

# Statistical inference for core-periphery structures

Eric Yanchenko\*, Srijan Sengupta† and Diganta Mukherjee‡

## Abstract

Core-periphery (CP) structure is an important meso-scale network property where nodes group into a small, densely interconnected core and a sparse periphery whose members primarily connect to the core rather than to each other. While this structure has been observed in numerous real-world networks, there has been minimal statistical formalization of it. In this work, we develop a statistical framework for CP structures by introducing a model-agnostic and generalizable population parameter which quantifies the strength of a CP structure at the level of the data-generating mechanism. We study this parameter under four canonical random graph models and establish theoretical guarantees for label recovery, including exact label recovery. Next, we construct intersection tests for validating the presence and strength of a CP structure under multiple null models, and prove theoretical guarantees for type I error and power. These tests provide a formal distinction between exogenous (or induced) and endogenous (or intrinsic) CP structure in heterogeneous networks, enabling a level of structural resolution that goes beyond merely detecting the presence of CP structure. The proposed methods show excellent performance on synthetic data, and our applications demonstrate that statistically significant CP structure is somewhat rare in real-world networks.

*Keywords:* Meso-scale structures, Core Periphery structures, Networks, Random graph models

---

\*Global Connectivity Program, Akita International University, Akita, 010-1211, Japan.  
eyanchenko@aiu.ac.jp

†Department of Statistics, North Carolina State University, Raleigh, NC, 27606. ssengup2@ncsu.edu

‡Sampling and Official Statistics Unit, Indian Statistical Institute, Kolkata, 700108, India. diganta@isical.ac.in

# 1 Introduction

The study of networks has exploded in popularity over the last few decades, thanks in part to their wide range of applications (Newman, 2010; Sengupta, 2023). Across these various domains, one common feature that has been observed is *core-periphery (CP) structure* where the network can be broken down into two groups: a small, densely interconnected *core* group and a sparse *periphery* group whose members primarily connect to the core rather than to each other (Borgatti and Everett, 2000; Csermely et al., 2013). CP structures have been observed in global trade (Krugman, 1996; Magone et al., 2016) and airport networks (Lordan and Sallan, 2017, 2019), where nations with large economies and airline hubs comprise the core, respectively. CP structures are fundamentally hierarchical, balancing the benefits of centralization (e.g., efficient coordination) with the reach and scalability of large, diversified and distributed systems. Understanding and modeling such architectures is critical in domains where control, resilience and strategic redundancy are essential. For example, if a major airline’s hub airport went offline due to a storm, this would cause much greater disruption to flight schedules than if a peripheral airport was nonoperational.

There are numerous algorithms and metrics to identify CP structure in a network. The most widely used approach is the Borgatti–Everett (BE) metric (Borgatti and Everett, 2000), which assumes a fixed binary template for “ideal” CP structure and maximizes similarity to this template. Other methods include Rombach et al. (2017) and Kojaku and Masuda (2017) which build off the BE metric, and Holme (2005) who defines the core as the  $k$ -core with largest closeness centrality. These methods lack statistical inference, however, and produce purely *descriptive* outputs: given any network, even one generated from a model with no underlying CP structure, they will still return a CP partition. Without a rigorous statistical basis for probabilistic error bounds or significance testing of these outputs, it is impossible to distinguish true CP structure from artifacts of random noise, algorithmic uncertainty and/or exogenous factors. This makes the resulting labels potentially misleading and undermines their utility for downstream decision-making (Yanchenko and Sengupta, 2023).

Beyond just detection, there has been minimal formulation of CP structure from a statistical inference perspective. The works that do consider statistical properties tend to be fairly narrow in their scope, focusing on just one or two data-generating models (e.g., Zhang et al., 2015; Naik et al., 2021; Gallagher et al., 2021). This means that the nuanced ways in which CP structures manifest in many popular random graph models have not been explored. Similarly, to the knowledge of the authors, there are not any theoretical guarantees that the true CP labels will be recovered by the existing methods. Hypothesis tests for the statistical significance of a CP structure with theoretical guarantees of type-1 (false positive) and type-2 (false negative) errors are also rare. Existing approaches often rely on bootstrap procedures to compute empirical  $p$ -values (Boyd et al., 2006; Kojaku and Masuda, 2017; Elliott et al., 2020), but these typically do not scale well to large networks. Moreover, these tests are generally limited to a single null model, usually the Erdős and Rényi (1959) (ER) model, which is overly simplistic. It has been shown in the context of other network properties like community structure (Yanchenko and Sengupta, 2024) and the small-world property (Lovekar et al., 2025) that many networks differ from the ER model without truly exhibiting the property of interest. That is, the ER null is so weak that nearly any real-world network will reject against it, regardless of true CP structure. Kojaku and Masuda (2017) further

stress the need for null models that account for degree heterogeneity, which the ER model lacks.

In this work, we seek to address these gaps by presenting a principled and comprehensive taxonomy of CP structure from a statistical inference perspective. We begin by defining a model-agnostic population parameter that quantifies the strength of the CP structure in the data-generating process. The generality of this parameter allows statistically rigorous comparison of CP structures across networks of varying size, sparsity and heterogeneity. We focus on four canonical random graph models: the Erdős–Rényi (ER) model, stochastic block model (SBM), Chung–Lu (CL) model and degree-corrected stochastic block model (DCBM). We find that the ER model, by definition, lacks CP structure, while the SBM exhibits *endogenous* (or intrinsic) CP structure under certain parameter regimes. In contrast, the CL model displays only *exogenous* (or externally imposed) CP structure, which is induced entirely by degree heterogeneity. Finally, the DCBM exhibits both endogenous and exogenous CP structure which can interact in non-trivial ways to either reinforce or suppress the overall CP pattern. We then prove a unified result showing that the sample analogue of our proposed parameter achieves perfect recovery (strong consistency) of the true CP labels under the SBM, CL, and DCBM. These results establish CP structure on firm and rigorous statistical foundations, analogous to the treatment of community detection in [Bickel and Chen \(2009\)](#).

Building on this foundation, we develop intersection-based hypothesis tests to detect and validate CP structure against two null models: ER and CL. The ER null enables detection of *any* statistically significant CP structure, while the CL null helps determine whether the observed CP structure *exceeds* what could be explained by degree heterogeneity alone. Through these tests, we formalize CP structure as an intersection criterion where *both* conditions must be satisfied: core–core connectivity must exceed core–periphery connectivity, and core–periphery connectivity must exceed periphery–periphery connectivity. This formulation helps distinguish true CP structure from other mesoscale patterns such as disassortative community structure. Together, these tests provide a simple and practical toolbox for analyzing real-world networks, offering a rich understanding of the underlying generative mechanisms and enabling a level of structural resolution that goes beyond merely detecting the presence of CP structure. We prove theoretical guarantees for type I error and power for both tests. Finally, extensive simulations demonstrate fast and accurate identification of CP structures as well as highly precise and reliable test performance. A comprehensive analysis of thirteen real-world networks reveals a provocative finding: contrary to conventional wisdom about the ubiquity of CP structures, statistically significant CP structure appears to be relatively rare.

While this paper shares some conceptual similarities with [Yanchenko and Sengupta \(2024\)](#), there are several key differences. Most importantly, the goal of this present work is broader, seeking a comprehensive taxonomy of CP structure across various random graph models, and developing an inference framework based on this taxonomy. Second, we prove the asymptotic recovery of the CP labels under three different scenarios. Finally, the hypothesis testing procedures diverge since this paper proposes an intersection test, requiring different proof techniques, and yields analytic cut-offs for multiple null models.

## 2 Model parameter and label recovery

### 2.1 Notation

For this work, we only consider simple, unweighted, undirected networks with no self-loops. Consider a network with  $n$  nodes. Let  $A$  denote the  $n \times n$  adjacency matrix where  $A_{ij} = 1$  if node  $i$  and node  $j$  have an edge, and 0 otherwise. We write  $A \sim P$  as shorthand for  $A_{ij} \mid P_{ij} \stackrel{\text{ind.}}{\sim} \text{Bernoulli}(P_{ij})$  for  $1 \leq i < j \leq n$ . We define a *core-periphery assignment* to be a vector  $\mathbf{c} \in \{0, 1\}^n$  such that  $c_i = 1$  if node  $i$  is in the core and 0 otherwise. Additionally, let  $k = \sum_{i=1}^n c_i$  denote the size of the core and  $\alpha_n = k/n$  its proportion.

### 2.2 Model parameter

To quantify the strength of the CP structure in the data-generating process, we define the parameter

$$\rho(P, \mathbf{c}) := \frac{\sum_{i < j} (P_{ij} - \bar{P}) \Delta_{ij}}{\frac{1}{2}n(n-1)\{\bar{P}(1-\bar{P})\bar{\Delta}(1-\bar{\Delta})\}^{1/2}}, \quad (1)$$

where  $\Delta_{ij}(\mathbf{c}) \equiv \Delta_{ij} = c_i + c_j - c_i c_j$ ,  $\bar{P} = \frac{1}{\binom{n}{2}} \sum_{i < j} P_{ij}$  is the mean edge probability and  $\bar{\Delta}$  is the entry-wise mean of the upper triangular part of matrix  $\Delta$ . The numerator captures the expected number of (centered) core-core and core-periphery edges, which should be large for models with a strong CP structure. The denominator normalizes this quantity by network size, sparsity and core size, allowing for objective comparison across networks and generative models. By construction,  $0 < \rho(P, \mathbf{c}) \leq 1$  when the average core-core and core-periphery connection probabilities exceeds the global average. In particular, if  $P$  represents an “idealized” CP structure where  $P_{ij} = 1$  if  $c_i = 1$  or  $c_j = 1$ , and 0 otherwise, then  $\rho(P, \mathbf{c}) = 1$ . Thus,  $\rho(P, \mathbf{c})$  quantifies CP structure in a manner consistent with our intuition. Importantly,  $\rho(P, \mathbf{c})$  is *model-agnostic* since it is not tied to any specific class of models and applies to any edge-independent random graph model. This generality makes it a flexible foundation for a unified statistical framework.

### 2.3 Statistical models and core-periphery taxonomy

While the CP parameter is universally applicable to any model  $P$ , it is instructive to understand its behavior under specific random graph models of interest. We consider the following four such models.

- Erdős-Rényi (ER) model: Here  $P_{ij} = p$  for all  $i, j$  (Erdős and Rényi, 1959), which means that the ER model tautologically does not exhibit a CP structure (Boyd et al., 2006; Kojaku and Masuda, 2017; Elliott et al., 2020). The metric defined in (1) aligns with this intuition since  $\rho(P, \mathbf{c}) = 0$  for all  $\mathbf{c}$  under the ER model (because in the numerator,  $P_{ij} - \bar{P}$  is always zero).
- CP-SBM: The SBM has been thoroughly studied in the context of community structure (Holland et al., 1983). Since CP structure is a meso-scale property of networks, the SBM is also a natural model for CP structure (e.g., Zhang et al., 2015; Elliott et al., 2020). We

consider a modification of this model for CP structure, called CP-SBM, where  $P_{ij} = \Omega_{c_i c_j}$  and  $\Omega$  is a  $2 \times 2$  symmetric blockmatrix such that

$$\Omega = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } p_{11} > p_{12} = p_{21} > p_{22}. \quad (2)$$

It is obvious that this model possesses inherent or *endogenous* CP structure (with a slight abuse of notation where  $c_i = 0$  is substituted by  $c_i = 2$  for the periphery group). To understand the behavior of  $\rho(P, \mathbf{c})$  under the CP-SBM, note that  $\Delta_{ij} = 1$  coincides with  $P_{ij} = p_{11}$  or  $p_{12}$ , which makes the numerator positive since  $p_{11} > p_{12} > p_{22}$ . More specifically, if (i)  $\alpha_n = 0.5$  or (ii)  $\alpha_n \rightarrow 0$ , then (i)  $\rho(P, \mathbf{c}) \sim \frac{(p_{11}-p_{22})+2(p_{12}-p_{22})}{4\sqrt{3P(1-P)}}$  and (ii)  $\rho(P, \mathbf{c}) \sim \sqrt{2\alpha_n} \cdot \frac{p_{12}-p_{22}}{\sqrt{P(1-P)}}$ . In both cases,  $\rho(P, \mathbf{c}) > 0$  and increases with increasing values of  $p_{11} - p_{22}$ , i.e., core-core – core-periphery connectivity;  $p_{12} - p_{22}$ , i.e., core-periphery – periphery-periphery connectivity; and core size  $\alpha_n$ . These properties accord with our intuition of CP structure and further validate (1) as a sensible CP model parameter.

- Chung-Lu (CL) model: This model allows for degree heterogeneity among nodes by introducing weight parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top$  and defining  $P_{ij} = \theta_i \theta_j$  ([Chung and Lu, 2002](#); [Dasgupta and Sengupta, 2022](#)). This model has no natural or endogenous core and periphery groups. For a given core size  $k$ , however, we can rank the nodes as  $\theta_{(1)} \geq \theta_{(2)} \geq \dots \geq \theta_{(k)} \geq \dots \theta_{(n)}$ , and define  $c_i = 1$  for  $\theta_{(1)}, \dots, \theta_{(k)}$  and  $c_i = 0$  for  $\theta_{(k+1)}, \dots, \theta_{(n)}$ . Then  $\rho(P, \mathbf{c}) \sim \frac{\{\alpha_n^2(\bar{\theta}_c^2 - \bar{\theta}^2) + \alpha_n(\bar{\theta}_c \bar{\theta}_p - \bar{\theta}^2)\}}{\{\bar{\theta}^2(1-\bar{\theta}^2)\Delta(1-\Delta)\}^{1/2}}$ , where  $\bar{\theta}_c$ ,  $\bar{\theta}_p$ , and  $\bar{\theta}$  are the mean  $\theta_i$  values in the core, periphery and overall network, respectively, allowing self-loops for simplicity. If we let  $\alpha_n \rightarrow 0$ , then the numerator becomes  $\alpha_n(\bar{\theta}_c \bar{\theta}_p - \bar{\theta}^2) > 0$  since  $\bar{\theta} \sim \bar{\theta}_p$  and  $\bar{\theta}_c > \bar{\theta}_p$ . Thus, the CL model yields  $\rho(P, \mathbf{c}) > 0$  which seems to indicate the presence of a CP structure, even though we said it is not endogenous. We call this *exogenous* (or externally induced) CP structure, in the sense that this is induced or imposed due to degree heterogeneity. A formal and statistically rigorous distinction between endogenous and exogenous CP structure is a central theme of this paper.
- CP-DCBM: The degree corrected block model (DCBM) ([Karrer and Newman, 2011](#)) extends the SBM to allow for degree heterogeneity, i.e.,  $P_{ij} = \theta_i \Omega_{c_i c_j} \theta_j$  where  $\boldsymbol{\theta}$  is again a weight parameter vector of length  $n$ . If  $\Omega$  satisfies (2), then this model too possesses endogenous CP structure, so we call it CP-DCBM. The CP-DCBM can be interpreted as a combination of the CP-SBM and the CL model.

In addition to endogenous CP structure, the CP-DCBM also possesses exogenous CP structure due to degree heterogeneity. The strength of the overall CP structure depends on how the exogenous component interacts with the endogenous CP structure. If core nodes have larger  $\theta_i$ , the CP signal is amplified; if periphery nodes have larger  $\theta_i$ , however, the exogenous component may weaken or even *cancel out* the endogenous structure. For example, if the core group vertices also have the highest  $\theta_i$  values, then the highest values of  $P_{ij} - \bar{P}$  correspond to  $\Delta_{ij} = 1$ , leading to a high value of  $\rho(P, \mathbf{c})$ . In this case, the exogenous component *reinforces* the endogenous component and the CP structure is amplified. More generally, if  $\alpha_n \rightarrow 0$ , then the CP-DCBM yields  $\rho(P, \mathbf{c}) \sim \sqrt{2\alpha_n} \cdot \frac{(\bar{\theta}_{1:k} p_{12} - \bar{\theta}_{k:n} p_{22}) \bar{\theta}_{k:n}}{\sqrt{P(1-P)}}$ , where

$\bar{\theta}_{1:k} = \frac{1}{k} \sum_{i=1}^k \theta_i$  and  $\bar{\theta}_{k:n} = \frac{1}{n-k} \sum_{i=k+1}^n \theta_i$ . If  $\bar{\theta}_{1:k} p_{12} > \bar{\theta}_{k:n} p_{22}$ , then  $\rho(P, \mathbf{c}) > 0$ . This condition implies, however, that simply having  $p_{12} > p_{22}$  does not guarantee  $\rho(P, \mathbf{c}) > 0$ . If  $p_{12} > p_{22}$  but  $\boldsymbol{\theta}$  is such that  $\bar{\theta}_{1:k} p_{12} < \bar{\theta}_{k:n} p_{22}$ , then  $\rho(P, \mathbf{c}) < 0$ . This occurs because the exogenous component is *misaligned*, suppressing the endogenous CP structure. As a final example, if  $p_{12} = \sqrt{p_{11}p_{22}}$  and  $\theta_i = \sqrt{p/p_{11}}$  for the core nodes while  $\theta_i = \sqrt{p/p_{22}}$  for the periphery nodes, then  $P_{ij} = p$  for all  $i, j$  and this CP-DCBM is equivalent to an ER model ( $\rho(P, \mathbf{c}) = 0$ ). In this case, the degree heterogeneity entirely cancels out the endogenous CP structure.

