A 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

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

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.

A.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 $\widetilde{z}$. However, for many community properties of interest, e.g., the community size test in Example 1.3, we cannot find such $\mathcal{N}$ and $\widetilde{z}$. In Example 1.3, we are interested in testing the community size and thus there is no representative nodes. See Figure 6 for illustration.

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

Definition A.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) $\tau \circ \sigma(z) := (\sigma(z(\tau(1))), \ldots, \sigma(z(\tau(n)))) = z'$ 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 A.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 A.1. See Figure 6(a) for an example of choosing σ and τ. On the other hand, the community sizes properties

$$\mathcal{C}_0 = \{z \in [K]^n : \text{all community sizes } = n/K\} \text{ and } \mathcal{C}_1 = \mathcal{C}_0^c, \tag{A.2}$$

are also symmetric by Definition A.1 but not by Definition 2.2. See Figure 6(b) for illustration. In fact, the following proposition shows that Definition 2.2 is a special case of Definition A.1.

Proposition A.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 (A.2) are generalized symmetric property pairs.

We defer the proof of the proposition to Appendix B.2. In Figure 6, we show how to choose concrete permutation transforms σ and τ for Examples 1.1 and 1.3.

A.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 A.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 7. From Figure 7(a), we can see that $B_{z_0} = B_{z_0}(d(\mathcal{C}_0, \mathcal{C}_1))$. Therefore, Definition A.2 is a generalization of Definition 2.5. For the uneven case,
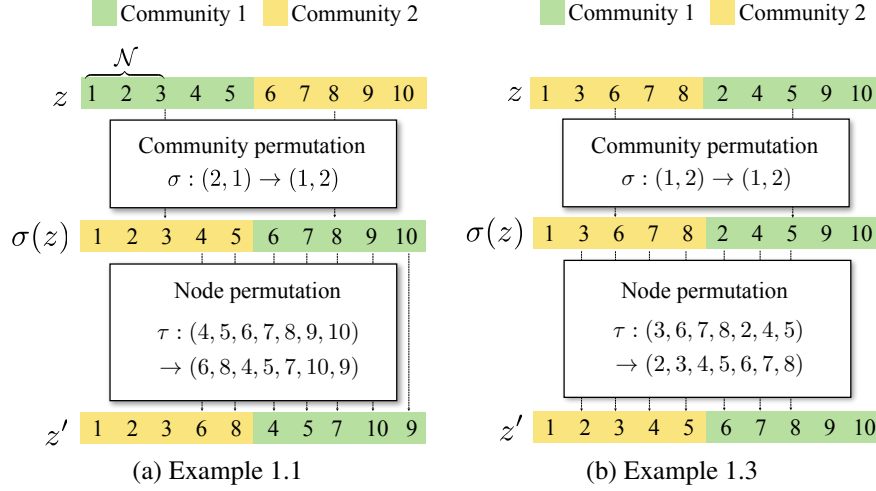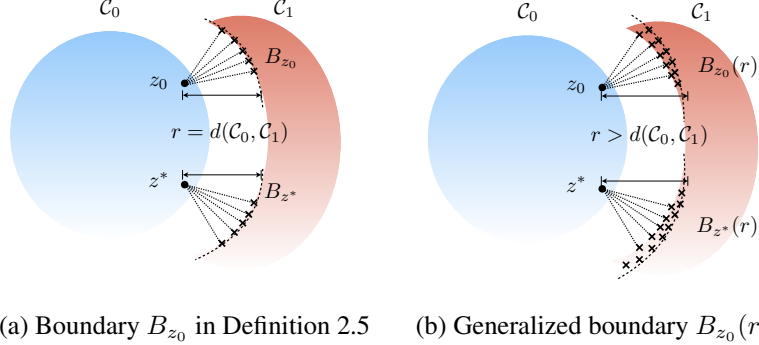
(a) Example 1.1   (b) Example 1.3

Figure 6: Permutation of null assignments in Example 1.1 and Example 1.3

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 $B_{z^*}(r)$.

(a) Boundary B_{z_0} in Definition 2.5   (b) Generalized boundary $B_{z_0}(r)$

Figure 7: The boundary B_{z_0} defined previously for even cases is in essence a ball centered at z_0 with radius $r = d(\mathcal{C}_0, \mathcal{C}_1)$.

