, z_1)} \mathbf{A}_{ij}} q^{\sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} (1-q)^{n_1 - \sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} \end{aligned}$$

and correspondingly,

$$\tilde{\psi}(\mathbf{A}) = \begin{cases} 0, & \text{if } L(z_0, \mathbf{A}) > L(z_1, \mathbf{A}); \\ 1, & \text{if } L(z_0, \mathbf{A}) \leq L(z_1, \mathbf{A}). \end{cases}$$

Then $\mathbb{P}_{z_0}(\tilde{\psi} = 1) = \mathbb{P}_{z_0}(L(z_0, \mathbf{A}) \leq L(z_1, \mathbf{A}))$ and $\mathbb{P}_{z_1}(\tilde{\psi} = 0) = \mathbb{P}_{z_1}(L(z_0, \mathbf{A}) > L(z_1, \mathbf{A}))$. Without loss of generality, we assume that $n_1(z_0, z_1) \geq n_2(z_0, z_1)$. Then, if we expend the size of $\mathcal{E}_2(z_0, z_1)$ to be the same as $\mathcal{E}_1(z_0, z_1)$, adding i.i.d entries $\{\mathbf{A}_{ij}, (i, j) \in \mathcal{E}_2^L(z_0, z_1) \setminus \mathcal{E}_2(z_0, z_1)\}$ conforming to the same distribution as $\{\mathbf{A}_{ij}, (i, j) \in \mathcal{E}_2(z_0, z_1)\}$, more information will be provided and the error rate will decrease, where $\mathcal{E}_2^L(z_0, z_1)$ denotes the set expended on $\mathcal{E}_2(z_0, z_1)$, and we have:

$$\begin{aligned} \tilde{L}(z_0, \mathbf{A}) &\propto p^{\sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} (1-p)^{n_1 - \sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} q^{\sum_{(i,j) \in \mathcal{E}_2^L(z_0, z_1)} \mathbf{A}_{ij}} (1-q)^{n_1 - \sum_{(i,j) \in \mathcal{E}_2^L(z_0, z_1)} \mathbf{A}_{ij}} \\ \tilde{L}(z_1, \mathbf{A}) &\propto p^{\sum_{(i,j) \in \mathcal{E}_2^L(z_0, z_1)} \mathbf{A}_{ij}} (1-p)^{n_1 - \sum_{(i,j) \in \mathcal{E}_2^L(z_0, z_1)} \mathbf{A}_{ij}} q^{\sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} (1-q)^{n_1 - \sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij}} \end{aligned}$$

Thus we can obtain lower bound on the minimax rate:

$$\begin{aligned} r(\tilde{\psi}) &= \mathbb{P}_{z_0}(L(z_0, \mathbf{A}) \leq L(z_1, \mathbf{A})) + \mathbb{P}_{z_1}(L(z_0, \mathbf{A}) > L(z_1, \mathbf{A})) \\ &\geq \mathbb{P}_{z_0}(\tilde{L}(z_0, \mathbf{A}) \leq \tilde{L}(z_1, \mathbf{A})) + \mathbb{P}_{z_1}(\tilde{L}(z_0, \mathbf{A}) > \tilde{L}(z_1, \mathbf{A})) \\ &= \mathbb{P}_{z_0}\left(\sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij} \leq \sum_{(i,j) \in \mathcal{E}_2^L(z_0, z_1)} \mathbf{A}_{ij}\right) + \mathbb{P}_{z_1}\left(\sum_{(i,j) \in \mathcal{E}_1(z_0, z_1)} \mathbf{A}_{ij} > \sum_{(i,j) \in \mathcal{E}_2^L(z_0, z_1)} \mathbf{A}_{ij}\right) \\ &\geq 2\mathbb{P}\left(\sum_{u=1}^{n_1} X^u \geq \sum_{u=1}^{n_1} Y^u\right). \end{aligned}$$

where $\{X_u\} \stackrel{\text{i.i.d}}{\sim} \text{Ber}(q)$, $\{Y^u\} \stackrel{\text{i.i.d}}{\sim} \text{Ber}(p)$, and $\{X^u\}$ are independent to $\{Y^u\}$.

Now $n_1 = d(\mathcal{C}_0, \mathcal{C}_1)$, and both p and q change with n_1 . We have $\mathbb{E}(|X^u - Y^u - \mathbb{E}[X^u - Y^u]|^3) \asymp p(1-q) + q(1-p)$. Since $0 < q < p < 1 - \delta$, we have $\delta p < p(1-q) + q(1-p) < 2p$. Thus $\mathbb{E}(|X^u - Y^u - \mathbb{E}[X^u - Y^u]|^3) \asymp p$. Similarly $\text{Var}(X^u - Y^u) \asymp p$. Thus,

$$\frac{\sum_{u=1}^{n_1} \mathbb{E}(|X^u - Y^u - \mathbb{E}[X^u - Y^u]|^3)}{\text{Var}(\sum_{u=1}^{n_1} \{X^u - Y^u\})^{3/2}} \asymp \frac{n_1[p(1-q) + q(1-p)]}{n_1^{3/2}(q(1-q) + p(1-p))^{3/2}} \asymp \frac{1}{\sqrt{n_1 p}} \rightarrow 0$$

Therefore, by the Lyapunov's Central Limit Theorem and the independence of $\{X^u\}$ and $\{Y^u\}$, as $n \rightarrow \infty$, we have $\sum_{u=1}^{n_1} X^u - \sum_{u=1}^{n_1} Y^u \xrightarrow{d} N(n_1(q-p), n_1q(1-q) + n_1p(1-p))$.

Therefore,

$$\mathbb{P}(\sum_{u=1}^{n_1} X^u \geq \sum_{u=1}^{n_1} Y^u) = \mathbb{P}(\sum_{u=1}^{n_1} X^u - \sum_{u=1}^{n_1} Y^u \geq 0) = 1 - \Phi\left(\frac{\sqrt{n_1}(p-q)}{\sqrt{p(1-p)+q(1-q)}}\right) + o(1).$$

When $\limsup_{n \rightarrow \infty} n_1 I(p, q) = O(1)$, we can see that $p - q = o(1)$. We have

$$\begin{aligned} I(p, q) &= -2 \log \left(1 - \frac{(\sqrt{p} - \sqrt{q})^2 + (\sqrt{1-p} - \sqrt{1-q})^2}{2} \right) \\ &= \left(\frac{(p-q)^2}{(\sqrt{p} + \sqrt{q})^2} + \frac{(p-q)^2}{(\sqrt{1-p} + \sqrt{1-q})^2} \right) (1 + o(1)) \\ &\geq \frac{\delta}{2} \frac{(p-q)^2}{p(1-p) + q(1-q)} (1 + o(1)). \end{aligned}$$