To summarize, the ER model does not have a CP structure, while the CP-SBM endogenously possesses CP structure. The CL model exhibits an exogenous version of CP structure, while the CP-DCBM possesses both endogenous and exogenous CP structure which could interact in complex ways.

Figure 1 provides a visual representation of the taxonomy of CP structure for these four models. Of particular interest is the scenario under the CP-DCBM where the endogenous CP structure is not suppressed by degree heterogeneity. This is the shaded region in Figure 1 and will be a major focus for the theoretical results in Theorems 2.1, 3.1 and 3.2. Across all models, the proposed metric  $\rho(P, \mathbf{c})$  captures CP strength in a way that aligns with both theoretical intuition and empirical behavior. This interpretability and versatile nature of  $\rho(P, \mathbf{c})$  makes it an especially well-suited population parameter for statistical inference.

## 2.4 Sample metric

In practice, we do not observe  $P$  or  $\mathbf{c}$ , but only the network adjacency matrix,  $A$ . Therefore, we need a sample version of the CP metric for statistical inference. To this end, for some  $\mathbf{c}$ , define

$$T(A, \mathbf{c}) := \frac{\sum_{i < j} (A_{ij} - \bar{A}) \Delta_{ij}}{\frac{1}{2} n(n-1) \{ \bar{A}(1-\bar{A})\bar{\Delta}(1-\bar{\Delta}) \}^{1/2}}. \quad (3)$$

Both the numerator and denominator of  $T(A, \mathbf{c})$  serve as sample analogs of those in the population-level metric  $\rho(P, \mathbf{c})$ , making  $T(A, \mathbf{c})$  a natural estimator of  $\rho(P, \mathbf{c})$ . We formalize the consistency of this estimator in Theorem 2.1. Furthermore,  $T(A, \mathbf{c})$  coincides with the widely used BE metric proposed by [Borgatti and Everett \(2000\)](#), thereby aligning our inference framework with established quantitative approaches in the network science literature. In practice, the core-periphery assignment vector,  $\mathbf{c}$ , is typically unknown, so we estimate it by maximizing  $T(A, \mathbf{c})$ , i.e.,

$$\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} \{ T(A, \mathbf{c}) \}. \quad (4)$$

We adopt the greedy algorithm from [Yanchenko \(2022\)](#) for the optimization in (4). Please see the Supplemental Materials for further details (Algorithm 1).

## 2.5 Consistent recovery of core-periphery labels

We are now ready to prove the first major theoretical contribution of this work: the sample metric in (3) and (4) achieves consistent recovery of the CP labels. All technical details for this and the ensuing section can be found in the Supplemental Materials.

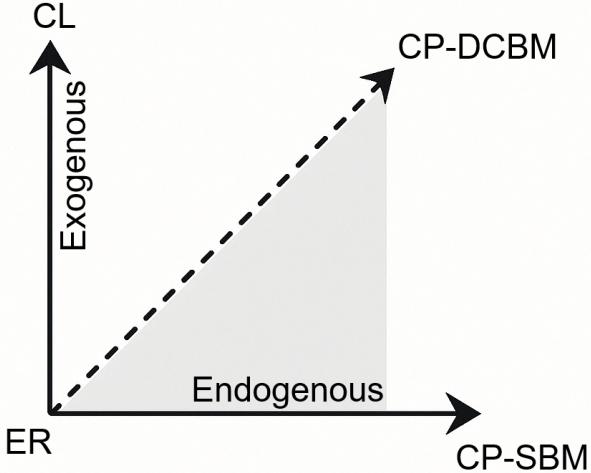


Figure 1: CP components and taxonomy: the  $x$ -axis represents *endogenous* CP structure (i.e., intrinsic, inherent, or innate), and the  $y$ -axis represents *exogenous* CP structure (i.e., externally induced or imposed), typically due to degree heterogeneity. The ER model lies at the origin, exhibiting neither endogenous nor exogenous CP structure. The CP-SBM has purely endogenous CP structure, whereas the CL model only exhibits exogenous CP structure. The CP-DCBM incorporates both endogenous and exogenous CP structure, but these two dimensions can interact in complex ways, potentially offsetting or reinforcing each other. The shaded region illustrates the regime under the CP-DCBM where endogenous structure dominates over exogenous effects. This scenario is rigorously formalized in Theorems 2.1, 3.1 and 3.2.

As per the convention in the literature (Bickel and Chen, 2009; Zhao et al., 2012; Sengupta and Chen, 2015, 2018; Chakrabarty et al., 2025; Bhadra et al., 2025), we introduce the parameter  $\varrho_n$  to control the sparsity of the network, i.e.,  $A \sim P_n = \varrho_n P$  where  $P$  does not depend on  $n$ . Additionally, we assume  $k$  is known and  $k/n := \alpha_n \rightarrow 0$  as  $n \rightarrow \infty$ . Our results can be extended to  $\alpha_n = \mathcal{O}(1)$  but we focus on  $\alpha_n \rightarrow 0$  given that the empirical literature generally has a smaller core than periphery (e.g., Naik et al., 2021). The consistency results also require the following sparsity assumption (A1) on the number of edges in the network and in the core:

$$\mathbf{A1.} \quad \alpha_n \rightarrow 0 \text{ and } n\varrho_n\alpha_n \rightarrow \infty \text{ as } n \rightarrow \infty. \quad (5)$$

Let  $\mathbf{c}^*$  and  $\hat{\mathbf{c}}$  denote the true and estimated core-periphery assignment vector, respectively. By “true” CP labels,  $\mathbf{c}^*$ , we mean the labels corresponding to the model descriptions in Section 2.3, e.g., the true block model labels for the SBM. In the proof of the following theorem, however, we show that this is also equivalent to  $\mathbf{c}^* = \arg \max_{\mathbf{c}} \{\rho(P, \mathbf{c})\}$ . We quantify the estimation error as the fraction of misclassified nodes, i.e.,

$$\xi_n(\hat{\mathbf{c}}) = \frac{1}{n} \sum_{i=1}^n I(\hat{c}_i \neq c_i^*). \quad (6)$$

In the following theorem, we first characterize an upper bound for  $\xi_n(\hat{\mathbf{c}})$  for each of the three models with endogenous and/or exogenous CP structure before further establishing that,

under an additional assumption, the method achieves *exact* recovery or *strong consistency*, i.e.,  $\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) \rightarrow 1$ . To the knowledge of the authors, these are the first results establishing label recovery guarantees of any CP metric.

**THEOREM 2.1.** *Let  $P_n$  be a CP-SBM, CL or CP-DCBM model. Then under (5),*

$$\mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \kappa(P)\alpha_n\sqrt{\frac{(8+\eta)\log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) \rightarrow 1 \text{ for any } \eta > 0 \text{ as } n \rightarrow \infty, \text{ where} \quad (7)$$

- If  $P_n$  is a CP-SBM, then  $\kappa(P) = \sqrt{p_{22}}/(p_{12} - p_{22})$ .
- If  $P_n$  is a CL model with  $\bar{\theta} = \frac{1}{n} \sum_{i=1}^n \theta_n$ ;  $\theta_{(1)} \geq \theta_{(2)} \geq \dots \geq \theta_{(k)} \geq \dots \geq \theta_{(n)}$ ; the “true” CP labels are defined as  $c_i^* = 1$  for the  $k$  vertices with largest  $\theta_i$  values and  $c_i^* = 0$  for the remaining  $(n-k)$  vertices; and under the assumption that  $\theta_{(k)}\theta_{(n)} > \theta_{(k+1)}\theta_{(k+2)}$ ; then  $\kappa(P) = \bar{\theta}/(\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)})$ .
- If  $P_n$  is a CP-DCBM, let  $\theta_c = \min\{\theta_i : c_i^* = 1\}$  be the lowest degree parameter in the core set, and  $\theta_{p,max} = \max\{\theta_i : c_i^* = 0\}$  and  $\theta_{p,min} = \min\{\theta_i : c_i^* = 0\}$  be the highest and lowest degree parameters in the periphery set, respectively; and under the assumption that  $p_{12}\theta_c\theta_{p,min} > \theta_{p,max}^2 p_{22}$ ; then  $\kappa(P) = \sqrt{\bar{p}}/(p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2)$ .

Furthermore, under the additional assumption that  $\alpha_n/\sqrt{\varrho_n} = o(1/\sqrt{n \log n})$ , the labels are strongly consistent for all three models, i.e.,

$$\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) \rightarrow 1 \text{ as } n \rightarrow \infty. \quad (8)$$

**Remark 1.** The error bound in (7) is  $o(\alpha_n)$ , which is stronger than the standard notion of *weak* consistency where  $\xi_n(\hat{\mathbf{c}}) \rightarrow 0$  in probability (Zhao et al., 2012). An error rate of  $o(\alpha_n)$  ensures that the number of mislabeled nodes vanishes not only as a fraction of  $n$ , but also as a fraction of  $k$ , the core size. In fact, since  $\alpha_n \rightarrow 0$ , even the naive estimate  $\hat{\mathbf{c}}^N$  that assigns all nodes to the periphery (i.e.,  $\hat{c}_i^N = 0$  for all  $i$ ) would satisfy weak consistency, as  $\xi_n(\hat{\mathbf{c}}^N) = \alpha_n \rightarrow 0$ . Thus, the sharper error bounds in Theorem 2.1 are crucial for meaningfully characterizing label recovery performance.

**Remark 2.** Strong consistency requires that  $\alpha_n/\sqrt{\varrho_n} = o(1/\sqrt{n \log n})$ , which is equivalent to saying that  $k = o\left(\sqrt{\frac{n\varrho_n}{\log n}}\right)$ . This has an intuitive interpretation that the maximum permissible size of the core for strong consistency depends on the network density. The sparser the network, the lower the ceiling for which strong consistency is guaranteed. When  $\varrho_n = \mathcal{O}(1/\log n)$ , i.e., expected node degrees are almost linearly increasing, strong consistency is guaranteed as long as the core size is  $k = o(\sqrt{n})$ . On the other hand, when  $\varrho_n = \mathcal{O}(\log^2 n/n)$ , i.e., expected node degrees increase at the rate  $\log^2 n$ , the ceiling is only  $k = o(\sqrt{\log n})$ .

**Remark 3.** We can also interpret Theorem 2.1 in light of the discussion in Section 2.3. The parameter  $\kappa(P)$  can be interpreted as the *inverse* signal-to-noise ratio (SNR): the lower the value of  $\kappa(P)$ , the stronger the presence of CP structure in the network, and hence

the lower the error. The CP-SBM has endogenous CP structure, and as long as the core-periphery edge probability  $p_{12}$  is higher than the periphery-periphery edge probability  $p_{22}$ , the proposed method successfully recovers the CP labels. As  $p_{12}$  approaches  $p_{22}$ , however, the core-periphery edges start behaving more similarly to the periphery-periphery edges, and it becomes harder to distinguish the core from the periphery, yielding a larger error bound.

Next, recall that the CL model has exogenous CP structure due to degree heterogeneity. Since  $\theta_{(k)}\theta_{(n)}$  is the lowest core-periphery edge probability and  $\theta_{(k+1)}\theta_{(k+2)}$  is the highest periphery-periphery edge probability, as long as  $\theta_{(k)}\theta_{(n)} > \theta_{(k+1)}\theta_{(k+2)}$ , then core-periphery connectivity is stronger than periphery-periphery connectivity, and lower values of  $\kappa(P)$  imply larger separation between the induced core and periphery, leading to more accurate label recovery. Finally, under the CP-DCBM, degree heterogeneity can interact with the endogenous CP structure in complex ways. The condition  $p_{12}\theta_c\theta_{p,min} > \theta_{p,max}^2 p_{22}$  ensures that the endogenous core is clearly separated from the periphery, i.e., degree heterogeneity does not dominate the underlying blockmodel structure. This condition is the mathematical version of the shaded region in Figure 1.

### 3 Hypothesis Testing

In this section, we expand the statistical framework to hypothesis tests for CP structure. From Section 2.3 and Figure 1, we know that the ER model has no CP structure, while the CP-SBM, CL, and CP-DCBM have endogenous and/or exogenous CP structure. Our first test detects any such CP structure against the ER null. Here,  $H_0$  posits that  $P$  is an ER model, while  $H_1$  states that  $P$  could be any of the three other models from Figure 1.

Next, recall that the CP-DCBM has endogenous CP structure which may be obfuscated by degree heterogeneity. Thus, our second test characterizes this endogenous CP structure against the CL null, formalizing the shaded region of Figure 1. Specifically,  $H_0$  corresponds to  $P$  as a CL model, while  $P$  is a CP-DCBM model with endogenous CP structure under  $H_1$ . In other words, this tests whether the network has a greater CP structure than could be explained by exogenous features alone. Note that our framework does not include a CL vs CP-SBM test, since neither model is a special case of the other.

Our proposed hypothesis tests differ from existing tests (e.g., [Boyd et al., 2010](#); [Kojaku and Masuda, 2017](#); [Elliott et al., 2020](#)) in at least three key ways. First, existing works focus on a single null model (ER or configuration ([Newman, 2018](#))) while we propose a test for both the ER and CL null models. This gives our approach greater flexibility and yields a richer picture of the network structure. Second, we propose an intersection test based on the observation that CP structure requires two features to be present: a core-core edge density greater than that of core-periphery, and a core-periphery edge density greater than that of periphery-periphery. Without this two-part test, non-CP structures, e.g., disassortative community structure, may mistakenly reject the null hypothesis. Finally, existing tests use a bootstrap procedure whereas we derive an analytic cutoff, allowing for better scaling to large networks. Please see the Supplemental Materials for further details on the existing hypothesis tests.

### 3.1 Erdős-Rényi null

The ER model is perhaps the most natural choice for a null model and has been used in previous works (Kojaku and Masuda, 2017; Elliott et al., 2020). For a network to have a CP structure, we claim that the core-core edge density must be greater than the core-periphery edge density as well as the core-periphery edge density being greater than that of the periphery-periphery. Because we must test these two conditions, we propose an intersection test. We use  $T_1(A) := T(A, \hat{\mathbf{c}}) = \max_{\mathbf{c}} \{T(A, \mathbf{c})\}$ , the maximized sample CP metric, and define

$$T_2(A) = \hat{p}_{11} - \hat{p}_{12} = \frac{1}{\binom{k}{2}} \sum_{i,j:\hat{c}_i=\hat{c}_j=1} A_{ij} - \frac{1}{k(n-k)} \sum_{i,j:\hat{c}_i=1,\hat{c}_j=0} A_{ij},$$

i.e., the estimated difference between the probability of a core-core and core-periphery edge. The intersection test is rejected when  $T_1 > C_1$  and  $T_2 > C_2$ , where  $C_1$  and  $C_2$  are defined in (9). The following theorem shows that the CP-SBM, CL, and CP-DCBM all have a statistically significantly stronger CP structure than that of the ER model under specific assumptions.

**THEOREM 3.1.** *Under  $H_0$ , i.e., when the network is generated from the ER model, the Type I error rate converges to 0, i.e., for any  $\eta > 0$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] < \eta$ , where the rejection thresholds are given by*

$$C_1 = \sqrt{\frac{\log(n\varrho_n\alpha_n)}{n}} \text{ and } C_2 = \frac{2\sqrt{2}\varrho_n \log(n)}{\sqrt{k}}. \quad (9)$$

*Under the assumption that  $\alpha_n \geq \frac{(\log(n \log n))^2}{n}$ , suppose that any of the following alternative hypotheses hold:*

- $H_1 : P$  is a CP-SBM where  $p_{11} > p_{12} = p_{21} > p_{22}$
- $H_1 : P$  is a Chung-Lu model where  $\theta_{(k)}\theta_{(n)} > \theta_{(k+1)}\theta_{(k+2)}$  and  $\theta_{(k-1)} > \theta_{(k+1)}$ .
- $H_1 : P$  is a CP-DCBM where  $\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c\theta_{p,min}}$  and  $\frac{p_{11}}{p_{12}} > \frac{\theta_{p,max}}{\theta_c}$ .

*Under each of these alternatives, the power of the test goes to one, i.e., for any  $\eta > 0$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] > 1 - \eta.$$

To our knowledge, this is the first hypothesis test for CP structure with an analytic cutoff, rather than relying on a bootstrap procedure, as well as the first to provide formal statistical guarantees.