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

$$W_n = \sup_{z \in B_{z_0}(r)} \sum_{1 \leq i < j \leq n} \big(\widehat{\mathbf{A}}_{ij} - \mathbb{E}_{\widehat{p}, \widehat{q}}(\widehat{\mathbf{A}}_{ij})\big)\big(\mathbb{1}[(i,j) \in \mathcal{E}_2(z_0, z)] - \mathbb{1}[(i,j) \in \mathcal{E}_1(z_0, z)]\big)e_{ij}, \quad \text{(A.3)}$$

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

Theorem A.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 (A.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

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

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

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

We defer the proof of theorem to Appendix C.2. The scaling assumptions in Theorem A.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 (A.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 $r = d(\mathcal{C}_0, \mathcal{C}_1) + C\widehat{p}K/(\widehat{p} - \widehat{q})$ 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 A.2 are true for all these examples.

Proposition A.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 8(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 $m+1, \ldots, m+m'$ 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 $m + 1, \ldots, m + m'$ collectively with another m' nodes from a different community for any $z_0 \in \mathcal{C}_0$. See Figure 8(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 8(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 B.3. The construction of $B_{z_0}(r_K)$ is visualized in Figure 8. We also summarize the results in Table 1.
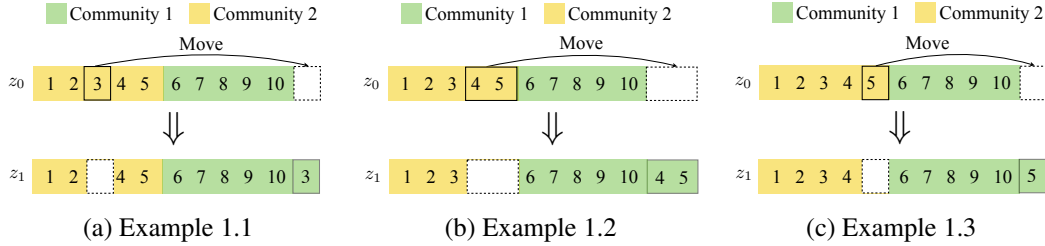

(a) Example 1.1 (b) Example 1.2 (c) Example 1.3

Figure 8: Construction of B_{z_0} in Proposition A.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 A.2.

Corollary A.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 A.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.,

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

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

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

A.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

	$d(z_0, \mathcal{C}_1)$	$|B_{z_0}(r_K)|$	$N(B_{z_0}(r_K), \sqrt{d(z_0, \mathcal{C}_1)})$
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 1: Important values for general cases of Examples 1.1-1.3.

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 A.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

$$\limsup_{n \to \infty} \frac{d(z_0, \mathcal{C}_1) I(p, q)}{\log N(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)})} < 1, \tag{A.4}$$

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

Remark A.1. If we choose $r = d(z_0, \mathcal{C}_1)$, as $B_{z_0} = B_{z_0}(d(z_0, \mathcal{C}_1))$, (A.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 A.6. Suppose $0 < q < p \leq 1 - \delta$ for some constant $\delta > 0$ and $\lim_{n \to \infty} d(\mathcal{C}_0, \mathcal{C}_1)p = \infty$. 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) $\lim_{n \to \infty} d(\mathcal{C}_0, \mathcal{C}_1)I(p, q) = \infty$, but there exists a $z_0 \in \mathcal{C}_0$ and some $r = d(z_0, \mathcal{C}_1) + O(1)$ such that $\limsup_{n \to \infty} d(z_0, \mathcal{C}_1)I(p, q)/\log N(B_{z_0}(r), 0) < 1$,

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

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

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

Proposition A.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 B.4. The results is also summarized in Table 1.

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 A.5 and Proposition A.7, we have the following lower bound of same community test in Example 1.1.

Corollary A.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

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

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

Applying Theorem A.6 and Proposition A.7, we have the following lower bound of same community test for groups in Example 1.2.

Corollary A.9. For $\mathcal{C}_0$ and $\mathcal{C}_1$ defined in (2.2), if $np \to \infty$, $0 < q < p < 1 - \delta$ for some $\delta > 0$ and

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

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

For Example 1.3, applying Theorem A.5 and Proposition A.7, we have the following result.

Corollary A.10. For $\mathcal{C}_0$ and $\mathcal{C}_1$ defined in (A.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

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

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

B 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.

B.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 $z_i, z_j \in \mathcal{P}\big(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}\big)$, 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 $z \in \mathcal{P}\big(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}\big)$ corresponds to a different swapped node among the first m nodes, and we have $N\big(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}\big) \leq m$. 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 $\mathcal{S} = \{s_1, s_2, ..., s_m\}$ 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, $k = 1, ..., m$, 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 $N\big(B_{z_0}, \sqrt{d(z_0, \mathcal{C}_1)}\big) = m$.
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 $\forall z_i, z_j \in B_{z_0}, |\mathcal{E}_{1,2}(z_0, z_i) \cap \mathcal{E}_{1,2}(z_0, z_j)| \geq m \wedge m'(n/K - m \wedge m')$.