Thus, if $\limsup_{n \rightarrow \infty} n_1 I(p, q) \leq \delta \Phi^{-1}(3/4)^2/2$, namely, $\limsup_{n \rightarrow \infty} \sqrt{n_1}(p-q)/\sqrt{p(1-p)+q(1-q)} \leq \limsup_{n \rightarrow \infty} \sqrt{2n_1 I(p, q)/\delta} \leq \Phi^{-1}(3/4)$, we have

$$\mathbb{P}(\sum_{u=1}^{n_1} X^u \geq \sum_{u=1}^{n_1} Y^u) \geq 1 - \Phi(\sqrt{n_1}(p-q)/\sqrt{p(1-p)+q(1-q)}) \geq 1/4.$$

and

$$r(\mathcal{C}_0, \mathcal{C}_1) \geq 2(1 - \Phi(\sqrt{n_1}(p-q)/\sqrt{p(1-p)+q(1-q)})) \geq 1/2.$$

C.1.2 PROOF OF THEOREM 5.6 (2)

When $d(z_0, \mathcal{C}_1)I(p, q) \rightarrow \infty$, if there exists a $z_0 \in \mathcal{C}_0$ and some $r = d(z_0, \mathcal{C}_1) + O(1)$ such that $\limsup_{n \rightarrow \infty} d(\mathcal{C}_0, \mathcal{C}_1)I(p, q)/\log N(B_{z_0}(r), 0) < 1$, then we take a 0-packing $\mathcal{P}(B_{z_0}(r), 0)$ (denoted $\mathcal{P}(0)$ for short) of the ball $B_{z_0}(r)$, and we have:

$$\begin{aligned} r(\mathcal{C}_0, \mathcal{C}_1) &= \min_{\psi} \left\{ \sup_{z \in \mathcal{C}_0} \mathbb{P}_z(\psi = 1) + \sup_{z \in \mathcal{C}_1} \mathbb{P}_z(\psi = 0) \right\} \geq \min_{\psi} \left\{ \mathbb{P}_{z_0}(\psi = 1) + \sup_{z \in \mathcal{P}(0)} \mathbb{P}_z(\psi = 0) \right\} \\ &= \min_{\psi} \left\{ \sum_{\mathbf{A}} \left(\mathbb{P}_{z_0}(\psi = 1 | \mathbf{A} = \mathbf{A}) \mathbb{P}_{z_0}(\mathbf{A} = \mathbf{A}) + \sup_{z \in \mathcal{P}(0)} \mathbb{P}_z(\psi = 0 | \mathbf{A} = \mathbf{A}) \mathbb{P}_z(\mathbf{A} = \mathbf{A}) \right) \right\} \\ &= \min_{\psi} \left\{ \sum_{\mathbf{A}} \left(\mathbb{1}(\psi(\mathbf{A}) = 1) \mathbb{P}_{z_0}(\mathbf{A} = \mathbf{A}) + \mathbb{1}(\psi(\mathbf{A}) = 0) \sup_{z \in \mathcal{P}(0)} \mathbb{P}_z(\mathbf{A} = \mathbf{A}) \right) \right\}, \end{aligned}$$

where the sum over \mathbf{A} is the summation over all possible realizations of the adjacency matrix \mathbf{A} . Thus the optimal method $\tilde{\psi}$ in this scenario should be:

$$\tilde{\psi}(\mathbf{A}) = \begin{cases} 0, & \text{if } L(z_0, \mathbf{A} = \mathbf{A}) \geq \sup_{z \in \mathcal{P}(0)} L(z, \mathbf{A} = \mathbf{A}); \\ 1, & \text{if } L(z_0, \mathbf{A} = \mathbf{A}) < \sup_{z \in \mathcal{P}(0)} L(z, \mathbf{A} = \mathbf{A}). \end{cases}$$

and we have $L(z_0, \mathbf{A} = A) < \sup_{z \in \mathcal{P}(0)} L(z, \mathbf{A} = A)$

$$\begin{aligned}
 r(\mathcal{C}_0, \mathcal{C}_1) &\geq \mathbb{P}_{z_0}(\tilde{\psi} = 1) + \sup_{z \in \mathcal{P}(0)} \mathbb{P}_z(\tilde{\psi} = 0) \\
 &= \mathbb{P}_{z_0} \left(L(z_0, \mathbf{A} = A) < \sup_{z \in \mathcal{P}(0)} L(z, \mathbf{A} = A) \right) \\
 &\quad + \sup_{z \in \mathcal{P}(0)} \mathbb{P}_z \left(L(z_0, \mathbf{A} = A) \geq \sup_{z \in \mathcal{P}(0)} L(z, \mathbf{A} = A) \right) \\
 &= \mathbb{P}_{z_0} \left(\sup_{z \in \mathcal{P}(0)} \log L(z, \mathbf{A} = A) - \log L(z_0, \mathbf{A} = A) > 0 \right) \\
 &\quad + \sup_{z \in \mathcal{P}(0)} \mathbb{P}_z \left(\sup_{z \in \mathcal{P}(0)} \log L(z, \mathbf{A} = A) - \log L(z_0, \mathbf{A} = A) \leq 0 \right).
 \end{aligned}$$

Similar with the case when $d(z_0, \mathcal{C}_1)I(p, q) = O(1)$, we can expand each $\mathcal{E}_2(z_0, z)$ to $\mathcal{E}_2^L(z_0, z)$ (or $\mathcal{E}_1(z_0, z)$ to $\mathcal{E}_1^L(z_0, z)$, we use the former notation for convenience) so that $\mathcal{E}_1(z_0, z)$ and $\mathcal{E}_2(z_0, z)$ are of equal sizes, and then we have

$$\begin{aligned}
 r(\mathcal{C}_0, \mathcal{C}_1) &\geq \mathbb{P}_{z_0} \left(\sup_{z \in \mathcal{P}(0)} \left(\sum_{(i,j) \in \mathcal{E}_2^L(z_0, z)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z_0, z)} \mathbf{A}_{ij} \right) > 0 \right) \\
 &\quad + \sup_{z \in \mathcal{P}(0)} \mathbb{P}_z \left(\sup_{z \in \mathcal{P}(0)} \left(\sum_{(i,j) \in \mathcal{E}_2^L(z_0, z)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z_0, z)} \mathbf{A}_{ij} \right) \leq 0 \right) \\
 &\geq \mathbb{P}_{z_0} \left(\sup_{z \in \mathcal{P}(0)} \left(\sum_{(i,j) \in \mathcal{E}_2^L(z_0, z)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z_0, z)} \mathbf{A}_{ij} \right) > 0 \right) \\
 &= \mathbb{P}_{z_0} \left(\sup_{z \in \mathcal{P}(0)} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u \right) > 0 \right),
 \end{aligned}$$

where $\{X_z^u\} \stackrel{\text{i.i.d}}{\sim} \text{Ber}(q)$, $\{Y_z^u\} \stackrel{\text{i.i.d}}{\sim} \text{Ber}(p)$, $\{X_{z_i}^u\} \perp \{X_{z_j}^u\}, i \neq j$, $\{Y_{z_i}^u\} \perp \{Y_{z_j}^u\}, i \neq j$ and $\{X_{z_i}^u\} \perp \{Y_{z_j}^u\}, \forall i, j$. By Lemma 5.2 in Zhang and Zhou (2016), we know that there exists $\eta \rightarrow 0$ such that

$$\mathbb{P}_{z_0} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u > 0 \right) \geq \exp \left(- (1 + \eta) d(z_0, \mathcal{C}_1) I(p, q) \right).$$