**Remark 4.** The proof relies on large-deviation bounds, union bounds over labelings, and asymptotic control of misclassification rates to establish the vanishing Type I error and asymptotic power of the test. Under  $H_0$ , we first prove that the probability of exceeding the threshold for test statistic  $T_1(A)$  vanishes asymptotically by leveraging Lemma 1 (from

proof of Theorem 2.1), which establishes an upper bound on the deviation of  $T(A, \hat{\mathbf{c}})$  from the expected correlation measure  $\rho(P_n, \hat{\mathbf{c}})$ . Since  $\rho(P_n, \mathbf{c}) = 0$  for all possible assignments under  $H_0$ , choosing  $C_1 = \varepsilon_n$  ensures that  $\mathbb{P}[T_1 > C_1] \rightarrow 0$ . For  $T_2(A)$ , the proof utilizes Bernstein's inequality to control the deviation of empirical edge probabilities from their expectations.

Under  $H_1$ , we first show that for  $T_1(A)$ , we have  $T(A, \hat{\mathbf{c}}) \geq T(A, \mathbf{c}^*)$ , where  $\mathbf{c}^*$  is the true community structure. A lower bound on  $\rho(P_n, \mathbf{c}^*)$  is obtained by decomposing it into dominant terms, and using Assumption A1, it is shown that  $\rho(P_n, \mathbf{c}^*) > 2\varepsilon_n$ . This guarantees that  $\mathbb{P}[T_1 > C_1] \rightarrow 1$ . Note that even though  $\rho(P_n, \mathbf{c}^*)$  is a correlation coefficient, because of the assumptions on the data-generating model, it goes to 0 instead of being  $\mathcal{O}(1)$ . For  $T_2(A)$ , the proof carefully handles two sources of error: statistical fluctuations and mislabeling errors. The statistical fluctuations are controlled using Bernstein's inequality as before, while the mislabeling errors are addressed using bounds on  $\xi_n(\hat{\mathbf{c}})$ , the fraction of misclassified nodes. The analysis relies on Theorem 2.1, which ensures that  $\xi_n(\hat{\mathbf{c}}) = o(\alpha_n)$  with high probability. This ensures that the true difference in probabilities,  $\varrho_n(p_{11} - p_{12})$ , dominates the error terms, leading to  $\mathbb{P}[T_2 > C_2] \rightarrow 1$ .

**Remark 5.** On a practical note, when implementing this test, since  $\varrho_n$  is unobserved, we recommend using  $\hat{p} = \frac{1}{n^2} \sum_{i,j} A_{ij}$  as an empirical proxy. Additionally,  $\log(n\varrho_n\alpha_n) = \log(k\varrho_n)$  and  $C_1$  are only defined if  $k\varrho_n > 1$ . If  $k\varrho_n < 1$ , the test cannot be used as either the network is too sparse or the core is too small for the hypothesis testing procedure to be reliable.

### 3.2 Chung-Lu null

Using a CL null model, we again propose an intersection test using the same test statistics,  $T_1(A)$  and  $T_2(A)$ , but modify the rejection threshold for  $T_1(A)$ . Under the null, we estimate the degree parameters as  $\hat{\theta}_i = d_i/\sqrt{2m}$ , where  $d_i = \frac{1}{n} \sum_j A_{ij}$  is the degree of node  $i$  and  $m = \sum_{i < j} A_{ij}$  is the total number of edges. We estimate the null probabilities as  $\hat{P}_{ij} = \hat{\theta}_i \hat{\theta}_j = d_i d_j / (2m)$ . The CL model is a special case of the CP-DCBM when  $p_{11} = p_{12} = p_{22}$ , so we test

$$H_0 : \boldsymbol{\theta}, p_{11} = p_{12} = p_{22} \text{ (CL model)} \text{ vs. } H_1 : \boldsymbol{\theta}, p_{11} > p_{12} > p_{22} \text{ (CP-DCBM).} \quad (10)$$

The following theorem formalizes the statistical consistency of the test.

**THEOREM 3.2.** *Under  $H_0$ , i.e., when the network is generated from the CL model, the Type I error rate converges to 0, i.e., for any  $\eta > 0$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] < \eta$ , where the rejection thresholds are given by*

$$C_1 = \rho(\hat{P}, \hat{\mathbf{c}}) + \tilde{\epsilon}_n + \varepsilon'_n, \text{ where } \tilde{\epsilon}_n = \frac{\sqrt{\alpha_n} \log(n\alpha_n)}{n^{1.5} \sqrt{\varrho_n}}, \text{ and } \varepsilon'_n = \frac{\sqrt{\log(n\alpha_n \varrho_n)}}{n}. \quad (11)$$

*Under  $H_1$ , i.e., when the network is generated from the CP-DCBM, consider the following three assumptions.*

1.  $\varepsilon'_n + \tilde{\epsilon}_n = o(\alpha_n^{1.5} \varrho_n^{0.5})$ , where  $\tilde{\epsilon}_n = \frac{\sqrt{\alpha_n} \log(n\alpha_n)}{n^{1.5} \sqrt{\varrho_n}}$ ,  $\varepsilon'_n = \frac{\sqrt{\log(n\alpha_n \varrho_n)}}{n}$ .

2. There exists  $\epsilon > 0$  such that  $\frac{p_{12}^2 - p_{11}p_{22}}{n^4\alpha_n^2\varrho_n} (S_{10}^2 + S_{01}^2 - S_{11}S_{00}) > \epsilon$ , where  $S_{kl} = \sum_{(i,j): c_i^* = k, c_j^* = l} \theta_i \theta_j$ , for  $k, l = 0, 1$ .
3. Similar to Theorem 2.1, let  $\theta_c = \min\{\theta_i : c_i^* = 1\}$  be the lowest degree parameter in the core set, and  $\theta_{p,max} = \max\{\theta_i : c_i^* = 0\}$  and  $\theta_{p,min} = \min\{\theta_i : c_i^* = 0\}$  be the highest and lowest degree parameters in the periphery set, respectively. Then  $\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c \theta_{p,min}}$ .

Under these three assumptions, the power of the test goes to one, i.e., for any  $\eta > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] > 1 - \eta.$$

**Remark 6.** The proof is similar to that of Theorem 3.1, but requires extra care thanks to the added CL model parameters. Comparing the  $C_1$  cutoffs from the ER and CL null is enlightening as to the effect of the CL null. First, notice that  $\epsilon'_n = \sqrt{n}C_1$  for the  $C_1$  cutoff from the ER null model. Additionally,  $\rho(\hat{P}, \hat{c}) > 0$  under the CL null model, but 0 under the ER model. So it is easy to see that the rejection threshold is much larger for the CL null compared to that of the ER null. This finding is as we might expect since the CL null can be considered a more complicated null model, making rejection more difficult. Indeed, rejecting the CL null implies that the ER null is also rejected, but the opposite does not always hold.

**Remark 7.** For practitioners, we first recommend testing against the ER null. If this test is rejected, the user should next test against the CL null. If this test is not rejected, it means that the network exhibits a CL model structure and/or the observed CP structure is merely a result of degree heterogeneity. On the other hand, if the CL null is rejected, then the network has a stronger CP structure than what could be induced by degree heterogeneity alone, and is therefore likely to be endogenous. Together, these tests provide a rich understanding of the underlying CP structure and enable a level of interpretation and structural resolution that goes beyond merely detecting the presence of CP structure.

## 4 Simulations

### 4.1 Detection

In the first set of simulations, we compare the Algorithm 1 against competing methods for determining the core of a network. We compare the BE metric with several well-known metrics: Rombach ([Rombach et al., 2017](#)), Curcuringu ([Cucuringu et al., 2016](#)) and BayesSBM. BayesSBM fits the stochastic block model using a Gibbs sampler and in this way is a similar to Zhang and Moore ([2014](#)) and Gallagher et al. ([2021](#)). All other methods use the `cpnet` Python package and `Reticulate` package to run in R. Please see later in the Supplemental Materials for more details on these competing methods. In particular, the method from ([Rombach et al., 2017](#)) has a comparative advantage since it is given knowledge of the true value of  $k$ , so its performance should be taken with a grain of salt. Again, we generate networks, apply the detection algorithms and report the classification accuracy and run-time, while varying  $\tilde{\rho}(P) = \max_c\{\rho(P, c)\}$  and  $n$ . Unless otherwise stated, we fix  $k = 0.1n$ , i.e., 10% of the nodes are in the core.

### 4.1.1 Stochastic block model

We first generate networks from a CP-SBM. Let  $n = 1000$  and fix  $p_{22} = n^{-1} = 0.001$ . Then we vary  $p_{12} = 0.002, 0.0044, \dots, 0.026$  and  $p_{11} = 2p_{12}$  which induces  $\tilde{\rho}(P) \in (0.01, 0.13)$ . The results are in Figure 2. Initially, Rombach has the largest classification accuracy, but for all  $\tilde{\rho}(P) \geq 0.05$ , the proposed algorithm is the most accurate method. BayesSBM yields perfect accuracy around  $\tilde{\rho}(P) = 0.08$  but does not perform as well as Algorithm 1. The Rombach metric begins strong, but plateaus without reaching 100% accuracy. Curcuringu starts with a very low accuracy until  $\tilde{\rho}(P) > 0.08$  at which point it quickly increases. In terms of computation time, Algorithm 1 and Curcuringu are significantly faster than the other two approaches. We then fix  $p_{11} = 0.015$ ,  $p_{12} = 0.0075$  and  $p_{22} = 0.001$  ( $\tilde{\rho}(P) = 0.06$ ) and vary  $n = 500, 750, \dots, 2500$ . The results are in Figure 2b. Again, the BE metric clearly performs the best in terms of both classification accuracy, and has a comparable computation time to that of Curcuringu. BayesSBM is able to reach 100% classification accuracy while Rombach again plateaus before that point.

### 4.1.2 Chung-Lu model

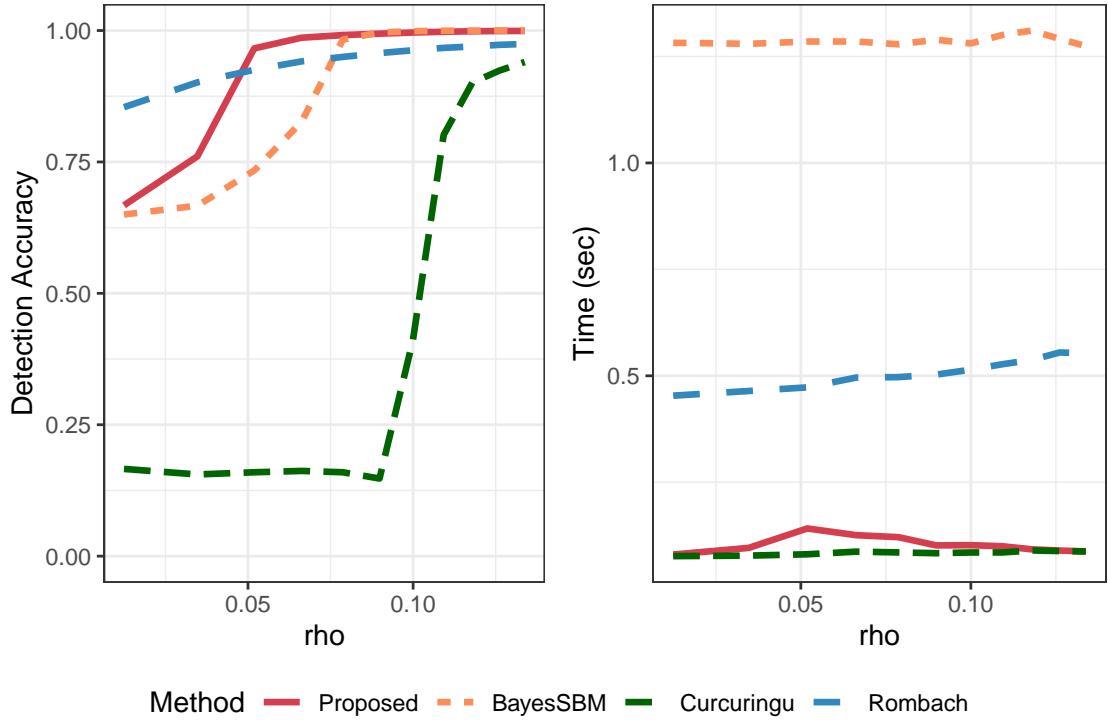
Next, we compare the algorithms on networks generated from a CL model. Let  $n = 1000$  and without loss of generality, let  $\theta_1, \dots, \theta_k$  be the weight parameters for the core nodes, and  $\theta_{k+1}, \dots, \theta_n$  for the periphery nodes. We sample  $\theta_1, \dots, \theta_k \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.1 + \delta/2, 0.2 + \delta/2)$  and  $\theta_{k+1}, \dots, \theta_n \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.1 - \delta/2, 0.2 - \delta/2)$  for  $\delta = 0.00, 0.03, \dots, 0.10$ . As  $\delta$  increases, the separation between the weight parameters in the core and periphery also increases, which should make identifying the core easier. The results are in Figure 3a.<sup>1</sup> Initially, Rombach has the best detection accuracy, followed by the proposed method, but they are comparable for  $\tilde{\rho}(P) \geq 0.06$ . BayesSBM performs slightly worse than the proposed method, while Curcuringu has consistently the lowest accuracy. In Figure 3b, we fix  $\delta = 0.16$  ( $\tilde{\rho}(P) = 0.06$ ) and vary  $n = 500, 750, \dots, 2000$ . Again, the proposed method and Rombach have similar detection accuracy for all  $n$ . In both settings, the proposed method and Curcuringu are noticeably faster than the other two methods.

### 4.1.3 Degree-corrected block model

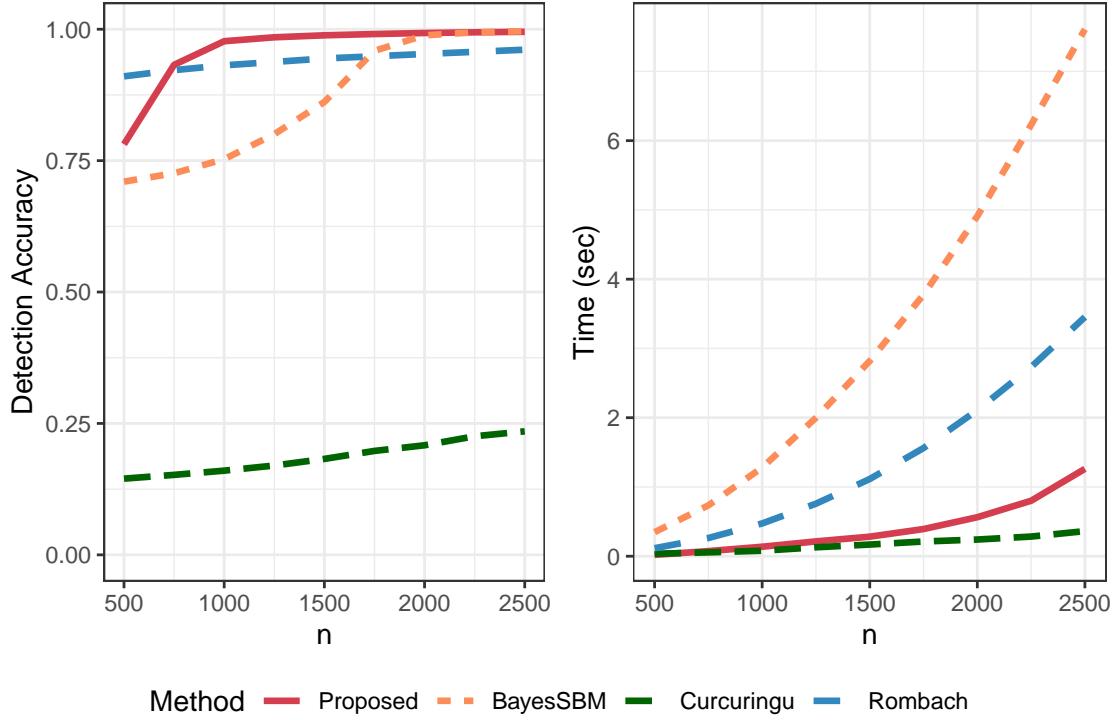
We also compare the algorithms on networks generated from a CP-DCBM. Let  $n = 1000$  and generate  $\theta_i \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.6, 0.8)$ . We set  $p_{22} = 0.001$  and vary  $p_{12} = 0.002, \dots, 0.08$  with  $p_{11} = 2p_{12}$  such that  $\tilde{\rho}(P) \in (0.00, 0.11)$ . The results are in Figure 4a. While the proposed algorithm and BayesSBM have similar classification accuracy, Rombach performs well at first before being overtaken around  $\tilde{\rho}(P) = 0.06$ . The accuracy for Curcuringu is low for all parameter settings. In Figure 4b, we report the results for  $p_{11} = 0.01$ ,  $p_{12} = 0.005$  and  $p_{22} = 0.001$  ( $\tilde{\rho}(P) = 0.06$ ) with  $n = 500, 750, \dots, 2000$ . The trends are similar to the other settings; Algorithm 1 has the highest classification accuracy for all  $n \geq 1000$ . BayesSBM's accuracy increases with  $n$  while those of Curcuringu and Rombach are roughly constant and

---

<sup>1</sup>These simulations were run on a different computer than the other settings, so the computation times may not be comparable across sub-sections.