B.2 PROOF OF PROPOSITION A.1

To prove that Definition 2.2 is a special case of Definition A.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 A.1 based on $\mathcal{N}$ and $\widetilde{z}$. Here we use Figure 6 to illustrate the construction. Given any $z, z' \in \mathcal{C}_0$, we first construct σ. Since $z_\mathcal{N} \simeq z'_\mathcal{N} \simeq \widetilde{z}_\mathcal{N}$, 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 6, 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 6. Using σ and τ constructed above, we can check that $\tau \circ \sigma(z) = z'$. We now check the last condition in Definition A.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 A.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 (A.2), we can see from Figure 6(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 A.1, the property pair in (A.2) is a generalized symmetric property pair.

B.3 PROOF OF PROPOSITION A.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 8 (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 $m + 1, \ldots, m + m'$ collectively to a different community (see Figure 8 (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 $m + 1, \ldots, m + m'$. 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 $m + 1, \ldots, m + m'$ 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 $|B_{z_0}(r_K)| = n(K - 1) = O(n)$.

B.4 PROOF OF PROPOSITION A.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 $z_i, z_j \in \mathcal{P}\big(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)}\big)$, 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 $N\big(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)}\big) \leq n$. 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 $N\big(B_{z_0}(r), \sqrt{d(z_0, \mathcal{C}_1)}\big) = n$.

C 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 E.

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

C.1 PROOF OF PROPOSITION 2.3

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

Lemma C.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 $z_1, z_2, \ldots, z_{|B_z(r)|}$. Define a $|B_z(r)|$-dimensional vector $\boldsymbol{L}_z$ as

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

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

We defer the proof of Lemma C.1 to Section E.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

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

$$= g(p, q) \sup_{z_k \in B_{z^*}(r)} \left\{ \sum_{i<j} \left\{ \left(\mathbf{A}_{ij} - \mathbb{E}(\mathbf{A}_{ij}) \right)\left(\mathbb{1}[(i,j) \in \mathcal{E}_2(z^*, z_k)] - \mathbb{1}[(i,j) \in \mathcal{E}_1(z^*, z_k)] \right) \right\} \right\}$$

$$+ g(p, q)\mu_0 + \delta_n$$

$$= g(p, q)\sigma_0 \sup_{k \in [|B_{z^*}(r)|]} \left\{ \frac{1}{\sigma_0} \sum_{i<j} (\boldsymbol{X}_{ij})_k \right\} + g(p, q)\mu_0 + \delta_n.$$

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

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

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

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

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

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

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

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

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

And thus we have

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

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

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

where $\widetilde{Z}' \sim N(0, \boldsymbol{\Sigma}_{z_0}/\sigma_0^2)$, and $\boldsymbol{\Sigma}_{z_0} = \text{Cov}(\boldsymbol{L}_{z_0})$. By Lemma C.1 we know that $\boldsymbol{\Sigma}_{z^*}$ and $\boldsymbol{\Sigma}_{z_0}$ are equal up to permutation. Therefore, the claim follows. We may also note 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 $\widehat{L}_0 := \sup_{z_k \in B_{z^*}(r)} \Big\{ g(\widehat{p}, \widehat{q}) \big(\sum_{(i,j) \in \mathcal{E}_2(z^*, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z_k)} \mathbf{A}_{ij} \big) \Big\}$ with plugged-in estimators $\widehat{p}, \widehat{q}$.

C.2 PROOF OF THEOREM 3.2

In fact, Proposition A.1 shows that the symmetric community property pairs defined in Section 2 are general symmetric property pairs under the general framework, and Theorem A.2 is a generalization of Theorem 3.2 under uneven cluster sizes. Thus we can just prove the more general Theorem A.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 C.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 A.2, when $z^* \in \mathcal{C}_0$

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

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

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

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 $B_{z_0}(r_K)$ 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 C.2 to Appendix E.2.

Now we are ready to present the proof of Theorem A.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

$$
q_\alpha = g(\widehat{p}, \widehat{q}) C_W(\alpha) + g(\widehat{p}, \widehat{q}) \widehat{\mu}_0, \tag{C.3}
$$

then it can be seen that the two events $\{\widehat{\mathrm{LRT}} \geq q_\alpha\}$ and $\{p_W \leq \alpha\}$ are equivalent. Therefore, it suffices to show that $\lim_{n\to\infty} \sup_{z^* \in \mathcal{C}_0} \mathbb{P}(\widehat{\mathrm{LRT}} \leq q_\alpha) = \alpha$. 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 $\widehat{p}, \widehat{q}$, and the third part is to illustrate the multiplier bootstrap as a valid approximation of the LRT quantile.