When $\limsup_{n \rightarrow \infty} d(z_0, \mathcal{C}_1) I(p, q) / \log |\mathcal{P}(0)| < 1$, for sufficiently large n we have $(1 + \eta) d(z_0, \mathcal{C}_1) I(p, q) \leq \log |\mathcal{P}(0)|$, and since $x > 1 - (1/2)^x$ for $x > 0$, we have that for n large enough

$$\begin{aligned}
 \mathbb{P}_{z_0} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u > 0 \right) &\geq \exp \left(- (1 + \eta) d(z_0, \mathcal{C}_1) I(p, q) \right) \\
 &\geq \exp \left(- \log |\mathcal{P}(0)| \right) = 1/|\mathcal{P}(0)| \geq 1 - (1/2)^{1/|\mathcal{P}(0)|},
 \end{aligned}$$

and thus

$$\begin{aligned}
r(\mathcal{C}_0, \mathcal{C}_1) &\geq \mathbb{P}_{z_0} \left(\sup_{z \in \mathcal{P}(0)} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u \right) > 0 \right) = 1 - \mathbb{P}_{z_0} \left(\sup_{z \in \mathcal{P}(0)} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u \right) \leq 0 \right) \\
&= 1 - \Pi_{z \in \mathcal{P}(0)} \mathbb{P}_{z_0} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u \leq 0 \right) \geq 1 - \left\{ (1/2)^{1/|\mathcal{P}(0)|} \right\}^{|\mathcal{P}(0)|} = 1/2.
\end{aligned}$$

The statement is true for any 0-packing of the ball $B_{z_0}(r)$, and thus the statement follows.

C.1.3 PROOF OF THEOREM 5.5

Under the regime $1/\rho_n = o(n^{1-c_2})$, we take one $\sqrt{d(z_0, \mathcal{C}_1)}$ -packing $\mathcal{P}(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)})$ (denoted $\tilde{\mathcal{P}}$ for short) of the ball $B_{z_0}(r)$, similar with the proof of Theorem 3.2, by Corollary 2.1 in Chernozhukov et al. (2013), we have:

$$\begin{aligned}
r(\mathcal{C}_0, \mathcal{C}_1) &\geq \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u \right) > 0 \right) \\
&= \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0, \mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} Y_z^u \right) > \delta_n \right) \\
&= \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0, \mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} Y_z^u + d(z_0, \mathcal{C}_1)(p - q) \right) > d(z_0, \mathcal{C}_1)(p - q) + \delta_n \right) \\
&= \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \xi_z > \frac{d(z_0, \mathcal{C}_1)(p - q)}{\sigma_d} + \delta_n/\sigma_d \right) + o(1).
\end{aligned}$$

where $\delta_n = \sup_{z \in \tilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0, \mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} Y_z^u \right) - \sup_{z \in \tilde{\mathcal{P}}} \left(\sum_{u=1}^{n_1(z_0, z)} X_z^u - \sum_{u=1}^{n_1(z_0, z)} Y_z^u \right) = O(1)$ and $\sigma_d = \sqrt{d(z_0, \mathcal{C}_1)(p(1-p) + q(1-q))}$, and $\{\xi_z\}_{z \in \tilde{\mathcal{P}}}$ are standard Gaussian variables with the same covariance matrix as $\{(\sum_{u=1}^{d(z_0, \mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} Y_z^u + d(z_0, \mathcal{C}_1)(p - q))/\sigma_d\}_{z \in \tilde{\mathcal{P}}}$.

By Lemma 2.1 in Chernozhukov et al. (2013), combined with the fact that $d(z_0, \mathcal{C}_1) = \Omega_P(n)$ and $1/\rho_n = o(n^{1-c_2})$, we have that

$$\left| \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \xi_z > \frac{d(z_0, \mathcal{C}_1)(p - q)}{\sigma_d} + \delta_n/\sigma_d \right) - \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \xi_z > \frac{d(z_0, \mathcal{C}_1)(p - q)}{\sigma_d} \right) \right| \lesssim \frac{\delta_n}{\sigma_d} \sqrt{\log n} = o(1).$$

We let $\{\tilde{X}_z^u\}_{u,z}$ be i.i.d Ber(q) random variables and $\{\tilde{Y}_z^u\}_{u,z}$ be i.i.d Ber(p) random variables, and $\{\tilde{X}_z^u\}_{u,z}$ and $\{\tilde{Y}_z^u\}_{u,z}$ are independent of each other. Then for each $z \in \tilde{\mathcal{P}}$, $\sum_{u=1}^{d(z_0, \mathcal{C}_1)} \tilde{X}_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} \tilde{Y}_z^u + d(z_0, \mathcal{C}_1)(p - q)$ shares the same distribution with $\sum_{u=1}^{d(z_0, \mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} Y_z^u + d(z_0, \mathcal{C}_1)(p - q)$. We let $\{\tilde{\xi}_z\}_{z \in \tilde{\mathcal{P}}}$ be the corresponding Gaussian analog of $\{(\sum_{u=1}^{d(z_0, \mathcal{C}_1)} \tilde{X}_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} \tilde{Y}_z^u + d(z_0, \mathcal{C}_1)(p - q))/\sigma_d\}_{z \in \tilde{\mathcal{P}}}$.

$\sum_{u=1}^{d(z_0, \mathcal{C}_1)} \tilde{Y}_z^u + d(z_0, \mathcal{C}_1)(p - q)) / \sigma_d \}_{z \in \tilde{\mathcal{P}}}$. Then we have:

$$|\text{Cov}(\xi_{z_i}, \xi_{z_j}) - \text{Cov}(\tilde{\xi}_{z_i}, \tilde{\xi}_{z_j})| = \begin{cases} 0 & \text{if } i = j, \\ O(\frac{1}{\sqrt{d(z_0, \mathcal{C}_1)}}) & \text{if } i \neq j. \end{cases}$$

Thus by Lemma 3.1 in Chernozhukov et al. (2013), we have $\Delta_0 = O(1/\sqrt{d(z_0, \mathcal{C}_1)})$, and