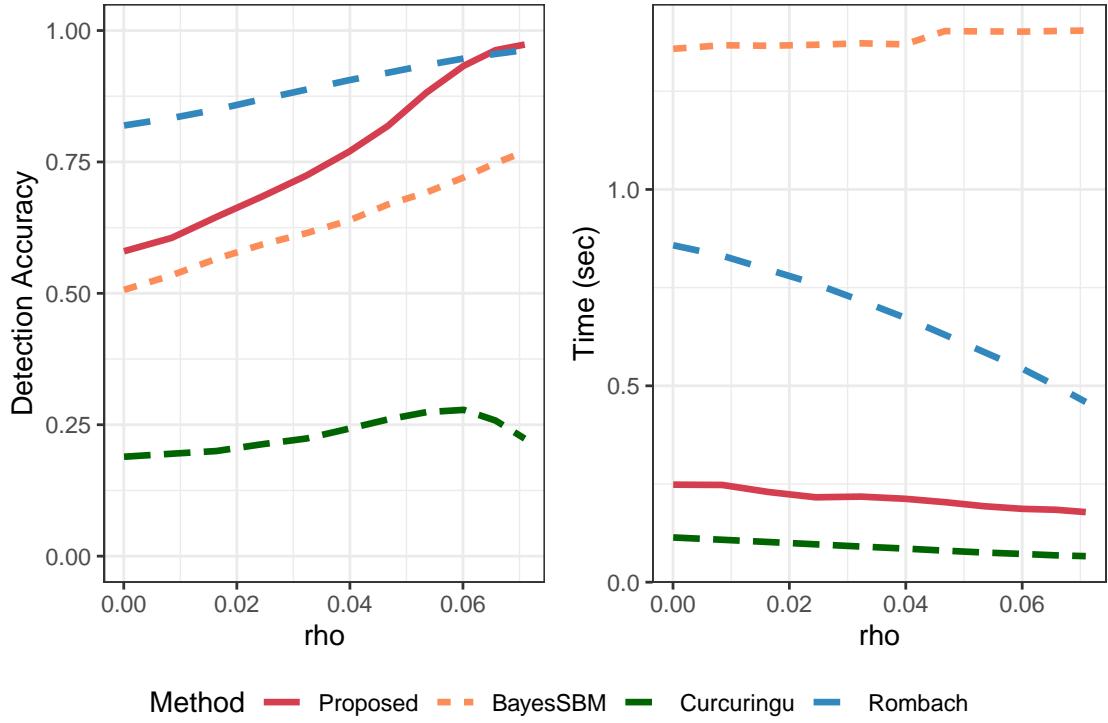


(a) Detection accuracy and computing time for networks generated from a CP-SBM with increasing  $\tilde{\rho}(P)$ .

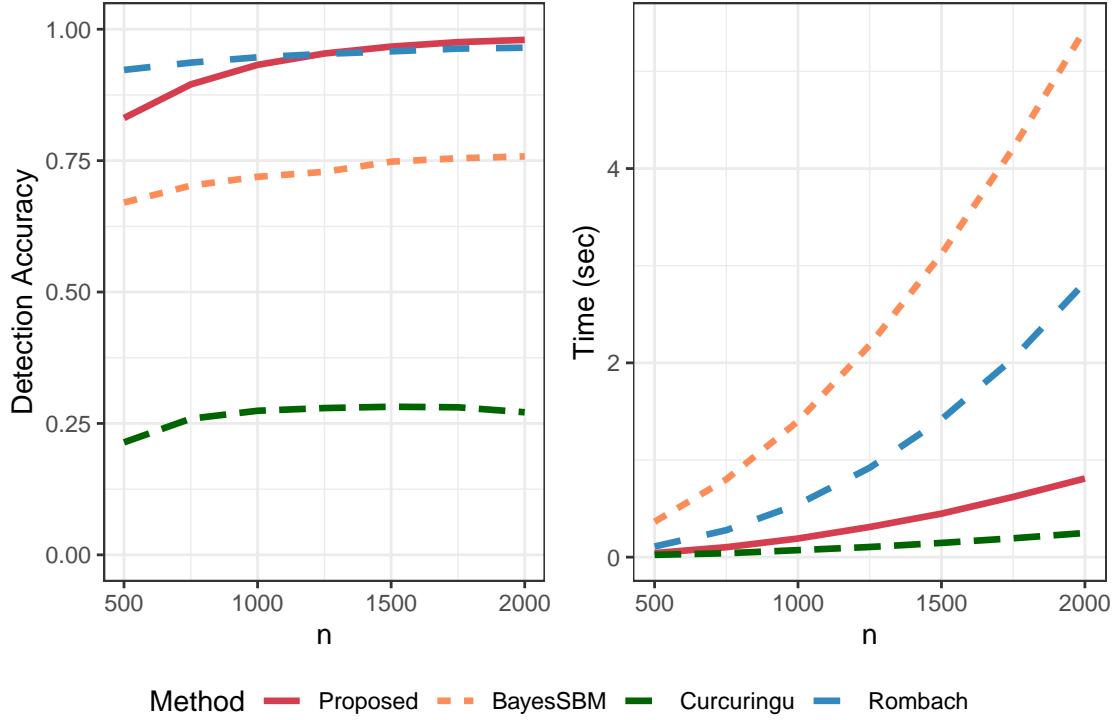


(b) Detection accuracy and computing time for networks generated from a CP-SBM with increasing  $n$ .

Figure 2: Core-periphery identification results with CP-SBM-generated networks.



(a) Detection accuracy and computing time for networks generated from a CL model with increasing  $\tilde{\rho}(P)$ .



(b) Detection accuracy and computing time for networks generated from a CL model with increasing  $n$ .

Figure 3: Core-periphery identification results with CL-generated networks.

do not reach 100%. Again, Curcuringu and the proposed method consistently have the fastest computing time.

## 4.2 Hypothesis testing

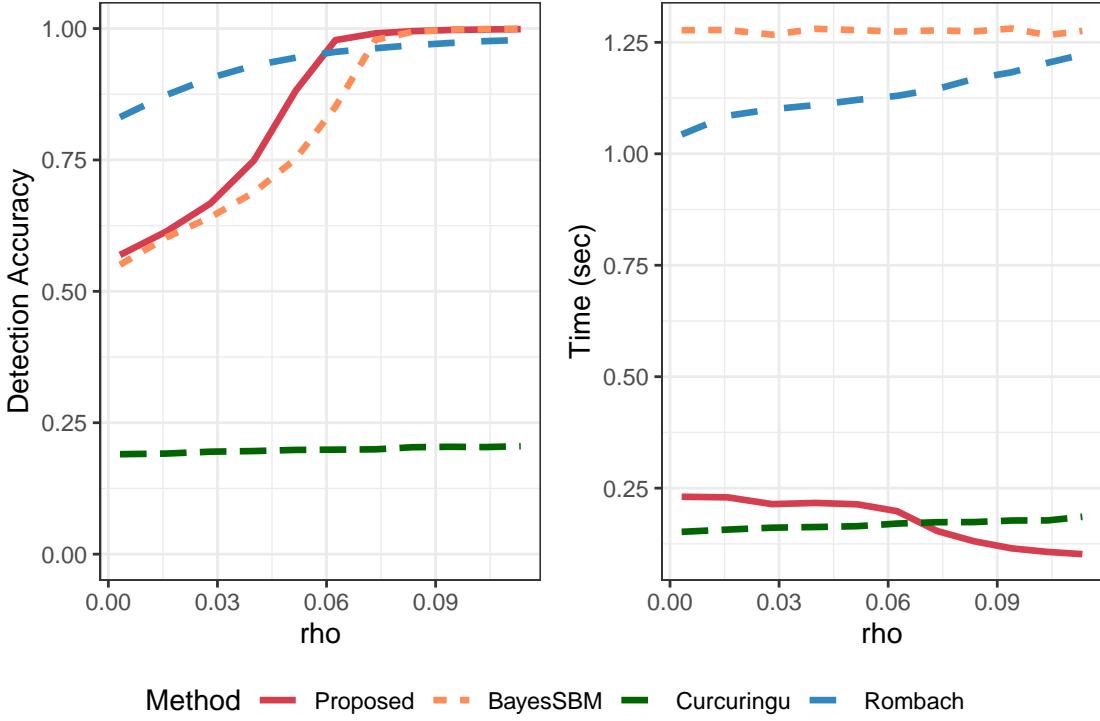
We turn our attention to analyzing the performance of the proposed hypothesis tests in Theorems 3.1 and 3.2. As there are no other existing tests for CP structure with statistical guarantees, we cannot make a fair comparison with our method, so we only consider the proposed tests. We conduct simulations testing against the ER and CL null models and various alternative models.

### 4.2.1 ER vs. CP-SBM

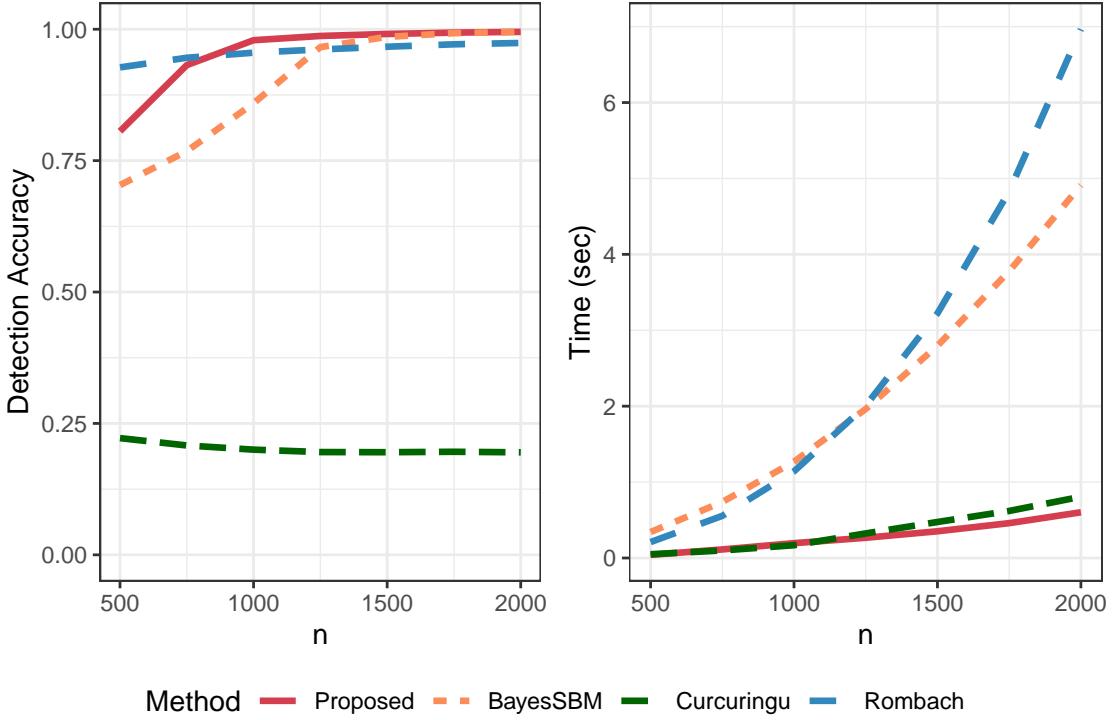
We leverage Theorem 3.1 to test the CP-SBM alternative model against an ER null. In particular, let  $n = 1000$  and  $p_{22} = 0.01$ . We vary  $p_{12} = 0.01, \dots, 0.05$  of length 10 and set  $p_{11} = 2p_{12}$ , implying  $\tilde{\rho}(P) \in (0.00, 0.13)$ , and generate networks from a CP-SBM. For  $p_{12} > 0.01$ , all networks have a CP structure, so we expect a large rejection rate. The results are in Figure 5a. As the strength of the CP structure in the network increases (as measured by  $\tilde{\rho}(P)$ ), the rejection rate of the test also increases. Next, we fix  $p_{11} = 0.05, p_{12} = 0.02$  and  $p_{22} = 0.01$ , which implies  $\tilde{\rho}(P) = 0.04$ , and vary  $n = 1000, 1250, \dots, 2500$ . These networks also have a CP structure so we expect a large rejection rate. The results are in Figure 5b. As  $n$  increases, so too does the power of the test. Since our theoretical results hold asymptotically, it is sensible that the test performs better for large  $n$ . Finally, we generate networks *without* a CP structure to ensure that the test maintains a small rejection rate. In particular, we generate networks with disassortative community structure where  $p_{11} = p_{22} = 0.01$  and  $p_{12} = 0.01, \dots, 0.10$  of length 5, and fix  $n = 1000$ . These parameters yield  $\tilde{\rho} \in (0.00, 0.21)$ . The results are in Figure 5c. The proposed test consistently fails to reject these networks since they do not have a CP structure.

### 4.2.2 ER vs. CL

We can also use Theorem 3.1 to test the ER null model against a CL model with CP structure. To generate the CL networks, we again let  $n = 1000$  and sample  $\theta_1, \dots, \theta_k \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.2 + \delta/2, 0.4 + \delta/2)$  and  $\theta_{k+1}, \dots, \theta_n \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.2 - \delta/2, 0.4 - \delta/2)$  for  $\delta = 0.00, 0.04, \dots, 0.40$ . The rejection rates are in Figure 6a. We see that rejection rates increase around  $\tilde{\rho}(P) = 0.08$  and reach a power of by  $\tilde{\rho}(P) = 0.11$ . Next, we fix  $\delta = 0.22$  and vary  $n = 1000, 1200, \dots, 2000$ . These results are in Figure 6b and show an increasing rejection rate as  $n$  increases. Finally, we generate CL networks where the assumption for Theorem 3.1 is not met. We fix  $n = 1000$ , let  $\theta_i \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.3 - \delta/2, 0.3 + \delta/2)$  for  $\delta = 0.00, 0.04, \dots, 0.40$  and assign the  $k$  nodes with largest value of  $\theta_i$  to the core. The results are in Figure 6c. For all values of  $\tilde{\rho}(P)$ , the rejection rate is 0. The reason for this can be seen in Table 1. Here, we report the average value of  $\theta_{(k)}\theta_{(n)}$  and  $\theta_{(k+1)}\theta_{(k+2)}$  for each value of  $\delta$ . Recall that one of the assumptions for the test in Theorem 3.1 is that  $\theta_{(k)}\theta_{(n)} > \theta_{(k+1)}\theta_{(k+2)}$ . Looking at this table, however, it is clear that this assumption is not being met, leading to the low rejection rate. This result empirically validates the theoretical condition of Theorem 3.1

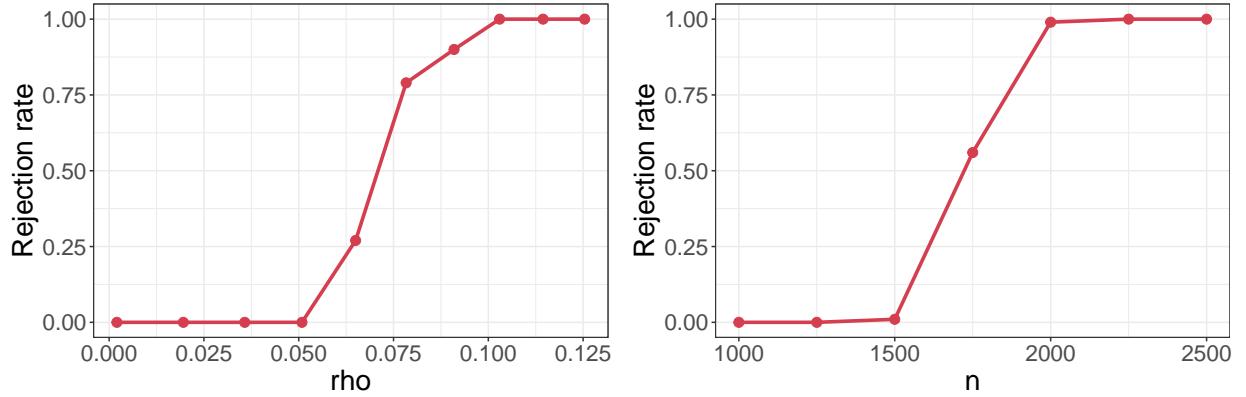


(a) Detection accuracy and computing time for networks generated from a CP-DCBM with increasing  $\tilde{\rho}(P)$ .



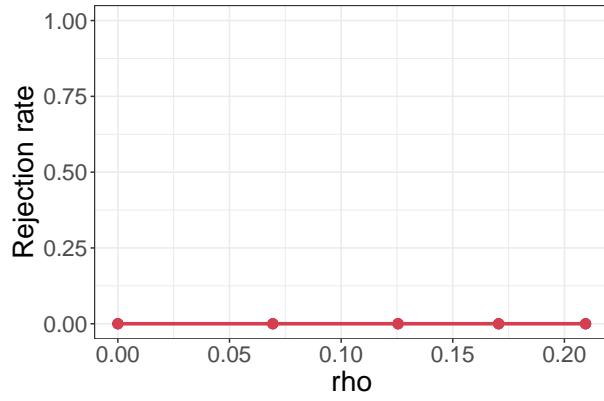
(b) Detection accuracy and computing time for networks generated from a CP-DCBM with increasing  $n$ .