C.2.1 DERIVATION OF THE LEADING TERM FOR LRT

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

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

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 $\widehat{\mathrm{LRT}}$ 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

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

where $\delta_n = \sup_{z_k \in B_{z^*}(r)} \Big\{ \log \big((1-\widehat{q})/(1-\widehat{p})\big) \Big(n_1(z^*, z_k) - n_2(z^*, z_k) \Big) \Big\} + O_P(\rho_n) = O_P(\rho_n)$, and $\widehat{L}_0 = g(\widehat{p}, \widehat{q}) \sup_{z_k \in B_{z^*}(r)} \big(\sum_{(i,j) \in \mathcal{E}_2(z^*, z_k)} \mathbf{A}_{ij} - \sum_{(i,j) \in \mathcal{E}_1(z^*, z_k)} \mathbf{A}_{ij} \big)$. From Proposition 2.3 we have that $\lim_{n\to\infty} \sup_{t \in \mathbb{R}} |\mathbb{P}(\widehat{L}_0 < t) - \mathbb{P}(\widehat{L}_0(z_0) < t)| = 0$ for any $z_0 \in \mathcal{C}_0$. Therefore, it suffices for us to prove that $\mathbb{P}(\widehat{\mathrm{LRT}} \geq q_\alpha) = \alpha + o(1)$ 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.

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

From previous section we know that

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

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

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

and denote

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

where $(\widehat{\boldsymbol{X}}_{ij})_k = (\widehat{\boldsymbol{A}}_{ij} - \mathbb{E}_{\widehat{p},\widehat{q}}(\widehat{\boldsymbol{A}}_{ij}))(\mathbb{1}[(i,j) \in \mathcal{E}_2(z_0, z_k)] - \mathbb{1}[(i,j) \in \mathcal{E}_1(z_0, z_k)])$ and the adjacency matrix $\widehat{\boldsymbol{A}}$ 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 $\boldsymbol{X}_{ij}$ and $\widehat{\boldsymbol{X}}_{ij}$ 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

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

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

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(\widetilde{W}_n \leqslant t | \widehat{\boldsymbol{X}}_{ij}) - \mathbb{P}(\Xi_0' \leqslant t) \right| = o_P(1).$$

We let $\boldsymbol{\Sigma}^{\Xi_0}$ and $\boldsymbol{\Sigma}^{\Xi_0'}$ 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:

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

Accordingly,

$$
\begin{aligned}
\boldsymbol{\Sigma}^{\Xi_0'}_{k,l} &= \frac{1}{\widehat{\sigma}_0^2} \mathrm{Cov}(\sum_{i<j} (\widehat{\boldsymbol{X}}_{ij})_k, \sum_{i<j} (\widehat{\boldsymbol{X}}_{ij})_l) \\
&= \frac{|\mathcal{E}_2(z_0, z_k) \cap \mathcal{E}_2(z_0, z_l)| \widehat{q}(1-\widehat{q}) + |\mathcal{E}_1(z_0, z_k) \cap \mathcal{E}_1(z_0, z_l)| \widehat{p}(1-\widehat{p})}{d(\mathcal{C}_0, \mathcal{C}_1)(\widehat{p}(1-\widehat{p}) + \widehat{q}(1-\widehat{q}))} \\
&= \frac{|\mathcal{E}_2(z_0, z_k) \cap \mathcal{E}_2(z_0, z_l)| q(1-q) + |\mathcal{E}_1(z_0, z_k) \cap \mathcal{E}_1(z_0, z_l)| p(1-p) + O_P(d(\mathcal{C}_0, \mathcal{C}_1)\sqrt{\rho_n}/n)}{d(\mathcal{C}_0, \mathcal{C}_1)(p(1-p) + q(1-q)) + O_P(d(\mathcal{C}_0, \mathcal{C}_1)\sqrt{\rho_n}/n)}.
\end{aligned}
$$

Then we have

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

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

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

and thus

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

and in turn we have

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

C.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 $\widehat{\boldsymbol{X}}_{ij}$, 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 $\widehat{\mathrm{LRT}}$. Give a constant

$t > \sqrt{2c_0}$, we have

$$\mathbb{P}(\widetilde{W}_n \geqslant t\sqrt{\log n}|\widehat{\mathbf{X}}_{ij}) = \mathbb{P}(\Xi_0' \geqslant t\sqrt{\log n}) + o_P(1)$$

$$\leq \sum_{k \in [|B_{z_0}(r)|]} \mathbb{P}\left(\left\{\frac{1}{\widehat{\sigma}_0}\sum_{i<j}\{\widehat{\xi}_{ij}\}_k\right\} \geqslant t\sqrt{\log n}\right) + o_P(1)$$

$$\lesssim |B_{z_0}(r)|e^{-\frac{t^2}{2}\log n} + o_P(1) = O_P\left(n^{c_0 - t^2/2}\right) + o_P(1) = o_P(1).$$