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \xi_z > t \right) - \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \tilde{\xi}_z > t \right) \right| \leq C \Delta_0^{1/3} (\log |\tilde{\mathcal{P}}| / \Delta_0)^{2/3} = o(1),$$

and thus

$$\begin{aligned} & \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0, \mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} Y_z^u + d(z_0, \mathcal{C}_1)(p - q) \right) > d(z_0, \mathcal{C}_1)(p - q) \right) \\ &= \mathbb{P}_{z_0} \left(\sup_{z \in \tilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0, \mathcal{C}_1)} \tilde{X}_z^u - \sum_{u=1}^{d(z_0, \mathcal{C}_1)} \tilde{Y}_z^u + d(z_0, \mathcal{C}_1)(p - q) \right) > d(z_0, \mathcal{C}_1)(p - q) \right) + o(1). \end{aligned}$$

Then similar with previous proof, we have when $\limsup \lim_{n \rightarrow \infty} d(z_0, \mathcal{C}_1)I(p, q) / \log |\tilde{\mathcal{P}}| < 1$,

$$r(\mathcal{C}_0, \mathcal{C}_1) \geq \mathbb{P}_{z_0} \left(\sup_{z \in B(r_K)} \left(\sum_{u=1}^{n_1(z_0, z)} \tilde{X}_z^u - \sum_{u=1}^{n_1(z_0, z)} \tilde{Y}_z^u \right) > 0 \right) + o(1) \geq 1/2 + o(1).$$

Also the results hold for any $\sqrt{d(z_0, \mathcal{C}_1)}$ -packing of the ball $B_{z_0}(r)$. Therefore, we proved the claim.

Appendix D. Proof of technical lemmas

Now we will provide proofs for the technical lemmas used for the proof of Theorem 5.2.

D.1 Proof of Lemma 2.2

It suffices for us to prove Lemma B.1, the more general version of Lemma 2.2. Due to the structure of \mathbf{L}_z , it suffices for us to prove that the edge-wise distance between assignments are permutation-invariant.

For any given $z_0 \in \mathcal{C}_0$ and $z_1, z'_1 \in \mathcal{C}_1$, we have:

$$\begin{aligned}
n_1(z_0, z_1) &= \sum_{i < j, i, j \in [n]} \mathbb{1}(z_0(i) = z_0(j), z_1(i) \neq z_1(j)) \\
&= \sum_{i < j, i, j \in [n]} \mathbb{1}(\sigma(z_0(i)) = \sigma(z_0(j)), \sigma(z_1(i)) \neq \sigma(z_1(j))) \\
&= \sum_{\tau(i) < \tau(j), i, j \in [n]} \mathbb{1}(\tau \circ \sigma(z_0)(\tau(i)) = \tau \circ \sigma(z_0)(\tau(j)), \tau \circ \sigma(z_1)(\tau(i)) \neq \tau \circ \sigma(z_1)(\tau(j))) \\
&= n_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z_1)).
\end{aligned}$$

Then very similarly we have $n_2(z_0, z_1) = n_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z_1))$ and thus $d(z_0, z_1) = n_1(z_0, z_1) \vee n_2(z_0, z_1) = n_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z_1)) \vee n_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z_1)) = d(\tau \circ \sigma(z_0), \tau \circ \sigma(z_1))$. This suggests that the permutation $\tau \circ \sigma$ does not change the distance between assignments. Also,

$$\begin{aligned}
|\mathcal{E}_1(z_0, z_1) \cap \mathcal{E}_1(z_0, z'_1)| &= \sum_{i < j, i, j \in [n]} \mathbb{1}(z_0(i) = z_0(j), z_1(i) \neq z_1(j), z'_1(i) \neq z'_1(j)) \\
&= \sum_{i < j, i, j \in [n]} \mathbb{1}(\tau \circ \sigma(z_0)(i) = \tau \circ \sigma(z_0)(j), \tau \circ \sigma(z_1)(i) \neq \tau \circ \sigma(z'_1)(j), \tau \circ \sigma(z'_1)(i) \neq \tau \circ \sigma(z_1)(j)) \\
&= |\mathcal{E}_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z_1)) \cap \mathcal{E}_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z'_1))|.
\end{aligned}$$

And similarly,

$$|\mathcal{E}_2(z_0, z_1) \cap \mathcal{E}_2(z_0, z'_1)| = |\mathcal{E}_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z_1)) \cap \mathcal{E}_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z'_1))|.$$

Thus the cardinality of the intersection of the sets $\mathcal{E}_i, i = 1, 2$ is also invariant under the permutation $\tau \circ \sigma$.

Now for any $z_0, z'_0 \in \mathcal{C}_0$, if $\tau \circ \sigma(z_0) = z'_0$, and $d(z_0, z_1) = d(z_0, \mathcal{C}_1)$, from previous results we have $d(z'_0, \tau \circ \sigma(z_1)) = d(z_0, \mathcal{C}_1)$. If there exists an assignment $z'_1 \in \mathcal{C}_1$ such that $d(z'_0, z'_1) < d(z'_0, \tau \circ \sigma(z_1))$, then $d(z_0, (\tau \circ \sigma)^{-1}(z'_1)) = d(z'_0, z'_1) < d(z_0, \mathcal{C}_1)$ due to the fact that $\tau \circ \sigma$ is a one to one mapping. Since $z_0 = (\tau \circ \sigma)^{-1}(z'_0)$, we know that \mathcal{C}_1 is closed under $(\tau \circ \sigma)^{-1}$ and $(\tau \circ \sigma)^{-1}(z'_1) \in \mathcal{C}_1$. This is contradictory to the fact that $z_1 = \operatorname{argmin}_{z \in \mathcal{C}_1} d(z_0, z)$. Therefore, $d(z'_0, \mathcal{C}_1) = d(z'_0, \tau \circ \sigma(z_1)) = d(z_0, \mathcal{C}_1)$.

Similarly, if $z_1 \in B_{z_0}(r)$, then $\tau \circ \sigma(z_1) \in B_{z'_0}(r)$. If $z'_1 \in B_{z'_0}(r)$, then $(\tau \circ \sigma)^{-1}(z'_1) \in B_{z_0}(r)$. Therefore, $\tau \circ \sigma$ is a one to one mapping from $B_{z_0}(r)$ to $B_{z'_0}(r)$, and $|B_{z_0}(r)| = |B_{z'_0}(r)|$.

Now for a given radius r , we find the permutation $\tau \in S_{|B_{z'_0}(r)|}$ such that $\tau(z_i) = \tau \circ \sigma(z_i) = z'_i$ for $z_i \in B_{z_0}(r)$ and $z'_i \in B_{z'_0}(r)$.