Figure 4: Core-periphery identification results with CP-DCBM-generated networks.



(a) CP-SBM networks with varying  $\tilde{\rho}(P)$ .

(b) CP-SBM networks with varying  $n$ .



(c) SBM networks with disassortative community structure (no CP).

Figure 5: Rejection rates for networks generated from the CP-SBM being tested against the ER null.

$\delta$	0.00	0.04	0.08	0.12	0.16	0.20	0.24	0.28	0.32	0.36	0.40
$\theta_{(k)}\theta_{(n)}$	0.09	0.09	0.09	0.08	0.08	0.08	0.07	0.07	0.06	0.05	0.05
$\theta_{(k+1)}\theta_{(k+2)}$	0.09	0.10	0.11	0.12	0.13	0.14	0.16	0.17	0.18	0.20	0.21

Table 1: Average value of certain weight parameters when testing ER vs. CL model with  $\theta_i \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.3 - \delta/2, 0.3 + \delta/2)$  for  $\delta = 0.00, 0.04, \dots, 0.40$ .

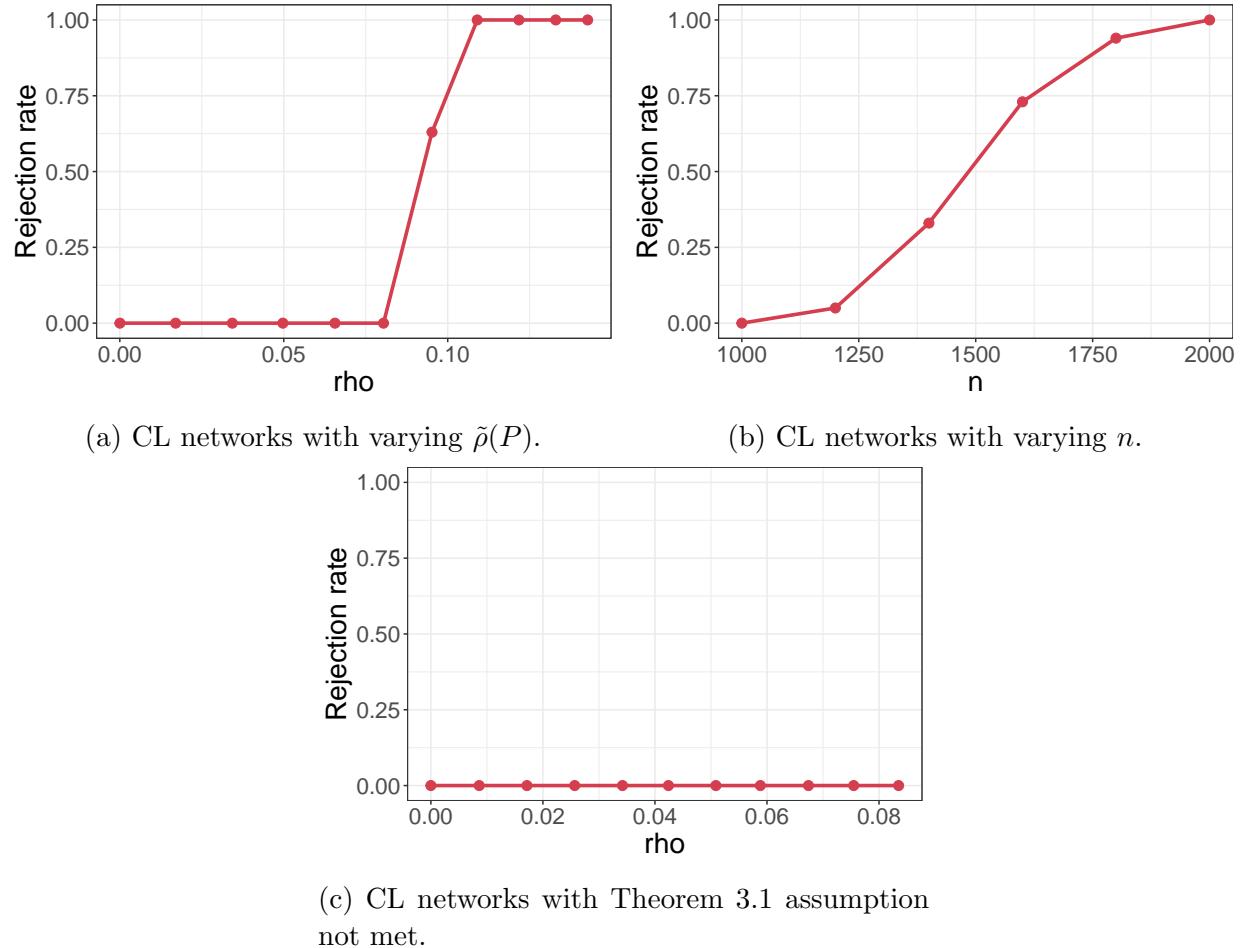


Figure 6: Rejection rates for networks generated from the CL model being tested against the ER null.

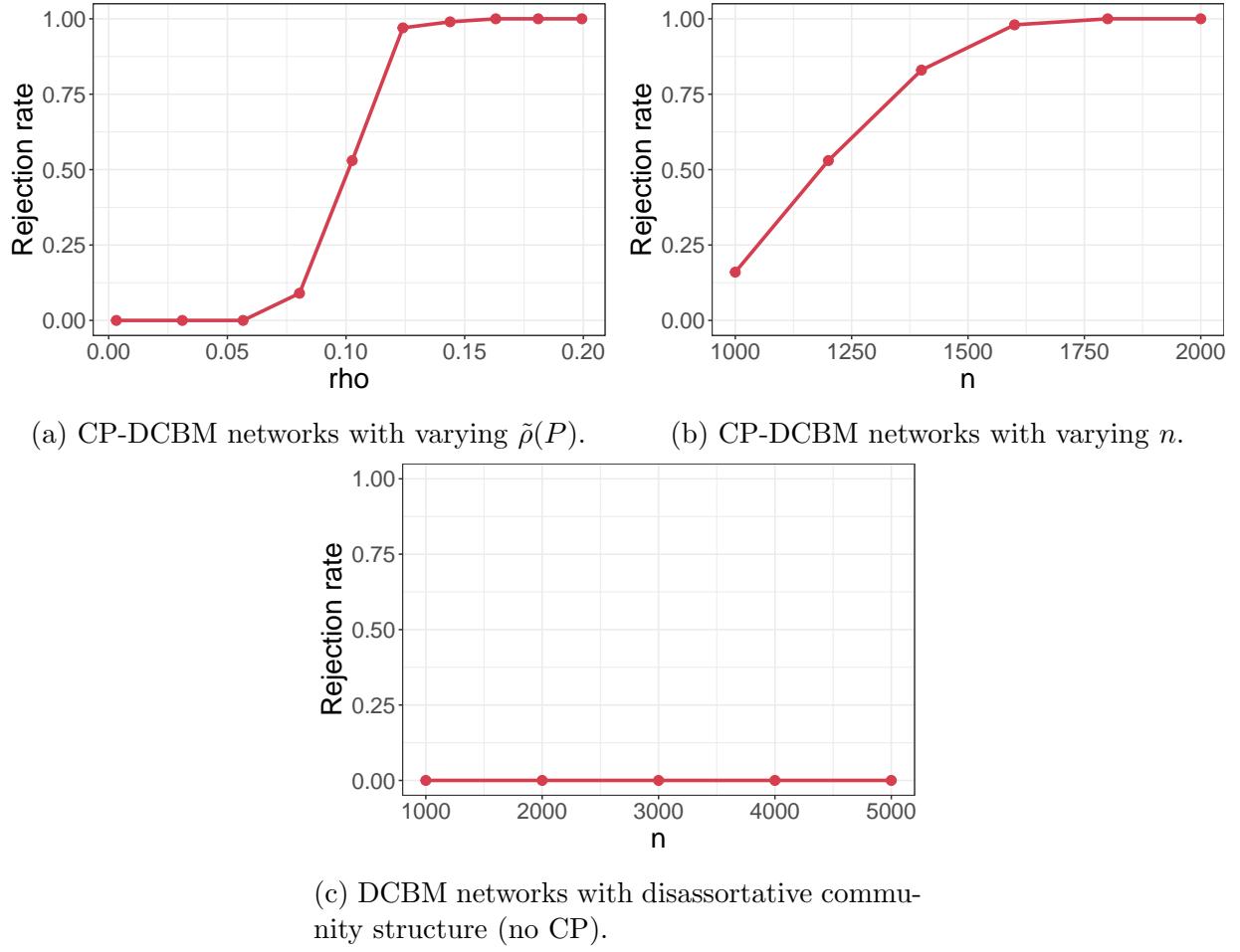


Figure 7: Rejection rates for networks generated from the CP-DCBM being tested against the ER null.

#### 4.2.3 ER vs. CP-DCBM

We conclude testing of the ER null model by using the CP-DCBM as the alternative model. We first fix  $n = 1000$  and generate network from a CP-DCBM with a CP structure where  $\theta_i \sim \text{Uniform}(0.6, 0.8)$ ,  $p_{22} = 0.05$  and vary  $p_{12} = 0.05, \dots, 0.25$  of length 10 with  $p_{11} = 2p_{12}$ . This implies that  $\tilde{\rho}(P) \in (0.00, 0.20)$ . When  $p_{12} > p_{22}$ , by Theorem 3.1, we expect to see a large rejection rate. The results are in Figure 8a and we can see that the rejection rate increases with increasing  $\tilde{\rho}(P)$ . Next, we fix  $p_{11} = 2p_{12} = 0.24$  and  $p_{22} = 0.05$  which yields  $\tilde{\rho}(P) = 0.08$ , and vary  $n = 1000, 1200, \dots, 2000$ . The results are in Figure 8b which demonstrate an increasing rejection rate with  $n$ . Lastly, we generate disassortative DCBM community structure (no CP) to ensure the test yields a small rejection rate. We let  $\theta_i \stackrel{\text{iid.}}{\sim} \text{Uniform}(0.6, 0.8)$ ,  $p_{11} = 0.05$ ,  $p_{12} = 0.10$  and  $p_{22} = 0.05$  such that  $\tilde{\rho}(P) = 0.05$ . We increase  $n = 1000, 2000, \dots, 5000$  with results in Figure 8c. We can see that even for large  $n$ , the rejection rate is still 0.

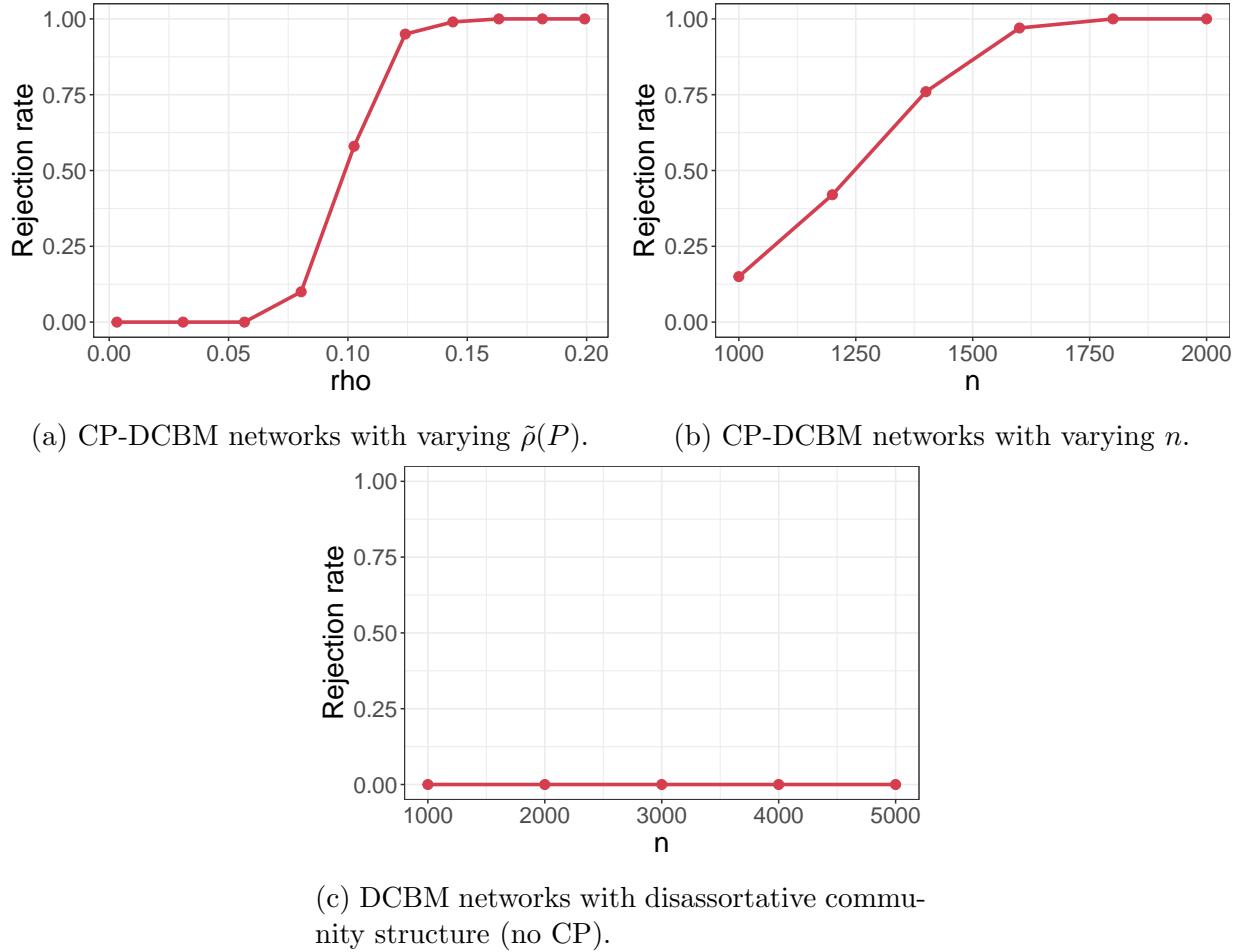


Figure 8: Rejection rates for testing the CL null against the CP-DCBM.

#### 4.2.4 CL vs. CP-DCBM

Finally, we test the CL null model against the CP-DCBM alternative model using the thresholds derived in Theorem 3.2. We use the same parameter settings as in the ER vs. CP-DCBM scenario with results in Figure 8a. The results are quite similar to those using the ER null. We do notice, however, that the rejection rate increases slightly slower with  $n$  using the CL null.

## 5 Real-world data

Finally, we apply the proposed hypothesis tests to thirteen real-world networks. Please see the Supplemental Materials for details on these datasets. For each network, we apply Algorithm 1 to find the CP labels  $\hat{\mathbf{c}}$  and test statistics  $T_1(A)$  and  $T_2(A)$ , and apply the hypothesis tests from Theorems 3.1 and 3.2 to test for statistically significant CP structure. The testing results are reported in Table 2, with the numeric values in the SI.

All thirteen networks yield  $T_1(A) > C_1$  for the ER null, but only seven networks reject the intersection test, including all of the biological networks. This means that these networks

have a greater CP structure than we would otherwise expect from a network generated from the ER model, and also highlight the importance of the intersection test. Had we only considered the first part of the test based on  $T_1(A)$ , we would have concluded that all networks had a CP structure. By explicitly checking  $T_2(A) > C_2$ , however, we find that many networks do not reject the null hypothesis. This indicates that these networks may possess disassortative community structure, instead of a CP structure. Thus, our test not only gives a rejection result, but also indicates why the test failed to reject.

Theorem 3.1 tells us that when the ER null is rejected, it could be because of a CP-SBM, CL or CP-DCBM model, so it is also important to test against the CL null. Of the seven networks which rejected the ER null hypothesis, only four now reject the CL null model (BritishMP, PolBlogs, CopenBT and Biological 1). For the other three networks (Facebook, Biological 2 and 3), this means that they do not possess an endogenous CP structure, and their previous rejection was likely due to degree heterogeneity. We stress that failing to reject the CL null does not necessarily imply that the network was generated from a CL model. Rather, it means that the data-generating mechanism has weaker CP structure than a CL model, but this could arise from a variety of models. Thus, the test cannot determine what model generated the network, but only compare the strength of CP structure in the data-generating models.