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

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

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,

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

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

$$|\mathbb{P}(\widehat{\mathrm{LRT}} \geq q_\alpha) - \alpha| \leq |\mathbb{P}(T_0 \geq C_{\widetilde{W}_n}(\alpha) + \Delta_n) - \mathbb{P}(\widetilde{W}_n \geq C_{\widetilde{W}_n}(\alpha) + \Delta_n)|$$
$$+ |\mathbb{P}(\widetilde{W}_n \geq C_{\widetilde{W}_n}(\alpha) + \Delta_n) - \mathbb{P}(\widetilde{W}_n \geq C_{\widetilde{W}_n}(\alpha))| + o(1)$$
$$\leq \mathbb{P}(|\widetilde{W}_n - C_{\widetilde{W}_n}(\alpha)| \leq \Delta_n) + o_P(1).$$

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 $\sup_k |\mathbb{E}(\sigma_k^2) - 1| \leq |\widehat{\sigma}_0^2/\widehat{\sigma}_0^2 - 1| + o_P(1) = o_P(1)$. 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

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

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

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

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

As for the Type I error, from the preceding proof we see that $\mathbb{P}(\mathrm{LRT} \geq q_\alpha) = \alpha + o_P(1)$, 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

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

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

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

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

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

D 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 A.5 and Theorem A.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 A.5 is actually based on the proof of Theorem A.6 under a stronger regime. Therefore, we will prove the two theorems together: we will first prove Theorem A.6 under more general conditions, and then we will apply the proof of Theorem A.6 to the proof of Theorem A.5 under stronger conditions.

D.1 PROOF OF THEOREM A.5 AND THEOREM A.6

The proof proceeds in the following order: we first prove the results under the two conditions of Theorem A.6, namely the proof of Theorem A.6 (1) and the proof of Theorem A.6 (2), then we provide the proof of Theorem A.5.

D.1.1 PROOF OF THEOREM A.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 $\{\mathbf{A}_{ij}, (i, j) \in \mathcal{E}_1(z_0, z_1) \bigcup \mathcal{E}_2(z_0, z_1)\}$. 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 $\inf_\psi \{ \sup_{z \in \mathcal{C}_0} \mathbb{P}_z(\psi = 1) + \sup_{z \in \mathcal{C}_1} \mathbb{P}_z(\psi = 0) \}$, the optimal method $\widetilde{\psi}$ 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:

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

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

and correspondingly,

$$\widetilde{\psi}(\mathbf{A}) = \left\{ \begin{array}{ll} 0, & \text{if } L(z_0, \mathbf{A}) > L(z_1, \mathbf{A}); \\ 1, & \text{if } L(z_0, \mathbf{A}) \leq L(z_1, \mathbf{A}). \end{array} \right.$$

Then $\mathbb{P}_{z_0}(\widetilde{\psi} = 1) = \mathbb{P}_{z_0}(L(z_0, \mathbf{A}) \leq L(z_1, \mathbf{A}))$ and $\mathbb{P}_{z_1}(\widetilde{\psi} = 0) = \mathbb{P}_{z_1}(L(z_0, \mathbf{A}) > L(z_1, \mathbf{A}))$. 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 $\{\mathbf{A}_{ij}, (i, j) \in \mathcal{E}_2^L(z_0, z_1) \backslash \mathcal{E}_2(z_0, z_1)\}$ 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:

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

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

Thus we can obtain lower bound on the minimax rate:

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

where $\{X_u\} \overset{\text{i.i.d}}{\sim} \text{Ber}(q), \{Y^u\} \overset{\text{i.i.d}}{\sim} \text{Ber}(p)$, 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,

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

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

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

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

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

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

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

D.1.2 PROOF OF THEOREM A.6 (2)

When $d(z_0, \mathcal{C}_1)I(p,q) \to \infty$, if there exists a $z_0 \in \mathcal{C}_0$ and some $r = d(z_0, \mathcal{C}_1) + O(1)$ such that $\limsup_{n\to\infty} d(\mathcal{C}_0, \mathcal{C}_1)I(p,q)/\log N(B_{z_0}(r),0) < 1$, 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:

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

$$= \min_\psi\left\{\sum_A\left(\mathbb{P}_{z_0}(\psi=1|\mathbf{A}=A)\mathbb{P}_{z_0}(\mathbf{A}=A) + \sup_{z\in\mathcal{P}(0)}\mathbb{P}_z(\psi=0|\mathbf{A}=A)\mathbb{P}_z(\mathbf{A}=A)\right)\right\}$$

$$= \min_\psi\left\{\sum_A\left(\mathbb{1}(\psi(A)=1)\mathbb{P}_{z_0}(\mathbf{A}=A) + \mathbb{1}(\psi(A)=0)\sup_{z\in\mathcal{P}(0)}\mathbb{P}_z(\mathbf{A}=A)\right)\right\},$$

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

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

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

$$r(\mathcal{C}_0,\mathcal{C}_1) \geq \mathbb{P}_{z_0}(\widetilde{\psi}=1) + \sup_{z\in\mathcal{P}(0)}\mathbb{P}_z(\widetilde{\psi}=0)$$

$$= \mathbb{P}_{z_0}\left(L(z_0,\mathbf{A}=A) < \sup_{z\in\mathcal{P}(0)} L(z,\mathbf{A}=A)\right)$$

$$+ \sup_{z\in\mathcal{P}(0)}\mathbb{P}_z\left(L(z_0,\mathbf{A}=A) \geq \sup_{z\in\mathcal{P}(0)} L(z,\mathbf{A}=A)\right)$$

$$= \mathbb{P}_{z_0}\left(\sup_{z\in\mathcal{P}(0)} \log L(z,\mathbf{A}=A) - \log L(z_0,\mathbf{A}=A) > 0\right)$$

$$+ \sup_{z\in\mathcal{P}(0)}\mathbb{P}_z\left(\sup_{z\in\mathcal{P}(0)} \log L(z,\mathbf{A}=A) - \log L(z_0,\mathbf{A}=A) \leq 0\right).$$

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

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

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

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

When $\limsup_{n\to\infty} d(z_0,\mathcal{C}_1)I(p,q)/\log|\mathcal{P}(0)| < 1$, 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

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

and thus

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

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

D.1.3 PROOF OF THEOREM A.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 $\widetilde{\mathcal{P}}$ 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:

$$r(\mathcal{C}_0, \mathcal{C}_1) \geq \mathbb{P}_{z_0}\left(\sup_{z \in \widetilde{\mathcal{P}}} \left(\sum_{u=1}^{n_1(z_0,z)} X_z^u - \sum_{u=1}^{n_1(z_0,z)} Y_z^u \right) > 0 \right)$$

$$= \mathbb{P}_{z_0}\left(\sup_{z \in \widetilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0,\mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0,\mathcal{C}_1)} Y_z^u \right) > \delta_n \right)$$

$$= \mathbb{P}_{z_0}\left(\sup_{z \in \widetilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0,\mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0,\mathcal{C}_1)} Y_z^u + d(z_0,\mathcal{C}_1)(p-q) \right) > d(z_0,\mathcal{C}_1)(p-q) + \delta_n \right)$$

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

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

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

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

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

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

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

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

and thus

$$\mathbb{P}_{z_0}\left(\sup_{z \in \widetilde{\mathcal{P}}} \left(\sum_{u=1}^{d(z_0,\mathcal{C}_1)} X_z^u - \sum_{u=1}^{d(z_0,\mathcal{C}_1)} Y_z^u + d(z_0,\mathcal{C}_1)(p-q) \right) > d(z_0,\mathcal{C}_1)(p-q) \right)$$

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

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

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

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.

E   PROOF OF TECHNICAL LEMMAS

Now we will provide proofs for the technical lemmas used for the proof of Theorem A.2.

E.1   PROOF OF LEMMA 2.2

It suffices for us to prove Lemma C.1, the more general version of Lemma 2.2. Due to the structure of $\boldsymbol{L}_z$, 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:

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

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

$$
|\mathcal{E}_1(z_0, z_1) \cap \mathcal{E}_1(z_0, z_1')| = \sum_{i<j, i,j\in[n]} \mathbb{1}\big(z_0(i) = z_0(j), z_1(i) \neq z_1(j), z_1'(i) \neq z_1'(j)\big)
$$

$$
= \sum_{i<j, i,j\in[n]} \mathbb{1}\big(\tau \circ \sigma(z_0)(i) = \tau \circ \sigma(z_0)(j), \tau \circ \sigma(z_1)(i) \neq \tau \circ \sigma(z_1')(j), \tau \circ \sigma(z_1')(i) \neq \tau \circ \sigma(z_1)(j)\big)
$$