When the true assignment is z_0 , the (k, l) -th entry of the covariance matrix for the vector \mathbf{L}_{z_0} can be expressed as

$$\begin{aligned}
 \text{Cov}(\mathbf{L}_{z_0})_{kl} &= g(p, q)^2 \left(|\mathcal{E}_2(z_0, z_k) \cap \mathcal{E}_2(z_0, z_l)| q(1 - q) + |\mathcal{E}_1(z_0, z_k) \cap \mathcal{E}_1(z_0, z_l)| p(1 - p) \right) \\
 &= g(p, q)^2 \left(|\mathcal{E}_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z_k)) \cap \mathcal{E}_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z_l))| q(1 - q) \right. \\
 &\quad \left. + |\mathcal{E}_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z_k)) \cap \mathcal{E}_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z_l))| p(1 - p) \right) \\
 &= \text{Cov} \left(g(p, q) \left(\sum_{\mathcal{E}_2(\tau(z'_0), \tau(z'_k))} \mathbf{A}_{ij} - \sum_{\mathcal{E}_1(\tau(z'_0), \tau(z'_k))} \mathbf{A}_{ij} \right), g(p, q) \left(\sum_{\mathcal{E}_2(\tau(z'_0), \tau(z'_l))} \mathbf{A}_{ij} - \sum_{\mathcal{E}_1(\tau(z'_0), \tau(z'_l))} \mathbf{A}_{ij} \right) \right) \\
 &= \text{Cov}(\mathbf{L}_{z'_0})_{\tau(k)\tau(l)}.
 \end{aligned}$$

Hence we finish the proof.

D.2 Proof of Lemma B.2

The proof mainly follows from Lemma 2.3 and Lemma 2.6 in Wang and Bickel (2017) with modifications for the function $F(\cdot)$ and $G(\cdot)$. We provide the sketch of proof as following: We first define the count statistics as proposed in Wang and Bickel (2017):

$$\mathbf{A}_{i,j} | (z(i) = a, z(j) = b) \sim \text{Ber}(\mathbf{H}_{a,b}), i \neq j, a, b \in [K].$$

where $\mathbf{H}_{a,b} = p = \lambda_1 \rho_n$ if $a = b$, and $\mathbf{H}_{a,b} = q = \lambda_2 \rho_n$ if $a \neq b$. $\mathbf{H} = \rho_n \mathbf{S}$.

$$\mathbf{O}_{a,b}(z) = \sum_{i=1}^n \sum_{j \neq i} \mathbb{1}(z(i) = a, z(j) = b) \mathbf{A}_{ij},$$

$$L = \sum_{i=1}^n \sum_{j=i+1}^n \mathbf{A}_{ij}, \mu_n = n^2 \rho_n.$$

For two assignments z, z' , The confusion matrix is:

$$\mathbf{R}_{k,a}(z, z') = n^{-1} \sum_{i=1}^n \mathbb{1}(z(i) = k, z'(i) = a).$$

By definition, we have $|n_k(z) - n/K| \leq c_K, \forall z \in \mathcal{C}_0 \cup \mathcal{C}_1, \forall k = 1, 2, \dots, K$. We let $\tilde{n}(z)$ denote the number of within-cluster edges, and assume

$$n_k(z) = n/K + a_k, |a_k| \leq c_K, k = 1, 2, \dots, K,$$

$$\sum_{k=1}^K a_k = 0.$$

Then

$$\begin{aligned}\tilde{n}(z) &= \frac{\sum_{k=1}^K (n/K + a_k)^2 - n}{2} = \frac{n^2/K - n}{2} + \frac{\sum_{k=1}^K a_k^2}{2} \\ &\leq \frac{n^2/K - n}{2} + Kc_K^2/2.\end{aligned}$$

Therefore, $\forall z, z' \in \mathcal{C}_1$ we have $\tilde{n}(z) + n_2(z, z') - n_1(z, z') = \tilde{n}(z')$, $|n_2(z, z') - n_1(z, z')| = |\tilde{n}(z) - \tilde{n}(z')| \leq Kc_K^2/2$. Thus, we denote z^* as the true assignment, and $\forall z \in \mathcal{C}_0 \cup \mathcal{C}_1$ we have

$$\begin{aligned}\log f(\mathbf{A}; z, \hat{p}, \hat{q}) &= \frac{1}{2} \left(\sum_{a,b=1}^K \mathbf{O}_{a,b}(z) \log \frac{\hat{\mathbf{H}}_{a,b}}{1 - \hat{\mathbf{H}}_{a,b}} \right) + \tilde{n}(z) \log(1 - \hat{p}) + (n(n-1)/2 - \tilde{n}(z)) \log(1 - \hat{q}) \\ &= \frac{1}{2} \left(\sum_{a,b=1}^K \mathbf{O}_{a,b}(z) (\log \hat{\mathbf{S}}_{a,b} + \log \rho_n - \log(1 - \hat{\mathbf{H}}_{a,b})) \right) + C(z^*) + O_P(\rho_n) \\ &= \frac{\mu_n}{2} \left(\sum_{a,b=1}^K \frac{\mathbf{O}_{a,b}(z)}{\mu_n} \{\log \hat{\mathbf{S}}_{a,b} + O_P(\rho_n)\} \right) + \log \rho_n L + C(z^*) + O_P(\rho_n).\end{aligned}$$

where $C(z^*) = \tilde{n}(z^*) \log(1 - \hat{p}) + (n(n-1)/2 - \tilde{n}(z^*)) \log(1 - \hat{q})$. We let $F(\mathbf{O}(z)/\mu_n) = \sum_{a,b=1}^K \frac{\mathbf{O}_{a,b}(z)}{\mu_n} \log \frac{\hat{\mathbf{S}}_{a,b}}{1 - \hat{\mathbf{H}}_{a,b}}$ and $F(\mathbf{RSR}^\top(z)) = \sum_{a,b=1}^K (\mathbf{RSR}^\top(z))_{a,b} \log \frac{\hat{\mathbf{S}}_{a,b}}{1 - \hat{\mathbf{H}}_{a,b}}$, where $\mathbf{R}(z) = \mathbf{R}(z, z^*)$ and $\mathbf{RSR}^\top(z) = \mathbf{R}(z, z^*) \mathbf{S} \mathbf{R}^\top(z, z^*)^\top$. We denote $\tilde{\mathcal{C}} \subseteq \mathcal{C}_0 \cup \mathcal{C}_1$ as some subset of assignments, and we let \mathcal{V}_G denote the set of $z \in \tilde{\mathcal{C}}$ that maximizes $F(\mathbf{RSR}^\top(z))$. Obviously $F(\cdot)$ is Lipschitz, for $\epsilon_n \rightarrow 0$ slowly,

$$\begin{aligned}& \left| F(\mathbf{O}(z)/\mu_n) - F(\mathbf{RSR}^\top(z)) \right| \\ & \leq C \cdot \sum_{k,l} \left| \mathbf{O}_{k,l}(z)/\mu_n - (\mathbf{RSR}^\top(z))_{k,l} \right| \\ & = O_P(\epsilon_n).\end{aligned}$$

We choose some positive $\delta_n \rightarrow 0$ slowly enough such that $\delta_n/\epsilon_n \rightarrow \infty$. We take any $Z' \in \mathcal{V}_G$, then we define

$$J_{\delta_n} = \left\{ z \in [K]^n : F(\mathbf{RSR}^\top(z)) - F(\mathbf{RSR}^\top(Z')) < -\delta_n \right\}.$$