## 6 Discussion

In this work, we provided a principled treatment of core-periphery structure from a statistical inference perspective. Our model-agnostic population parameter quantifies the underlying CP structure and allows for a taxonomy of endogenous vs. exogenous CP structure. We proved novel theoretical results showing exact recovery of the true CP labels while also proposing hypothesis tests against the ER and CL null models. Throughout the work, we highlighted the nuances of distinguishing endogenous and exogenous CP structure, a difficulty that most clearly arises for the CL model. In general, we leave it to the user to determine whether the exogenous CP structures present in CL networks should be considered *true* CP structure. Indeed, in the real data analysis, we found that about half the networks rejected the ER null model but only half of these also rejected the CL null. This result can be interpreted in one of two ways. If we believe that the CL model is a reasonable null model, then this means that CP structure is relatively rare in real-world networks. Conversely, these results can reinforce a finding in [Kojaku and Masuda \(2018\)](#) which is that the ER null is too easy to reject, whereas the CL null is too difficult. Under this interpretation, many real-world networks do have a CP structure, but the CL null is too high of a threshold to clear, thus leading to many false negatives. These results also show that developing other meaningful null models to represent “no CP structure” is an important open research problem. Future work could also include extending our methods to more complicated CP structures, i.e., more than two blocks ([Kojaku and Masuda, 2018](#)) or layered structures ([Gallagher et al., 2021](#)).

**Conflicting Interests** The authors have no conflicting interests to declare.

Network	$n$	$\hat{p}$	$\alpha_n$	$T_1$	$T_2$	Erdos-Renyi			Chung-Lu		
						$T_1 > C_1$	$T_2 > C_2$	$\cap$	$T_1 > C_1$	$T_2 > C_2$	$\cap$
FRIENDSHIP											
UK faculty	81	0.18	0.20	0.26	0.34	✓					
Facebook	4039	0.01	0.07	0.10	0.17	✓	✓	✓			✓
COMM.											
Email	167	0.24	0.13	0.42	0.41	✓			✓		
BritishMP	381	0.08	0.16	0.25	0.40	✓	✓	✓	✓	✓	✓
Congress	475	0.09	0.14	0.18	0.16	✓			✓		
CITATION											
PolBlogs	1224	0.02	0.07	0.21	0.23	✓	✓	✓	✓	✓	✓
COLLAB.											
DBLP	2203	0.47	0.24	0.26	0.23	✓			✓		
CONTACT											
Hospital	75	0.41	0.29	0.44	0.38	✓			✓		
School	252	0.29	0.21	0.22	0.34	✓					
CopenBT	672	0.09	0.19	0.17	0.19	✓	✓	✓	✓	✓	✓
BIOLOGICAL											
Biological 1	2220	0.02	0.18	0.13	0.10	✓	✓	✓	✓	✓	✓
Biological 2	3289	0.02	0.08	0.13	0.27	✓	✓	✓			✓
Biological 3	4040	0.01	0.09	0.10	0.13	✓	✓	✓			✓

Table 2: Real-data analysis results where  $n$  is the number of nodes,  $\hat{p}$  is the average edge density and  $\alpha_n$  is the proportion of nodes assigned to the core. The remaining values are defined in Theorem 3.1 and 3.2.

**Code and data sharing** The code for all methods is available on the author’s GitHub: <https://github.com/eyanchenko/CPinference>. All datasets are freely available online.

**Funding** Srijan Sengupta’s work was supported in part by the National Science Foundation grant DMS-2413327 and the NIH grant 7R01LM013309.

## References

- Adamic, L. A. and Glance, N. (2005). The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43.
- Bhadra, S., Pensky, M., and Sengupta, S. (2025). Scalable community detection in massive networks via predictive assignment. *arXiv preprint arXiv:2503.16730*.
- Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073.
- Borgatti, S. P. and Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, 21(4):375–395.
- Boyd, J. P., Fitzgerald, W. J., and Beck, R. J. (2006). Computing core/periphery structures and permutation tests for social relations data. *Social Networks*, 28(2):165–178.
- Boyd, J. P., Fitzgerald, W. J., Mahutga, M. C., and Smith, D. A. (2010). Computing continuous core/periphery structures for social relations data with minres/svd. *Social Networks*, 32(2):125–137.
- Chakrabarty, S., Sengupta, S., and Chen, Y. (2025). Subsampling based community detection for large networks. *Statistica Sinica*, 35(3):1–42.
- Cho, A., Shin, J., Hwang, S., Kim, C., Shim, H., Kim, H., Kim, H., and Lee, I. (2014). Wormnet v3: a network-assisted hypothesis-generating server for caenorhabditis elegans. *Nucleic acids research*, 42(W1):W76–W82.
- Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882.
- Csermely, P., London, A., Wu, L.-Y., and Uzzi, B. (2013). Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2):93–123.
- Csárdi, G., Nepusz, T., Traag, V., Horvát, S., Zanini, F., Noom, D., and Müller, K. (2025). *igraph: Network Analysis and Visualization in R*. R package version 2.1.4.
- Cucuringu, M., Rombach, P., Lee, S. H., and Porter, M. A. (2016). Detection of core–periphery structure in networks using spectral methods and geodesic paths. *European Journal of Applied Mathematics*, 27(6):846–887.
- Dasgupta, A. and Sengupta, S. (2022). Scalable estimation of epidemic thresholds via node sampling. *Sankhya A*, 84:321–344.
- Elliott, A., Chiu, A., Bazzi, M., Reinert, G., and Cucuringu, M. (2020). Core–periphery structure in directed networks. *Proceedings of the Royal Society A*, 476(2241):20190783.
- Erdős, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297.

- Fink, C. G., Fullin, K., Gutierrez, G., Omodt, N., Zinnecker, S., Sprint, G., and McCulloch, S. (2023). A centrality measure for quantifying spread on weighted, directed networks. *Physica A*.
- Gallagher, R. J., Young, J.-G., and Welles, B. F. (2021). A clarified typology of core-periphery structure in networks. *Science advances*, 7(12):eabc9800.
- Gao, J., Liang, F., Fan, W., Sun, Y., and Han, J. (2009). Graph-based consensus maximization among multiple supervised and unsupervised models. *Advances in neural information processing systems*, 22.
- Greene, D. and Cunningham, P. (2013). Producing a unified graph representation from multiple social network views. In *Proceedings of the 5th annual ACM web science conference*, pages 118–121.
- Holland, P., Laskey, K., and Leinhardt, S. (1983). Stochastic blockmodels: first steps. *Social Networks*, 5:109–137.
- Holme, P. (2005). Core-periphery organization of complex networks. *Phys. Rev. E*, 72:046111.
- Ji, M., Sun, Y., Danilevsky, M., Han, J., and Gao, J. (2010). Graph regularized transductive classification on heterogeneous information networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 570–586. Springer.
- Karrer, B. and Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83:016107.
- Kojaku, S. and Masuda, N. (2017). Finding multiple core-periphery pairs in networks. *Physical Review E*, 96(5):052313.
- Kojaku, S. and Masuda, N. (2018). Core-periphery structure requires something else in the network. *New Journal of Physics*, 20(4):043012.
- Krugman, P. (1996). *The self organizing economy*. John Wiley & Sons.
- Leskovec, J. and Krevl, A. (2014). SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>.
- Leskovec, J. and Mcauley, J. (2012). Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25.
- Lordan, O. and Sallan, J. M. (2017). Analyzing the multilevel structure of the european airport network. *Chinese Journal of Aeronautics*, 30(2):554–560.
- Lordan, O. and Sallan, J. M. (2019). Core and critical cities of global region airport networks. *Physica A: Statistical mechanics and its applications*, 513:724–733.
- Lovekar, K., Sengupta, S., and Paul, S. (2025). Testing for the network small-world property. *Electronic Journal of Statistics*, 19(1):1453–1506.
- Magone, J. M., Laffan, B., and Schweiger, C. (2016). *Core-periphery relations in the European Union: Power and conflict in a dualist political economy*. Routledge.
- Michalski, R., Palus, S., and Kazienko, P. (2011). Matching organizational structure and social network extracted from email communication. In *Business Information Systems: 14th International Conference, BIS 2011, Poznań, Poland, June 15-17, 2011. Proceedings* 14, pages 197–206. Springer.
- Naik, C., Caron, F., and Rousseau, J. (2021). Sparse networks with core-periphery structure. *Electronic Journal of Statistics*, 15(1):1814 – 1868.
- Nepusz, T., Petrőczi, A., Négyessy, L., and Bazsó, F. (2008). Fuzzy communities and the

- concept of bridgeness in complex networks. *Physical Review E*, 77(1):016107.
- Newman, M. (2018). *Networks*. Oxford university press.
- Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
- Rombach, P., Porter, M. A., Fowler, J. H., and Mucha, P. J. (2017). Core-periphery structure in networks (revisited). *SIAM Review*, 59(3):619–646.
- Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *AAAI*.
- Sapiezynski, P., Stopczynski, A., Lassen, D. D., and Lehmann, S. (2019). Interaction data from the copenhagen networks study. *Scientific Data*, 6(1):315.
- Sengupta, S. (2023). Statistical network analysis: Past, present, and future. *arXiv preprint arXiv:2311.00122*.
- Sengupta, S. and Chen, Y. (2015). Spectral clustering in heterogeneous networks. *Statistica Sinica*, pages 1081–1106.
- Sengupta, S. and Chen, Y. (2018). A block model for node popularity in networks with community structure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(2):365–386.
- Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaggiotto, M., Van den Broeck, W., Régis, C., Lina, B., et al. (2011). High-resolution measurements of face-to-face contact patterns in a primary school. *PloS one*, 6(8):e23176.
- Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Regis, C., Kim, B.-A., Comte, B., and Voirin, N. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS one*, 8:e73970.
- Yanchenko, E. (2022). A divide-and-conquer algorithm for core-periphery identification in large networks. *Stat*, page e475.
- Yanchenko, E. and Sengupta, S. (2023). Core-periphery structure in networks: a statistical exposition. *Statistic Surveys*, 17:42–74.
- Yanchenko, E. and Sengupta, S. (2024). A generalized hypothesis test for community structure in networks. *Network Science*, 12(2):122–138.
- Zhang, P. and Moore, C. (2014). Scalable detection of statistically significant communities and hierarchies, using message passing for modularity. *Proceedings of the National Academy of Sciences*, 111(51):18144–18149.
- Zhang, X., Martin, T., and Newman, M. E. J. (2015). Identification of core-periphery structure in networks. *Physical Review E*, 91:032803.
- Zhao, Y., Levina, E., and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292.

## Supplemental Materials

### Table of Contents

7. Greedy algorithm for CP identification
8. Discussion of existing hypothesis tests
9. Competing methods description
10. Real-data analysis results
11. Proofs for label recovery
12. Proofs for hypothesis test

## 7 Algorithm

We adopt the greedy algorithm of [Yanchenko \(2022\)](#) to maximize the test statistic. The algorithm works as follows. First, we randomly generate initial CP labels. Then one at a time and for each node  $i$ , we swap the label of the node and compute the test statistic with these new labels. The new label is kept if that test statistic is larger than before, otherwise we keep the original labels. This process repeats until the labels are unchanged for an entire cycle of all  $n$  nodes. Note that the node order is randomly shuffled each iteration and multiple starting labels can be run simultaneously to avoid finding a local max. The steps are outlined in Algorithm 1.

---

**Algorithm 1** Greedy

---

**Result:** Core-periphery labels  $\mathbf{c}$

**Input:**  $n \times n$  adjacency matrix  $A$

Initialize labels  $\mathbf{c}$

$run = 1$

**while**  $run > 0$  **do**

$run = 0$

Randomly order nodes

**for**  $i$  in  $1, \dots, n$  **do**

Swap label of node  $i$ :  $\mathbf{c}^* = \mathbf{c}$ ,  $c_i^* = 1 - c_i$

**if**  $T(A, \mathbf{c}^*) > T(A, \mathbf{c})$  **then**

$\mathbf{c} \leftarrow \mathbf{c}^*$

$run = 1$

**end**

**end**

**end**

---

## 8 Existing hypothesis tests

We briefly describe three existing hypothesis tests for CP structure found in the literature. [Boyd et al. \(2010\)](#) adopt the BE metric to quantify the strength of the CP structure in the original network. They then propose a bootstrap test, generating bootstrap networks by re-wiring edges while preserving the total number of edges as the original network. The BE metric is computed on these bootstrap samples and an empirical  $p$ -value is then computed. Note that re-wiring edges and preserving the total number of edges is analogous to the ER null as the ER model preserves the *expected* number of edges on the network. [Elliott et al. \(2020\)](#) extend the BE definition of CP structure to directed networks. This metric is then used to carry out a bootstrap testing procedure where the bootstrap networks are generated from an ER model. Lastly, in [Kojaku and Masuda \(2017\)](#), the authors show that if a configuration null model is used, then it is impossible to generate networks with a CP structure. Thus, they propose introducing at least one more block into the network. By defining a modularity-like metric to quantify the CP structure in the network, the authors generate configuration model bootstrap samples to carry out the hypothesis test.

## 9 Methods

We explain the details of the Rombach and BayesSBM methods used in the simulation study.

### Rombach

This method comes from [Rombach et al. \(2017\)](#) and the output is a scalar coreness  $C_i$  for each node  $i = 1, \dots, n$ . To convert these to binary labels, we sort the nodes from largest to smallest based on  $C_i$ . Then we choose the nodes with  $k$  largest coreness value where  $k$  is the true core size. This gives this method a comparative advantage as it has partial knowledge of the data-generating mechanism (known  $k$ ).

### BayesSBM

This method computes the core labels using a Bayesian paradigm. The network  $A$  is assumed to be from a two-block SBM with likelihood,

$$L(A|\mathbf{p}, \mathbf{c}) = \prod_{i < j} p_{c_i c_j}^{A_{ij}} (1 - p_{c_i c_j})^{1 - A_{ij}}.$$

We can approximate this as a Poisson likelihood such that

$$L(A|\mathbf{p}, \mathbf{c}) = \prod_{i < j} \frac{e^{-p_{c_i c_j}} A_{ij}^{p_{c_i c_j}}}{A_{ij}}.$$

Then we place a uniform prior on  $\mathbf{c}$ , i.e.,

$$\pi(\mathbf{c}) = \frac{1}{2^n}$$

and let

$$\pi(\mathbf{p}) \propto \mathbb{I}(p_{11} > p_{12} > p_{22})$$

which enforces a CP structure into the fit, similar to [Gallagher et al. \(2021\)](#). Then we use Gibbs sampling to approximate the posterior of  $\mathbf{p}$  and  $\mathbf{c}$  and assign nodes to the core if their posterior probability of being in the core is greater than 50%, i.e.,  $P(c_i = 1|A, \mathbf{p}) > 0.5$ .

## 10 Real-data analysis

### Data sets

In order to examine the proposed test's properties under diverse settings, we chose networks of various sizes, densities and from various domains including: friendship ([Nepusz et al., 2008](#); [Leskovec and Mcauley, 2012](#)), communication ([Michalski et al., 2011](#); [Greene and Cunningham, 2013](#); [Fink et al., 2023](#)), citation ([Adamic and Glance, 2005](#)), collaboration ([Gao et al., 2009](#); [Ji et al., 2010](#)), contact ([Vanhems et al., 2013](#); [Stehlé et al., 2011](#); [Sapiezynski et al., 2019](#)) and biological networks ([Cho et al., 2014](#)). All data sets were download using

Network	$T_1$	$T_2$	Erdos-Renyi		Chung-Lu	
			$C_1$	$C_2$	$C_1$	$C_2$
FRIENDSHIP						
UK faculty	0.262	0.340	0.114	0.553	0.262	0.553
Facebook	0.097	0.170	0.017	0.015	0.103	0.015
COMMUNICATION						
Email	0.423	0.411	0.099	0.724	0.382	0.724
BritishMP	0.248	0.400	0.065	0.181	0.247	0.181
Congress	0.185	0.164	0.062	0.193	0.176	0.193
CITATION						
PolBlogs	0.207	0.232	0.024	0.047	0.189	0.047
COLLABORATION						
DBLP	0.261	0.231	0.050	0.448	0.246	0.448
CONTACT						
Hospital	0.441	0.380	0.171	1.069	0.398	1.069
School	0.218	0.341	0.106	0.614	0.229	0.614
CopenBT	0.170	0.185	0.061	0.154	0.169	0.154
BIOLOGICAL						
Biological 1	0.126	0.103	0.031	0.024	0.124	0.024
Biological 2	0.127	0.276	0.021	0.022	0.137	0.022
Biological 3	0.101	0.134	0.018	0.011	0.106	0.011

Table 3: Real-data analysis results where all values are defined in Theorem 3.1 and 3.2.

(Leskovec and Krevl, 2014; Rossi and Ahmed, 2015; Csárdi et al., 2025). The networks and summary statistics are listed in Table 2, where we remove all edge weights and direction as well as self-loops. Additionally, the DBLP network was pre-processed to only consider two groups of researchers (Sengupta and Chen, 2018).

## Results

In Table 3, we report the observed values of the test statistics for the 13 real-world networks from Section 4, as well as the cutoff and rejection decision. This table provides more details than that in the main text, as it shows not only the binary rejection decision, but also the raw values.

## 11 Technical proofs for label recovery

We state and prove four lemmas in this section. The first lemma is a general result that will be used to prove the subsequent results. Lemmas 2-4 are supporting results for the CL model, the CP-SBM, and the CP-DCBM, respectively. Finally, we state and prove Theorem 2.1.