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

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 $\boldsymbol{L}_{z_0}$ can be expressed as

$$\text{Cov}(\boldsymbol{L}_{z_0})_{kl} = g(p, q)^2 \Big(|\mathcal{E}_2(z_0, z_k) \cap \mathcal{E}_2(z_0, z_l)| q(1 - q) + |\mathcal{E}_1(z_0, z_k) \cap \mathcal{E}_1(z_0, z_l)| p(1 - p) \Big)$$

$$= g(p, q)^2 \Big(|\mathcal{E}_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z_k)) \cap \mathcal{E}_2(\tau \circ \sigma(z_0), \tau \circ \sigma(z_l))| q(1 - q)$$

$$+ |\mathcal{E}_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z_k)) \cap \mathcal{E}_1(\tau \circ \sigma(z_0), \tau \circ \sigma(z_l))| p(1 - p) \Big)$$

$$= \text{Cov} \Big(g(p, q) \big(\sum_{\mathcal{E}_2(\tau(z_0'), \tau(z_k'))} \mathbf{A}_{ij} - \sum_{\mathcal{E}_1(\tau(z_0'), \tau(z_k'))} \mathbf{A}_{ij} \big), g(p, q) \big(\sum_{\mathcal{E}_2(\tau(z_0'), \tau(z_l'))} \mathbf{A}_{ij} - \sum_{\mathcal{E}_1(\tau(z_0'), \tau(z_l'))} \mathbf{A}_{ij} \big) \Big)$$

$$= \text{Cov}(\boldsymbol{L}_{z_0'})_{\tau(k)\tau(l)}.$$

Hence we finish the proof.

E.2   PROOF OF LEMMA C.2

The proof mainly follows from Lemma 2.3 and Lemma 2.6 in Wang & 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 & 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 $a \neq b$. $\mathbf{H} = \rho_n \mathbf{S}$.

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

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

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

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

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

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

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

Then

$$\widetilde{n}(z) = \frac{\sum_{k=1}^{K}(n/K + a_k)^2 - n}{2} = \frac{n^2/K - n}{2} + \frac{\sum_{k=1}^{K} a_k^2}{2}$$

$$\leq \frac{n^2/K - n}{2} + K c_K^2 / 2.$$

29

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

$$\log f(\mathbf{A}; z, \widehat{p}, \widehat{q}) = \frac{1}{2} \Big(\sum_{a,b=1}^{K} \boldsymbol{O}_{a,b}(z) \log \frac{\widehat{\mathbf{H}}_{a,b}}{1 - \widehat{\mathbf{H}}_{a,b}} \Big) + \widetilde{n}(z) \log(1 - \widehat{p}) + \big(n(n-1)/2 - \widetilde{n}(z) \big) \log(1 - \widehat{q})$$

$$= \frac{1}{2} \Big(\sum_{a,b=1}^{K} \boldsymbol{O}_{a,b}(z) \big(\log \widehat{\mathbf{S}}_{a,b} + \log \rho_n - \log(1 - \widehat{\mathbf{H}}_{a,b}) \big) \Big) + C(z^*) + O_P(\rho_n)$$

$$= \frac{\mu_n}{2} \Big(\sum_{a,b=1}^{K} \frac{\boldsymbol{O}_{a,b}(z)}{\mu_n} \{ \log \widehat{\mathbf{S}}_{a,b} + O_P(\rho_n) \} \Big) + \log \rho_n L + C(z^*) + O_P(\rho_n).$$

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

$$\Big| F\big(\boldsymbol{O}(z)/\mu_n\big) - F(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z)) \Big|$$

$$\leq C \cdot \sum_{k,l} \Big| \boldsymbol{O}_{k,l}(z)/\mu_n - \big(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z) \big)_{k,l} \Big|$$

$$= O_P(\epsilon_n).$$

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

$$J_{\delta_n} = \Big\{ z \in [K]^n : F(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z)) - F(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(Z')) < -\delta_n \Big\}.$$

Then we have

$$\sum_{z \in J_{\delta_n}} e^{\log f(\mathbf{A}; z, \widehat{p}, \widehat{q})} \leq f(\mathbf{A}; Z', \widehat{p}, \widehat{q}) K^n e^{O_P(\mu_n \epsilon_n) - \mu_n \delta_n/2 + O_P(\rho_n)}$$

$$= f(\mathbf{A}; Z', \widehat{p}, \widehat{q}) o_P(1).$$

For $z \in \widetilde{\mathcal{C}} \backslash \{ J_{\delta_n} \cup \mathcal{V}_G \}$, $|F(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z)) - F(\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(Z'))| \to 0$ and $\|\boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z) - \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(Z')\|_\infty \to 0$. Treating $\boldsymbol{R}(z)$ as a vector, choosing $z_\perp$ be such that $\boldsymbol{R}(z_\perp) := \min_{\boldsymbol{R}(z_0): z_0 \in \mathcal{V}_G} \|\boldsymbol{R}(z) - \boldsymbol{R}(z_0)\|_2$ for a given $z \in \widetilde{\mathcal{C}} \backslash \{ J_{\delta_n} \cup \mathcal{V}_G \}$. Due to the consistency of $\widehat{p}, \widehat{q}$, the function $F(\cdot)$ is a linear function with constant coefficients. We know that with probability $1 - o(1)$:

$$\frac{\partial F \big((1 - \epsilon) \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z_\perp) + \epsilon \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z) \big)}{\partial \epsilon} \bigg|_{\epsilon = 0^+} < 0.$$