Then we have

$$\begin{aligned}\sum_{z \in J_{\delta_n}} e^{\log f(\mathbf{A}; z, \hat{p}, \hat{q})} &\leq f(\mathbf{A}; Z', \hat{p}, \hat{q}) K^n e^{O_P(\mu_n \epsilon_n) - \mu_n \delta_n / 2 + O_P(\rho_n)} \\ &= f(\mathbf{A}; Z', \hat{p}, \hat{q}) o_P(1).\end{aligned}$$

For $z \in \tilde{\mathcal{C}} \setminus \{J_{\delta_n} \cup \mathcal{V}_G\}$, $|F(\mathbf{RSR}^\top(z)) - F(\mathbf{RSR}^\top(Z'))| \rightarrow 0$ and $\|\mathbf{RSR}^\top(z) - \mathbf{RSR}^\top(Z')\|_\infty \rightarrow 0$. Treating $\mathbf{R}(z)$ as a vector, choosing z_\perp be such that $\mathbf{R}(z_\perp) := \min_{\mathbf{R}(z_0): z_0 \in \mathcal{V}_G} \|\mathbf{R}(z) - \mathbf{R}(z_0)\|_2$

for a given $z \in \tilde{\mathcal{C}} \setminus \{J_{\delta_n} \cup \mathcal{V}_G\}$. Due to the consistency of \hat{p}, \hat{q} , the function $F(\cdot)$ is a linear function with constant coefficients. We know that with probability $1 - o(1)$:

$$\left. \frac{\partial F((1 - \epsilon)\mathbf{RSR}^\top(z_\perp) + \epsilon\mathbf{RSR}^\top(z))}{\partial \epsilon} \right|_{\epsilon=0^+} < 0.$$

Given a matrix A , we denote the matrix maximum norm $\|A\|_\infty = \max_{jk} |A_{jk}|$. Letting $\bar{z} = \min_{\sigma(z)} |\sigma(z) - z_\perp|$, and $\mathbf{X}(z) = \mathbf{O}(z)/\mu_n - \mathbf{RSR}^\top(z)$, we have

$$\begin{aligned} & \mathbb{P} \left(\max_{z \notin \mathcal{S}(z_\perp)} \|\mathbf{X}(\bar{z}) - \mathbf{X}(z_\perp)\|_\infty > \epsilon |\bar{z} - z_\perp|/n \right) \\ & \leq \sum_{m=1}^n \mathbb{P} \left(\max_{z: z=\bar{z}, |\bar{z}-z_\perp|=m} \|\mathbf{X}(z) - \mathbf{X}(z_\perp)\|_\infty > \epsilon \frac{m}{n} \right) \\ & \leq \sum_{m=1}^n 2K^K n^m K^{m+2} \exp \left(-C \frac{m\mu_n}{n} \right) \rightarrow 0. \end{aligned}$$

where $\mathcal{S}(z) = \{\sigma(z) | \sigma \in S_K\}$. Since $\mathbf{RSR}^\top(\bar{z}) - \mathbf{RSR}^\top(z_\perp) = \Omega(|\bar{z} - z_\perp|)$, we have that

$$\frac{\mathbf{O}(\bar{z})}{\mu_n} - \frac{\mathbf{O}(z_\perp)}{\mu_n} = (1 + o_P(1)) \left(\mathbf{RSR}^\top(\bar{z}) - \mathbf{RSR}^\top(z_\perp) \right).$$

And thus we probability $1 - o(1)$ uniform on all z , we have

$$F(\mathbf{O}(\bar{z})/\mu_n) < F(\mathbf{O}(z_\perp)/\mu_n).$$

In turn, we have

$$\log f(\mathbf{A}; z, \hat{p}, \hat{q}) \leq \log f(\mathbf{A}; z_\perp, \hat{p}, \hat{q}) + O_P(\rho_n).$$

Since from Lemma A.1 in Wang and Bickel (2017) the high probability is uniform on all assignments, we have that, with probability $1 - o(1)$, for any $z \in \tilde{\mathcal{C}} \setminus \mathcal{V}_G$ we can find $z' \in \mathcal{V}_G$ such that $\log f(\mathbf{A}; z, p, q) = \log f(\mathbf{A}; z', p, q) + O_P(\rho_n)$, and therefore,

$$\sup_{z \in \tilde{\mathcal{C}}} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) = \sup_{z \in \mathcal{V}_G} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) + O_P(\rho_n).$$

Now we consider $F(\mathbf{RSR}^\top(z))$:

$$\begin{aligned} F(\mathbf{RSR}^\top(z)) &= \mathbb{E}(F(\mathbf{O}(z)/\mu_n)) \\ &= \frac{1}{\mu_n} \left(C_1(z^*) + \log \frac{\hat{\lambda}_1(1 - \hat{p})}{\hat{\lambda}_2(1 - \hat{q})} (n_2(z^*, z)q - n_1(z^*, z)p) \right) \\ &= \frac{1}{\mu_n} \left(C_1(z^*) + \log \frac{\hat{\lambda}_1(1 - \hat{p})}{\hat{\lambda}_2(1 - \hat{q})} (n_2(z^*, z) \vee n_1(z^*, z)(\lambda_2 - \lambda_1) + c_0(z))\rho_n \right), \end{aligned}$$

where $\hat{\lambda}_1 = \hat{p}/\rho_n$, $\hat{\lambda}_2 = \hat{q}/\rho_n$ and $C_1(z^*) = \log \frac{\hat{\lambda}_1}{1-\hat{p}} \tilde{n}(z^*)p + \log \frac{\hat{\lambda}_2}{1-\hat{q}} (n(n-1)/2 - \tilde{n}(z^*))q$, and $c_0(z) \leq \lambda_1 K c_K^2/2, \forall z \in \mathcal{C}_0 \cup \mathcal{C}_1$.

Thus when $z^* \in \mathcal{C}_0$ and $\tilde{\mathcal{C}} = \mathcal{C}_1$, it can be easily perceived that $\mathcal{V}_G \subseteq B_{z^*}(r_K)$ with high probability, and hence

$$\sup_{z \in \mathcal{C}_1} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) = \sup_{z \in B_{z^*}(r_K)} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) + O_P(\rho_n).$$

Moreover, when $z^* \in \tilde{\mathcal{C}}$, $\mathcal{V}_G \subseteq B_{z^*}(r^*)$ with high probability, where $r^* = \lambda_1 K c_K^2 / \{2(\lambda_1 - \lambda_2)\} = O(1)$. By Lemma 5.3 in Zhang and Zhou (2016), for any $z \in B_{z^*}(r^*)$, if $z \neq z^*$, then $d(z^*, z) = \Omega(n)$. Therefore, $B_{z^*}(r^*) = \{z^*\}$. In other words, we have

$$\sup_{z \in \tilde{\mathcal{C}}} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) = \log f(\mathbf{A}; z^*, \hat{p}, \hat{q}) + O_P(\rho_n). \quad (\text{D.1})$$

More concretely, if we take $\tilde{\mathcal{C}} = \mathcal{C}_0 \cup \mathcal{C}_1$, we have

$$\sup_{z \in \mathcal{C}_0 \cup \mathcal{C}_1} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) = \log f(\mathbf{A}; z^*, \hat{p}, \hat{q}) + O_P(\rho_n),$$

and if $z^* \in \mathcal{C}_1$, we have

$$\sup_{z \in \mathcal{C}_1} \log f(\mathbf{A}; z, \hat{p}, \hat{q}) = \log f(\mathbf{A}; z^*, \hat{p}, \hat{q}) + O_P(\rho_n).$$

D.3 Consistency of Probability Estimation

Recall the estimators \hat{p} and \hat{q} are defined in (2.9), and that $\hat{\lambda}_1 = \hat{p}/\rho_n$ and $\hat{\lambda}_2 = \hat{q}/\rho_n$. The following lemma shows that $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are consistent.