For the results in this section, we will use the identity that

$$\rho(P, \mathbf{c}) := \frac{\sum_{i < j} (P_{ij} - \bar{P}) \Delta_{ij}}{\frac{1}{2}n(n-1)\{\bar{P}(1-\bar{P})\bar{\Delta}(1-\bar{\Delta})\}^{1/2}} = \frac{\sum_{i < j} (P_{ij} - \bar{P})(\Delta_{ij} - \bar{\Delta})}{\frac{1}{2}n(n-1)\{\bar{P}(1-\bar{P})\bar{\Delta}(1-\bar{\Delta})\}^{1/2}},$$

since

$$\sum_{i < j} (P_{ij} - \bar{P}) = \bar{\Delta} \sum_{i < j} (P_{ij} - \bar{P}) = 0,$$

and, similarly,

$$T(A, \mathbf{c}) := \frac{\sum_{i < j} (A_{ij} - \bar{A}) \Delta_{ij}}{\frac{1}{2}n(n-1)\{\bar{A}(1-\bar{A})\bar{\Delta}(1-\bar{\Delta})\}^{1/2}} = T(A, \mathbf{c}) := \frac{\sum_{i < j} (A_{ij} - \bar{A})(\Delta_{ij} - \bar{\Delta})}{\frac{1}{2}n(n-1)\{\bar{A}(1-\bar{A})\bar{\Delta}(1-\bar{\Delta})\}^{1/2}}.$$

### Lemma 1.

LEMMA 1. Let  $\varepsilon_n = \alpha_n^{1/2} \varrho_n^{1/2} \sqrt{\frac{\log(n\varrho_n\alpha_n)}{n\varrho_n\alpha_n}}$ . From the definitions of  $\alpha_n$  and  $\varrho_n$  and assumption A1, clearly  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then under A1 we have

$$\mathbb{P}(\max_{\mathbf{c}} \{|T(A_n, \mathbf{c}) - \rho(P_n, \mathbf{c})|\} > \varepsilon_n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

*Proof.* Let  $\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} \{T(A, \mathbf{c})\}$ . Recall also that  $A \sim \varrho_n P$  and  $k/n = \mathcal{O}(\alpha_n)$ .

Now, consider some fixed  $\mathbf{c}$  and let  $X_n(\mathbf{c}) = \sum_{i < j} A_{ij}(\Delta_{ij} - \bar{\Delta}_n)$  and  $Y_n = \sum_{i < j} A_{ij}$ . Then

$$T(A_n, \mathbf{c}) = \frac{X_n}{\sqrt{Y_n(\binom{n}{2}) - Y_n}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}} := f(X_n, Y_n).$$

We will prove the desired result using two pieces. First, we show that  $X_n(\mathbf{c})$  and  $Y_n$  are “close” to their respective expectations,  $\mu_x(\mathbf{c})$  and  $\mu_y$ , for all  $\mathbf{c}$ . Then we will show that  $f(x, y)$  is “close” to  $f(x^*, y^*)$  when  $x \approx x^*$  and  $y \approx y^*$ .

First, Bernstein’s inequality states that if  $\mathsf{E}(X_i) = 0$  and  $|X_i| \leq M$  for  $i = 1, \dots, n$ , then

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) \leq \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^n \mathsf{E}(X_i^2) + \frac{1}{3}Mt}\right).$$

Thus,

$$\mathbb{P}(|X_n(\mathbf{c}) - \mu_x(\mathbf{c})| > t_1) \leq 2 \exp\left(-\frac{\frac{1}{2}t_1^2}{\sum_{i < j} P_{ij}(1 - P_{ij})(\Delta_{ij} - \bar{\Delta}_n)^2 + \frac{1}{3}t_1}\right) \sim \exp\left(-\frac{t_1^2}{\varrho_n \alpha_n n^2 + t_1}\right)$$

since  $P_{ij}(1 - P_{ij}) = \mathcal{O}(\varrho_n)$  and  $\sum_{i < j} (\Delta_{ij} - \bar{\Delta}_n)^2 = \mathcal{O}(\alpha_n n^2)$ . Additionally,

$$\mathbb{P}(|Y_n - \mu_y| > t_2) \leq 2 \exp\left(-\frac{\frac{1}{2}t_2^2}{\sum_{i < j} P_{ij}(1 - P_{ij}) + \frac{1}{3}t_2}\right) \sim \exp\left(-\frac{t_2^2}{\varrho_n n^2 + t_2}\right).$$

Let  $\gamma_n = \sqrt{\frac{\log(n\varrho_n\alpha_n)}{n\varrho_n\alpha_n}}$  and  $t_1 = \varrho_n\alpha_n n^2 \gamma_n$ . Note that from assumption A1, we have  $\gamma_n \rightarrow 0$  and  $n\varrho_n\alpha_n \gamma_n^2 \rightarrow \infty$  as  $n \rightarrow \infty$ . Then taking the union over  $2^n$  CP assignments,

$$\mathbb{P}(\max_{\mathbf{c}} |X_n - \mu_x(\mathbf{c})| > \varrho_n\alpha_n n^2 \gamma_n) \leq 2^{n+1} \exp\left(-\frac{\varrho_n^2 \alpha_n^2 n^4 \gamma_n^2}{2\varrho_n \alpha_n n^2}\right) \leq \exp(n \log 2 - \varrho_n \alpha_n n^2 \gamma_n^2) \rightarrow 0$$

since  $n\varrho_n\alpha_n \gamma_n^2 \rightarrow \infty$ . Additionally,  $Y_n$  does not depend on  $\mathbf{c}$  so letting  $t_2 = \varrho_n n^2 \gamma_n$ ,

$$\mathbb{P}(|Y_n - \mu_y| > \varrho_n n^2 \gamma_n) \leq \exp\left(-\frac{\varrho_n^2 n^4 \gamma_n^2}{\varrho_n n^2 + \varrho_n n^2 \gamma_n}\right) \leq \exp(-\varrho_n n^2 \gamma_n^2) \rightarrow 0.$$

Thus, for large  $n$ ,  $X_n(\mathbf{c})$  and  $Y_n$  are “close” to their respective expectations for all  $\mathbf{c}$ .

For the second step, recall that

$$f(x, y) = \frac{x}{\sqrt{y(n^2 - y)}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}}.$$

Then we want to show that  $|f(x, y) - f(x^*, y^*)|$  is upper bounded by a small constant when  $|x - x^*| < t_1$  and  $|y - y^*| < t_2$ . Consider the first-order Taylor expansion:

$$f(x, y) - f(x^*, y^*) = (x - x^*) \frac{\partial}{\partial x} f(x, y, n) \Big|_{x=x^*, y=y^*} + (y - y^*) \frac{\partial}{\partial y} f(x, y, n) \Big|_{x=x^*, y=y^*} + R_n,$$

where

$$R_n \sim (x - x^*, y - y^*) \begin{pmatrix} \partial_x^2 f & \partial_{xy} f \\ \partial_{xy} f & \partial_y^2 f \end{pmatrix} \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}.$$

We have that  $x^* = \mathcal{O}(\alpha_n \varrho_n n^2)$  and  $y^* = \mathcal{O}(\varrho_n n^2)$ . It's also easy to see that

$$\frac{\partial}{\partial x} f(x, y) \Big|_{x=x^*, y=y^*} = \frac{1}{\sqrt{y^*(n^2 - y^*)}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}} \sim \alpha_n^{-1/2} \varrho_n^{-1/2} n^{-2},$$

and

$$\frac{\partial}{\partial y} f(x, y) \Big|_{x=x^*, y=y^*} = \frac{x^*(n^2 - 2y^*)}{2\{y^*(n^2 - y^*)\}^{3/2}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}} \sim \alpha_n^{1/2} \varrho_n^{-1/2} n^{-2}.$$

From step 1,  $|x - x^*| < t_1 = \alpha_n \varrho_n n^2 \gamma_n$  and  $|y - y^*| < t_2 = \varrho_n n^2 \gamma_n$ , so

$$f(x, y) - f(x^*, y^*) \sim \alpha_n \varrho_n n^2 \gamma_n \cdot \frac{1}{\alpha_n^{1/2} \varrho_n^{1/2} n^2} + \varrho_n n^2 \gamma_n \frac{\alpha_n^{1/2}}{\varrho_n^{1/2} n^2} + R_n \sim \alpha_n^{1/2} \varrho_n^{1/2} \gamma_n + R_n.$$

Additionally,

$$R_n \sim \frac{\{y^*(n^2 - y^*)\}^{-5/2}}{\sqrt{\Delta_n(1 - \bar{\Delta}_n)}} (x - x^*, y - y^*) \begin{pmatrix} 0 & -2(n^2 - 2y^*)\{y^*(n^2 - y^*)\} \\ -2(n^2 - 2y^*)\{y^*(n^2 - y^*)\} & x^*(3n^4 - 8n^2y^* + 8(y^*)^2) \end{pmatrix} \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}$$

$$\sim \alpha_n^{1/2} \varrho_n^{1/2} \gamma_n^2$$

such that

$$f(x, y) - f(x^*, y^*) \sim \alpha_n^{1/2} \varrho_n^{1/2} \gamma_n.$$

We are now ready to show the desired result. Let

$$E = \left\{ \bigcap_{\mathbf{c}} \{|X(\mathbf{c}) - \mu_x(\mathbf{c})| < t_1\} \right\} \bigcap \{|Y - \mu_y| < t_2\}.$$

Then

$$\begin{aligned} \mathbb{P}(\max_{\mathbf{c}} \{|T(A_n, \mathbf{c}) - \rho(P_n, \mathbf{c})|\} > \varepsilon_n) &= \mathbb{P}(\max_{\mathbf{c}} \{|f(X, Y) - f(\mu_x, \mu_y)| > \varepsilon_n\}) \\ &\leq \mathbb{P}(\max_{\mathbf{c}} \{|f(X, Y) - f(\mu_x, \mu_y)| > \varepsilon_n | E\} \mathbb{P}(E) + \mathbb{P}(\max_{\mathbf{c}} \{|f(X, Y) - f(\mu_x, \mu_y)| > \varepsilon_n | E^c\} \mathbb{P}(E^c)) \end{aligned}$$

Now, by step 1,  $\mathbb{P}(E^c) \lesssim \exp(-\varrho_n n^2 \gamma_n^2) + \exp(n - \varrho_n \alpha_n n^2 \gamma_n^2)$ . Thus,

$$\begin{aligned} \mathbb{P}(\max_{\mathbf{c}} \{|T(A_n, \mathbf{c}) - \rho(P_n, \mathbf{c})|\} > \varepsilon_n) &\leq \mathbb{P}(\max_{\mathbf{c}} \{|f(X, Y) - f(\mu_x, \mu_y)| > \varepsilon_n | E\} + \mathbb{P}(E^c)) \\ &\lesssim \exp(-\varrho_n n^2 \gamma_n^2) + \exp(n - \varrho_n \alpha_n n^2 \gamma_n^2) \rightarrow 0 \end{aligned}$$

by step 2 and A1.  $\square$

## Lemma 2.

LEMMA 2. Let  $P_n$  be a two-block SBM where  $P_n = \varrho_n P$  and  $P_{ij} = \Omega_{c_i^* c_j^*}$  where

$$\Omega = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } p_{11} > p_{12} = p_{21} > p_{22}.$$

Then for any  $\mathbf{c} \neq \mathbf{c}^*$  such that the size of the core is the same for  $\mathbf{c}$  and  $\mathbf{c}^*$ ,  $\rho(P_n, \mathbf{c}) < \rho(P_n, \mathbf{c}^*)$ , and we have

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) = \xi_n(\mathbf{c}) \sqrt{\frac{\varrho_n}{2\alpha_n}} \frac{p_{12} - p_{22}}{\sqrt{p_{22}}} (1 + o(1)).$$

*Proof.* Consider

$$\rho(P_n, \mathbf{c}) = \frac{\sum_{i < j} \varrho_n (P_{ij} - \bar{p})(\Delta_{ij} - \bar{\Delta})}{\sqrt{\binom{n}{2} \varrho_n \bar{p} (1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta} (1 - \bar{\Delta})}}.$$

Now, let  $\Delta(\mathbf{c}^*)$  and  $\Delta(\mathbf{c})$  denote the  $\Delta$  matrices corresponding to the labels  $\mathbf{c}^*$  and  $\mathbf{c}$ , respectively. Since  $\mathbf{c}^*$  and  $\mathbf{c}$  have the same core size, we have

$$1 - \bar{\Delta}(\mathbf{c}^*) = 1 - \bar{\Delta}(\mathbf{c}) = \alpha_3 = \frac{(n-k)(n-k-1)}{n(n-1)} = 1 + o(1)$$

as  $k = o(n)$  from A1, and

$$\bar{\Delta}(\mathbf{c}^*) = \bar{\Delta}(\mathbf{c}) = 1 - \alpha_3 = \frac{2k(n-k) - k(k-1)}{n(n-1)} = 2\alpha_n(1 + o(1)).$$

Now, given

$$\Omega = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } p_{11} > p_{12} = p_{21} > p_{22},$$

we have

$$\bar{p} = \alpha_{n,k}^{(1)} p_{11} + \alpha_{n,k}^{(2)} p_{12} + \alpha_{n,k}^{(3)} p_{22},$$

where

$$\alpha_{n,k}^{(1)} = \frac{\binom{k}{2}}{\binom{n}{2}}, \alpha_{n,k}^{(2)} = \frac{k(n-k)}{\binom{n}{2}}, \text{ and } \alpha_{n,k}^{(3)} = \frac{\binom{n-k}{2}}{\binom{n}{2}}.$$

For notational simplicity, we will write  $\alpha_{n,k}^{(i)}$  as  $\alpha_i$  for  $i = 1, 2, 3$ . Note that  $\alpha_i > 0$  for  $i = 1, 2, 3$  and  $\sum_{i=1}^3 \alpha_i = 1$ . Also note that under A1,  $\bar{p} = p_{22}(1 + o(1))$ . Therefore, the denominator of  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$  is given by

$$\sqrt{\binom{n}{2} \varrho_n \bar{p} (1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta} (1 - \bar{\Delta})} = \binom{n}{2} \sqrt{2 \varrho_n \alpha_n p_{22} (1 - \varrho_n p_{22})} (1 + o(1)).$$

Now consider the numerators of  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$  and note that  $\Delta_{ij}(\cdot)$  is the only term that changes between  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$ . Therefore,

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) = \frac{\sum_{i < j} \varrho_n (P_{ij} - \bar{p}) (\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}))}{\binom{n}{2} \sqrt{2 \varrho_n \alpha_n p_{22} (1 - \varrho_n p_{22})}} (1 + o(1)).$$

Note that

$$\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}) = \begin{cases} +1 & \{c_i = 0\} \cap \{c_j = 0\} \cap \{\{c_i^* = 1\} \cup \{c_j^* = 1\}\} \\ -1 & \{c_i^* = 0\} \cap \{c_j^* = 0\} \cap \{\{c_i = 1\} \cup \{c_j = 1\}\} \\ 0 & \text{otherwise} \end{cases}.$$

Also,

$$P_{ij} - \bar{p} = \begin{cases} \alpha_2(p_{11} - p_{12}) + \alpha_3(p_{11} - p_{22}) & c_i^* = c_j^* = 1 \\ \alpha_3(p_{12} - p_{22}) - \alpha_1(p_{11} - p_{12}) & \{c_i^* = 1, c_j^* = 0\} \cup \{c_i^* = 0, c_j^* = 1\} \\ -\alpha_1(p_{11} - p_{22}) - \alpha_2(p_{12} - p_{22}) & c_i^* = c_j^* = 0 \end{cases}.$$

Now consider the pairs  $(i, j)$  such that  $i < j$  and  $(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) \neq 0$ . This can only happen when the pair  $(i, j)$  belongs to one of the following three sets:

- $S_1 = \{i < j\} \cap \{c_i = 0\} \cap \{c_j = 0\} \cap \{c_i^* = c_j^* = 1\}$
- $S_2 = \{i < j\} \cap \{c_i = 0\} \cap \{c_j = 0\} \cap \{\{c_i^* = 1, c_j^* = 0\} \cup \{c_i^* = 0, c_j^* = 1\}\}$
- $S_3 = \{i < j\} \cap \{c_i^* = 0\} \cap \{c_j^* = 0\} \cap \{\{c_i = 1\} \cup \{c_j = 1\}\}$

Therefore,

$$(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) = \begin{cases} \alpha_2(p_{11} - p_{12}) + \alpha_3(p_{11} - p_{22}) & (i, j) \in S_1 \\ \alpha_3(p_{12} - p_{22}) - \alpha_1(p_{11} - p_{12}) & (i, j) \in S_2 \\ \alpha_1(p_{11} - p_{22}) + \alpha_2(p_{12} - p_{22}) & (i, j) \in S_3 \\ 0 & \text{otherwise} \end{cases}.$$

Since  $\alpha_3 = O(1)$  but  $\alpha_1 = O(\alpha_n^2) = o(1)$ , we have  $\alpha_3(p_{12} - p_{22}) - \alpha_1(p_{11} - p_{12}) > 0$ , which means that  $(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) > 0$  for  $(i, j) \in S_1 \cup S_2 \cup S_3$ . Therefore  $\sum_{i < j} (P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) > 0$  for any  $\mathbf{c} \neq \mathbf{c}^*$ , which proves that  $\rho(P_n, \mathbf{c}^*) > \rho(P_n, \mathbf{c})$  for all  $\mathbf{c} \neq \mathbf{c}^*$ .