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 $\boldsymbol{X}(z) = \boldsymbol{O}(z)/\mu_n - \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\top(z)$, we have

$$\mathbb{P} \left(\max_{z \notin \mathcal{S}(z_\perp)} \|\boldsymbol{X}(\bar{z}) - \boldsymbol{X}(z_\perp)\|_\infty > \epsilon |\bar{z} - z_\perp|/n \right)$$

$$\leq \sum_{m=1}^{n} \mathbb{P} \left(\max_{z: z = \bar{z}, |\bar{z} - z_\perp| = m} \|\boldsymbol{X}(z) - \boldsymbol{X}(z_\perp)\|_\infty > \epsilon \frac{m}{n} \right)$$

$$\leq \sum_{m=1}^{n} 2 K^K n^m K^{m+2} \exp \left(-C \frac{m \mu_n}{n} \right) \to 0.$$

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

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

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

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

In turn, we have

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

Since from Lemma A.1 in Wang & Bickel (2017) the high probability is uniform on all assignments, we have that, with probability $1 - o(1)$, for any $z \in \widetilde{\mathcal{C}} \backslash \mathcal{V}_G$ 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,

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

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

$$F(\boldsymbol{RSR}^\top(z)) = \mathbb{E}(F(\boldsymbol{O}(z)/\mu_n))$$

$$= \frac{1}{\mu_n} \left(C_1(z^*) + \log \frac{\widehat{\lambda}_1(1 - \widehat{p})}{\widehat{\lambda}_2(1 - \widehat{q})} \left(n_2(z^*, z)q - n_1(z^*, z)p \right) \right)$$

$$= \frac{1}{\mu_n} \left(C_1(z^*) + \log \frac{\widehat{\lambda}_1(1 - \widehat{p})}{\widehat{\lambda}_2(1 - \widehat{q})} \left(n_2(z^*, z) \vee n_1(z^*, z)(\lambda_2 - \lambda_1) + c_0(z) \right) \rho_n \right),$$

where $\widehat{\lambda}_1 = \widehat{p}/\rho_n, \widehat{\lambda}_2 = \widehat{q}/\rho_n$ and $C_1(z^*) = \log \frac{\widehat{\lambda}_1}{1 - \widehat{p}} \widetilde{n}(z^*)p + \log \frac{\widehat{\lambda}_2}{1 - \widehat{q}}(n(n-1)/2 - \widetilde{n}(z^*))q$, 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 $\widetilde{\mathcal{C}} = \mathcal{C}_1$, it can be easily perceived that $\mathcal{V}_G \subseteq B_{z^*}(r_K)$ with high probability, and hence

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

Moreover, when $z^* \in \widetilde{\mathcal{C}}$, $\mathcal{V}_G \subseteq B_{z^*}(r^*)$ with high probability, where $r^* = \lambda_1 K c_K^2/\{2(\lambda_1 - \lambda_2)\} = O(1)$. By Lemma 5.3 in Zhang & 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

$$\sup_{z \in \widetilde{\mathcal{C}}} \log f(\mathbf{A}; z, \widehat{p}, \widehat{q}) = \log f(\mathbf{A}; z^*, \widehat{p}, \widehat{q}) + O_P(\rho_n). \tag{E.1}$$

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

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

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

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

E.3 CONSISTENCY OF PROBABILITY ESTIMATION

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

Lemma E.1. Under the same condition of Theorem A.2, we have

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

Proof. From Lemma 1 and Theorem 2 in Bickel et al. (2013), we know that $|\log\left(\widehat{p}/(1-\widehat{p})\right) - \log\left(p/(1-p)\right)| = O(1/\sqrt{n^2\rho_n})$ and $|\log\left(\widehat{q}/(1-\widehat{q})\right) - \log\left(q/(1-q)\right)| = O(1/\sqrt{n^2\rho_n})$. 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 $(\widehat{\nu}_1, \widehat{\nu}_2)$ and $(\widehat{p}, \widehat{q})$ should be $\widehat{\nu}_1 = \log\widehat{p}/(1-\widehat{p})$ and $\widehat{\nu}_2 = \log\widehat{q}/(1-\widehat{q})$. Then we have

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

and thus by previous results we have

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

$\square$