Lemma D.1. Under the same condition of Theorem 5.2, we have

$$|\hat{\lambda}_i - \lambda_i| = O(1/\sqrt{n^2 \rho_n}), \quad i = 1, 2$$

Proof. From Lemma 1 and Theorem 2 in Bickel et al. (2013), we know that $|\log(\hat{p}/(1-\hat{p})) - \log(p/(1-p))| = O(1/\sqrt{n^2 \rho_n})$ and $|\log(\hat{q}/(1-\hat{q})) - \log(q/(1-q))| = O(1/\sqrt{n^2 \rho_n})$. We let ν_1 and ν_2 denote the logit of p and q . Then since (ν_1, ν_2) is a one-to-one function of (p, q) , we know the relationship between $(\hat{\nu}_1, \hat{\nu}_2)$ and (\hat{p}, \hat{q}) should be $\hat{\nu}_1 = \log \hat{p}/(1-\hat{p})$ and

$\hat{\nu}_2 = \log \hat{q}/(1 - \hat{q})$. Then we have

$$\begin{aligned}
\hat{\nu}_i - \nu_i &= \log \frac{\hat{\lambda}_i \rho_n}{1 - \hat{\lambda}_i \rho_n} - \log \frac{\lambda_i \rho_n}{1 - \lambda_i \rho_n} \\
&= \log \hat{\lambda}_i \rho_n - \log \lambda_i \rho_n + \log(1 - \lambda_i \rho_n) - \log(1 - \hat{\lambda}_i \rho_n) \\
&= \log \left(1 + \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i} \right) - \log \left(1 + \frac{(\lambda_i - \hat{\lambda}_i) \rho_n}{1 - \lambda_i \rho_n} \right) \\
&= (1 + o(1)) \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i} + (1 + o(1)) \frac{(\hat{\lambda}_i - \lambda_i) \rho_n}{1 - \lambda_i \rho_n} \\
&\asymp \hat{\lambda}_i - \lambda_i
\end{aligned}$$

and thus by previous results we have

$$|\hat{\lambda}_i - \lambda_i| = O(1/\sqrt{n^2 \rho_n}), \quad i = 1, 2$$

□

References

- Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015*, pages 670–688. IEEE Computer Soc., Los Alamitos, CA, 2015. doi: 10.1109/FOCS.2015.47. URL <https://doi.org/10.1109/FOCS.2015.47>.
- Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory*, 62(1):471–487, 2016. ISSN 0018-9448. doi: 10.1109/TIT.2015.2490670. URL <https://doi.org/10.1109/TIT.2015.2490670>.
- Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.*, 48(3):1452–1474, 2020. ISSN 0090-5364. doi: 10.1214/19-AOS1854. URL <https://doi.org/10.1214/19-AOS1854>.
- Naman Agarwal, Afonso S. Bandeira, Konstantinos Koiliaris, and Alexandra Kolla. Multisection in the stochastic block model using semidefinite programming. In *Compressed sensing and its applications*, Appl. Numer. Harmon. Anal., pages 125–162. Birkhäuser/Springer, Cham, 2017.
- Edo M Airolidi, Thiago B Costa, and Stanley H Chan. Stochastic block-model approximation of a graphon: Theory and consistent estimation. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 692–700. Curran Associates, Inc., 2013. URL <http://papers.nips.cc/paper/>

- 5047-stochastic-blockmodel-approximation-of-a-graphon-theory-and-consistent-estimation.pdf.
- Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999. doi: 10.1038/43601. URL <https://doi.org/10.1038/43601>.
- Afonso S. Bandeira. Random Laplacian matrices and convex relaxations. *Found. Comput. Math.*, 18(2):345–379, 2018. ISSN 1615-3375. doi: 10.1007/s10208-016-9341-9. URL <https://doi.org/10.1007/s10208-016-9341-9>.
- Albert-László Barabási and Zoltán N. Oltvai. Network biology: understanding the cell’s functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004. doi: 10.1038/nrg1272. URL <https://doi.org/10.1038/nrg1272>.
- Peter Bickel, David Choi, Xiangyu Chang, and Hai Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.*, 41(4):1922–1943, 2013. ISSN 0090-5364. doi: 10.1214/13-AOS1124. URL <https://doi.org/10.1214/13-AOS1124>.
- Peter J. Bickel and Purnamrita Sarkar. Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(1):253–273, 2016. ISSN 1369-7412. doi: 10.1111/rssb.12117. URL <https://doi.org/10.1111/rssb.12117>.
- Christine Brun, François Chevenet, David Martin, Jérôme Wojcik, Alain Guénoche, and Bernard Jacq. Functional classification of proteins for the prediction of cellular function from a protein-protein interaction network. *Genome Biology*, 5(1):R6, 2003. doi: 10.1186/gb-2003-5-1-r6. URL <https://doi.org/10.1186/gb-2003-5-1-r6>.
- Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures Algorithms*, 49(3):503–532, 2016. ISSN 1042-9832. doi: 10.1002/rsa.20633. URL <https://doi.org/10.1002/rsa.20633>.
- Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819, 2013. ISSN 0090-5364. doi: 10.1214/13-AOS1161. URL <https://doi.org/10.1214/13-AOS1161>.
- D. S. Choi, P. J. Wolfe, and E. M. Airoidi. Stochastic blockmodels with a growing number of classes. *Biometrika*, 99(2):273–284, 2012. ISSN 0006-3444. doi: 10.1093/biomet/asr053. URL <https://doi.org/10.1093/biomet/asr053>.
- Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. Simple: Statistical inference on membership profiles in large networks, 2019.
- Joshua Faskowitz, Xiaoran Yan, Xi-Nian Zuo, and Olaf Sporns. Weighted stochastic block models of the human connectome across the life span. *Scientific Reports*, 8(1):12997, 2018. doi: 10.1038/s41598-018-31202-1. URL <https://doi.org/10.1038/s41598-018-31202-1>.

- Michael Y Galperin, David M Kristensen, Kira S Makarova, Yuri I Wolf, and Eugene V Koonin. Microbial genome analysis: the COG approach. *Briefings in Bioinformatics*, 20(4):1063–1070, 09 2017. ISSN 1477-4054. doi: 10.1093/bib/bbx117. URL <https://doi.org/10.1093/bib/bbx117>.
- Chao Gao and John Lafferty. Testing network structure using relations between small subgraph probabilities, 2017.
- Chao Gao, Zongming Ma, Anderson Y. Zhang, and Harrison H. Zhou. Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.*, 18:Paper No. 60, 45, 2017. ISSN 1532-4435.
- Chao Gao, Zongming Ma, Anderson Y. Zhang, and Harrison H. Zhou. Community detection in degree-corrected block models. *Ann. Statist.*, 46(5):2153–2185, 2018. ISSN 0090-5364. doi: 10.1214/17-AOS1615. URL <https://doi.org/10.1214/17-AOS1615>.
- Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, and Ulrike von Luxburg. Two-sample tests for large random graphs using network statistics, 2017.
- Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, and Ulrike von Luxburg. Two-sample hypothesis testing for inhomogeneous random graphs. *Ann. Statist.*, 48(4): 2208–2229, 2020. ISSN 0090-5364. doi: 10.1214/19-AOS1884. URL <https://doi.org/10.1214/19-AOS1884>.
- Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming: extensions. *IEEE Trans. Inform. Theory*, 62(10):5918–5937, 2016. ISSN 0018-9448. doi: 10.1109/TIT.2016.2594812. URL <https://doi.org/10.1109/TIT.2016.2594812>.
- Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic block-models: first steps. *Social Networks*, 5(2):109–137, 1983. ISSN 0378-8733. doi: 10.1016/0378-8733(83)90021-7. URL [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7).
- Vishesh Karwa, Debdeep Pati, Sonja Petrović, Liam Solus, Nikita Alexeev, Mateja Raič, Dane Wilburne, Robert Williams, and Bowei Yan. Exact tests for stochastic block models, 2016.
- Eugene V. Koonin, Natalie D. Fedorova, John D. Jackson, Aviva R. Jacobs, Dmitri M. Krylov, Kira S. Makarova, Raja Mazumder, Sergei L. Mekhedov, Anastasia N. Nikolskaya, B. Sridhar Rao, Igor B. Rogozin, Sergei Smirnov, Alexander V. Sorokin, Alexander V. Sverdlov, Sona Vasudevan, Yuri I. Wolf, Jodie J. Yin, and Darren A. Natale. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biology*, 5(2):R7, 2004. doi: 10.1186/gb-2004-5-2-r7. URL <https://doi.org/10.1186/gb-2004-5-2-r7>.
- Jing Lei. A goodness-of-fit test for stochastic block models. *Ann. Statist.*, 44(1):401–424, 2016. ISSN 0090-5364. doi: 10.1214/15-AOS1370. URL <https://doi.org/10.1214/15-AOS1370>.

- Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing*, pages 694–703. ACM, New York, 2014.
- Elchanan Mossel, Joe Neeman, and Allan Sly. Stochastic block models and reconstruction, 2012.
- Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for the planted bisection model. *Electronic Journal of Probability*, 21(0), 2016. ISSN 1083-6489. doi: 10.1214/16-ejp4185. URL <http://dx.doi.org/10.1214/16-EJP4185>.
- Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018. ISSN 0209-9683. doi: 10.1007/s00493-016-3238-8. URL <https://doi.org/10.1007/s00493-016-3238-8>.
- M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003. ISSN 0036-1445. doi: 10.1137/S003614450342480. URL <https://doi.org/10.1137/S003614450342480>.
- Matey Neykov, Junwei Lu, and Han Liu. Combinatorial inference for graphical models. *Ann. Statist.*, 47(2):795–827, 2019. ISSN 0090-5364. doi: 10.1214/17-AOS1650. URL <https://doi.org/10.1214/17-AOS1650>.
- Debnath Pal and David Eisenberg. Inference of protein function from protein structure. *Structure*, 13(1):121–130, 2005.
- Valentina Shumovskaia and Maxim Panov. Towards hypothesis testing for random graphs with community structure. In *ITIS Interdisciplinary School-Conference 2018*, pages 135–147, 2018.
- Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, Alexander Roth, Alberto Santos, Kalliopi P Tsafou, Michael Kuhn, Peer Bork, Lars J Jensen, and Christian von Mering. String v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(Database issue):D447–D452, 01 2015. doi: 10.1093/nar/gku1003. URL <https://pubmed.ncbi.nlm.nih.gov/25352553>.
- Timothée Tabouy, Pierre Barbillon, and Julien Chiquet. Variational Inference for Stochastic Block Models From Sampled Data. *J. Amer. Statist. Assoc.*, 115(529):455–466, 2020. ISSN 0162-1459. doi: 10.1080/01621459.2018.1562934. URL <https://doi.org/10.1080/01621459.2018.1562934>.
- Minh Tang, Avanti Athreya, Daniel L. Sussman, Vince Lyzinski, Youngser Park, and Carey E. Priebe. A semiparametric two-sample hypothesis testing problem for random graphs. *J. Comput. Graph. Statist.*, 26(2):344–354, 2017a. ISSN 1061-8600. doi: 10.1080/10618600.2016.1193505. URL <https://doi.org/10.1080/10618600.2016.1193505>.
- Minh Tang, Avanti Athreya, Daniel L. Sussman, Vince Lyzinski, and Carey E. Priebe. A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, 23

(3):1599–1630, 2017b. ISSN 1350-7265. doi: 10.3150/15-BEJ789. URL <https://doi.org/10.3150/15-BEJ789>.

Hafeez Ur Rehman, Nouman Azam, JingTao Yao, and Alfredo Benso. A three-way approach for protein function classification. *PloS one*, 12(2):e0171702–e0171702, 02 2017. doi: 10.1371/journal.pone.0171702. URL <https://pubmed.ncbi.nlm.nih.gov/28234929>.

Y. X. Rachel Wang and Peter J. Bickel. Likelihood-based model selection for stochastic block models. *Ann. Statist.*, 45(2):500–528, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1457. URL <https://doi.org/10.1214/16-AOS1457>.

Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

Yuhong Yang and Andrew Barron. Information-theoretic determination of minimax rates of convergence. *Ann. Statist.*, 27(5):1564–1599, 1999. ISSN 0090-5364. doi: 10.1214/aos/1017939142. URL <https://doi.org/10.1214/aos/1017939142>.

Se-Young Yun and Alexandre Proutiere. Accurate community detection in the stochastic block model via spectral algorithms, 2014.

Anderson Y. Zhang and Harrison H. Zhou. Minimax rates of community detection in stochastic block models. *Ann. Statist.*, 44(5):2252–2280, 2016. ISSN 0090-5364. doi: 10.1214/15-AOS1428. URL <https://doi.org/10.1214/15-AOS1428>.