Next, to precisely characterize  $\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c})$ , we need to count the number of pairs  $(i, j)$  in the sets  $S_1$ ,  $S_2$ , and  $S_3$ . From the definition of  $\xi_n(\mathbf{c})$ , there are  $n\xi_n(\mathbf{c})$  disagreements between  $\mathbf{c}$  and  $\mathbf{c}^*$ . Let  $m$  be the number of vertices such that  $c_i^* = 1$  and  $c_i = 0$ . Since  $\mathbf{c}$  and  $\mathbf{c}^*$  have the same core size,

$$m = \frac{n\xi_n(\mathbf{c})}{2}.$$

Without loss of generality, suppose  $c_i^* = 1$  for  $i = 1, \dots, k$  and  $c_i^* = 0$  for  $i = k+1, \dots, n$ . Then there are  $m$  indices,  $1 \leq i_1 < \dots < i_m \leq k$ , such that  $c_{i_1}^* = 1$  and  $c_{i_1} = 0$  for  $i = i_1, \dots, i_m$ . Fix  $i = i_1$  and consider the pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_1$ . This means  $c_j = 0$  and  $c_j^* = 1$ , which happens for  $m$  distinct values of  $j$ . But we should only count the pairs where  $i < j$ , which means we should exclude  $j = i_1$ , giving us  $(m-1)$  pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_1$ . Next, fix  $i = i_2$ . The same argument plays out, except now we should exclude  $j = i_1$  and  $j = i_2$  to ensure that  $i < j$ . Therefore there are  $(m-2)$  pairs of the form  $(i_2, j)$  such that  $(i_2, j) \in S_1$ . Proceeding in this manner through  $i_1, \dots, i_m$ , the total number of  $(i, j)$  pairs such that  $(i, j) \in S_1$  is given by

$$(m-1) + (m-2) + \dots + 0 = \frac{m(m-1)}{2}.$$

Next, fix  $i = i_1$  and consider the pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_2$ . Since  $c_i^* = 1$  for  $i = 1, \dots, k$  and  $c_i^* = 0$  for  $i = k+1, \dots, n$ , we do not need to consider the set  $\{c_i^* = 0, c_j^* = 1\}$  because  $i$  is always greater than  $j$  in this set. This means that  $c_j = 0$  and  $c_j^* = 0$ , the vertex  $j$  is a periphery vertex that is not misclassified. This happens for  $(n-k-m)$  distinct values of  $j$ . Proceeding in this manner through  $i_1, \dots, i_m$ , the total number of  $(i, j)$  pairs such that  $(i, j) \in S_2$  is given by

$$m(n-k-m).$$

Finally, consider the pairs of the form  $(i, j)$  such that  $(i, j) \in S_3$ . The positive and negative entries of  $\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})$  must cancel out, which means that the size of  $S_3$  is the

same as the combined size of  $S_1$  and  $S_2$ . Therefore, the number of  $(i, j)$  pairs such that  $(i, j) \in S_3$  is given by

$$(n - k - 1) + (n - k - 2) + \dots + (n - k - m) = m \left( n - k - \frac{m + 1}{2} \right).$$

$$(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) = \begin{cases} \alpha_2(p_{11} - p_{12}) + \alpha_3(p_{11} - p_{22}) & (i, j) \in S_1 \\ \alpha_3(p_{12} - p_{22}) - \alpha_1(p_{11} - p_{12}) & (i, j) \in S_2 \\ \alpha_1(p_{11} - p_{22}) + \alpha_2(p_{12} - p_{22}) & (i, j) \in S_3 \\ 0 & \text{otherwise} \end{cases}.$$

Therefore,

$$\begin{aligned} \sum_{i < j} (P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) &= \sum_{\{i, j\} \in S_1} (P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) \\ &\quad + \sum_{\{i, j\} \in S_2} (P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) \\ &\quad + \sum_{\{i, j\} \in S_3} (P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) \\ &= \frac{m(m-1)}{2} [\alpha_2(p_{11} - p_{12}) + \alpha_3(p_{11} - p_{22})] \\ &\quad + m(n - k - m) [\alpha_3(p_{12} - p_{22}) - \alpha_1(p_{11} - p_{12})] \\ &\quad + m \left( n - k - \frac{m+1}{2} \right) [\alpha_1(p_{11} - p_{22}) + \alpha_2(p_{12} - p_{22})] \\ &= \frac{n^2 \xi_n(\mathbf{c})}{2} (p_{12} - p_{22})(1 + o(1)) \end{aligned}$$

since  $\alpha_1 = O(\alpha_n^2)$ ,  $\alpha_2 = O(\alpha_n)$ , and  $m = n\xi_n(\mathbf{c})/2$ . Putting the numerator and denominator together,

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) = \frac{\varrho_n(p_{12} - p_{22}) n^2 \xi_n(\mathbf{c})}{2 \binom{n}{2} \sqrt{2\varrho_n \alpha_n p_{22}(1 - \varrho_n p_{22})}} (1 + o(1)) = \frac{p_{12} - p_{22}}{\sqrt{p_{22}}} \sqrt{\frac{\varrho_n}{2\alpha_n(1 - \varrho_n p_{22})}} \xi_n(\mathbf{c}) (1 + o(1)).$$

This completes the proof.  $\square$

Comment: Lemma 2 is a deterministic result about the population parameter that does not involve any probabilistic statements. The first part of the lemma states that the population parameter is uniquely maximized at the true labels. The second part characterizes the gap between the true labels and any other set of labels (with the same core size) with respect to the corresponding values of the population parameter. The proof is based on a counting argument that quantifies a precise connection between the error rate,  $\xi_n(\mathbf{c})$ , and the corresponding difference in the population parameter,  $\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c})$ .

### Lemma 3.

LEMMA 3. Let  $P_n$  be a Chung-Lu model where  $P_n = \varrho_n P$  and  $P_{ij} = \theta_i \theta_j$ , where  $\theta_1, \dots, \theta_n$  are the degree parameters. Let  $\theta_{(1)} \geq \theta_{(2)} \geq \dots \geq \theta_{(k)} \geq \theta_{(k+1)} \geq \dots \geq \theta_{(n-1)} \geq \theta_{(n)}$  be the ordered values of the degree parameters. Define the “true” core-periphery labels as  $c_i^* = 1$  for the  $k$  vertices with the largest  $\theta_i$  values and  $c_i^* = 0$  for the remaining  $(n - k)$  vertices. Then for any  $\mathbf{c} \neq \mathbf{c}^*$  such that the size of the core is the same for  $\mathbf{c}$  and  $\mathbf{c}^*$ , and under the assumption that  $\theta_{(k)}\theta_{(n)} > \theta_{(k+1)}\theta_{(k+2)}$ , we have  $\rho(P_n, \mathbf{c}) < \rho(P_n, \mathbf{c}^*)$ , and

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) \geq \xi_n(\mathbf{c}) \sqrt{\frac{\varrho_n}{2\alpha_n} \frac{\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)}}{\bar{\theta}}} (1 + o(1)).$$

*Proof.* We broadly follow the arguments from the proof of Lemma 2. Consider

$$\rho(P_n, \mathbf{c}) = \frac{\sum_{i < j} \varrho_n(P_{ij} - \bar{p})(\Delta_{ij} - \bar{\Delta})}{\sqrt{\binom{n}{2} \varrho_n \bar{p}(1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta}(1 - \bar{\Delta})}}.$$

Now, let  $\Delta(\mathbf{c}^*)$  and  $\Delta(\mathbf{c})$  denote the  $\Delta$  matrices corresponding to the labels  $\mathbf{c}^*$  and  $\mathbf{c}$ , respectively. Since  $\mathbf{c}^*$  and  $\mathbf{c}$  have the same core size, we have

$$1 - \bar{\Delta}(\mathbf{c}^*) = 1 - \bar{\Delta}(\mathbf{c}) = \alpha_3 = \frac{(n - k)(n - k - 1)}{n(n - 1)} = 1 + o(1)$$

as  $k = o(n)$  from A1, and

$$\bar{\Delta}(\mathbf{c}^*) = \bar{\Delta}(\mathbf{c}) = 1 - \alpha_3 = \frac{2nk - (k^2 + k)}{n(n - 1)} = 2\alpha_n(1 + o(1)).$$

Next, we have

$$\bar{p} = \frac{1}{\binom{n}{2}} \sum_{i < j} \theta_i \theta_j \sim \bar{\theta}^2.$$

Therefore, the denominator of  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$  is given by

$$\sqrt{\binom{n}{2} \varrho_n \bar{p}(1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta}(1 - \bar{\Delta})} = \binom{n}{2} \sqrt{2\varrho_n \alpha_n \bar{\theta}^2 (1 - \varrho_n \bar{\theta}^2)} (1 + o(1)).$$

Now consider the numerators of  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$  and note that  $\Delta_{ij}(\cdot)$  is the only term that changes between  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$ . Therefore,

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) = \frac{\sum_{i < j} \varrho_n(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}))}{\binom{n}{2} \sqrt{2\varrho_n \alpha_n \bar{\theta}^2 (1 - \varrho_n \bar{\theta}^2)}} (1 + o(1)).$$

Note that

$$\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}) = \begin{cases} +1 & \{c_i = 0\} \cap \{c_j = 0\} \cap \{\{c_i^* = 1\} \cup \{c_j^* = 1\}\} \\ -1 & \{c_i^* = 0\} \cap \{c_j^* = 0\} \cap \{\{c_i = 1\} \cup \{c_j = 1\}\} \\ 0 & \text{otherwise} \end{cases}.$$

Also,

$$P_{ij} - \bar{p} = \theta_i \theta_j - \bar{\theta}^2.$$

Now consider the pairs  $(i, j)$  such that  $i < j$  and  $(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) \neq 0$ . This can only happen when the pair  $(i, j)$  belongs to one of the following three sets:

- $S_1 = \{i < j\} \cap \{c_i = 0\} \cap \{c_j = 0\} \cap \{c_i^* = c_j^* = 1\}$
- $S_2 = \{i < j\} \cap \{c_i = 0\} \cap \{c_j = 0\} \cap \{\{c_i^* = 1, c_j^* = 0\} \cup \{c_i^* = 0, c_j^* = 1\}\}$
- $S_3 = \{i < j\} \cap \{c_i^* = 0\} \cap \{c_j^* = 0\} \cap \{\{c_i = 1\} \cup \{c_j = 1\}\}$

Next, to precisely characterize  $\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c})$ , we need to count the number of pairs  $(i, j)$  in the sets  $S_1$ ,  $S_2$ , and  $S_3$ . From the definition of  $\xi_n(\mathbf{c})$ , there are  $n\xi_n(\mathbf{c})$  disagreements between  $\mathbf{c}$  and  $\mathbf{c}^*$ . Let  $m$  be the number of vertices such that  $c_i^* = 1$  and  $c_i = 0$ . Since  $\mathbf{c}$  and  $\mathbf{c}^*$  have the same core size,

$$m = \frac{n\xi_n(\mathbf{c})}{2}.$$

Without loss of generality, suppose  $c_i^* = 1$  for  $i = 1, \dots, k$  and  $c_i^* = 0$  for  $i = k+1, \dots, n$ . Then there are  $m$  indices,  $1 \leq i_1 < \dots < i_m \leq k$ , such that  $c_{i_1}^* = 1$  and  $c_{i_1} = 0$  for  $i = i_1, \dots, i_m$ . Fix  $i = i_1$  and consider the pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_1$ . This means  $c_j = 0$  and  $c_j^* = 1$ , which happens for  $m$  distinct values of  $j$ . But we should only count the pairs where  $i < j$ , which means we should exclude  $j = i_1$ , giving us  $(m-1)$  pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_1$ . Next, fix  $i = i_2$ . The same argument plays out, except now we should exclude  $j = i_1$  and  $j = i_2$  to ensure that  $i < j$ . Therefore there are  $(m-2)$  pairs of the form  $(i_2, j)$  such that  $(i_2, j) \in S_1$ . Proceeding in this manner through  $i_1, \dots, i_m$ , the total number of  $(i, j)$  pairs such that  $(i, j) \in S_1$  is given by

$$(m-1) + (m-2) + \dots + 0 = \frac{m(m-1)}{2}.$$

Next, fix  $i = i_1$  and consider the pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_2$ . Since  $c_i^* = 1$  for  $i = 1, \dots, k$  and  $c_i^* = 0$  for  $i = k+1, \dots, n$ , we do not need to consider the set  $\{c_i^* = 0, c_j^* = 1\}$  because  $i$  is always greater than  $j$  in this set. This means that  $c_j = 0$  and  $c_j^* = 0$ , the vertex  $j$  is a periphery vertex that is not misclassified. This happens for  $(n-k-m)$  distinct values of  $j$ . Proceeding in this manner through  $i_1, \dots, i_m$ , the total number of  $(i, j)$  pairs such that  $(i, j) \in S_2$  is given by

$$m(n-k-m).$$

Finally, consider the pairs of the form  $(i, j)$  such that  $(i, j) \in S_3$ . The positive and negative entries of  $\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})$  must cancel out, which means that the size of  $S_3$  is the same as the combined size of  $S_1$  and  $S_2$ . Therefore, the number of  $(i, j)$  pairs such that  $(i, j) \in S_3$  is given by

$$(n-k-1) + (n-k-2) + \dots + (n-k-m) = m \left( n - k - \frac{m+1}{2} \right).$$

Note that  $\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}) = 1$  when  $(i, j) \in S_1 \cup S_2$  and  $\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}) = -1$  when  $(i, j) \in S_3$ . Also, the number of  $(i, j)$  pairs in  $S_1 \cup S_2$  is the same as the number of  $(i, j)$  pairs in  $S_3$ . Finally for any  $(i, j) \in S_1 \cup S_2$  we have  $P_{ij} \geq \theta_{(k)}\theta_{(n)}$  and for any  $(i, j) \in S_3$  we have  $P_{ij} \leq \theta_{(k+1)}\theta_{(k+2)}$ . Therefore,

$$\begin{aligned} \sum_{i < j} (P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) &= \sum_{\{i,j\} \in S_1 \cup S_2} (P_{ij} - \bar{p}) - \sum_{\{i,j\} \in S_3} (P_{ij} - \bar{p}) \\ &= \sum_{\{i,j\} \in S_1 \cup S_2} P_{ij} - \sum_{\{i,j\} \in S_3} P_{ij} \\ &\geq m(n - k - m) (\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)}) \\ &= \frac{n^2 \xi_n(\mathbf{c})}{2} (\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)}) (1 + o(1)). \end{aligned}$$

Putting the numerator and denominator together,

$$\begin{aligned} \rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) &\geq \frac{\varrho_n(\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)})n^2\xi_n(\mathbf{c})}{2\binom{n}{2}\sqrt{2\varrho_n\alpha_n\bar{\theta}^2(1-\varrho_n\bar{\theta}^2)}}(1+o(1)) \\ &= \frac{\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)}}{\bar{\theta}} \sqrt{\frac{\varrho_n}{2\alpha_n(1-\varrho_n\bar{\theta}^2)}} \xi_n(\mathbf{c})(1+o(1)). \end{aligned}$$

This completes the proof.  $\square$

Comment: The extra assumption  $\theta_{(k)}\theta_{(n)} > \theta_{(k+1)}\theta_{(k+2)}$  ensures that the product of the smallest degree parameters in the core and periphery is larger than the product of the two largest degree parameters in the periphery. Moreover, notice that if  $\theta_{(1)} = \dots = \theta_{(k)}$  and  $\theta_{(k+1)} = \dots = \theta_{(n)}$ , then the CL model becomes the SBM model and we recover the result in Lemma 1.

#### Lemma 4.

LEMMA 4. Let  $P_n$  be a CP-DCBM model where  $P_n = \varrho_n P$  and  $P_{ij} = \theta_i \Omega_{c_i^* c_j^*} \theta_j$ , where  $\theta_1, \dots, \theta_n$  are the degree parameters, and

$$\Omega = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } p_{11} > p_{12} = p_{21} > p_{22}.$$

Let  $\theta_c = \min\{\theta_i : c_i^* = 1\}$  be the lowest degree parameter in the core set, and  $\theta_{p,max} = \max\{\theta_i : c_i^* = 0\}$  and  $\theta_{p,min} = \min\{\theta_i : c_i^* = 0\}$  be the highest and lowest degree parameters in the periphery set, respectively. Then for any  $\mathbf{c} \neq \mathbf{c}^*$  such that the size of the core is the same for  $\mathbf{c}$  and  $\mathbf{c}^*$ , and under the assumption that

$$\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c \theta_{p,min}},$$

we have  $\rho(P_n, \mathbf{c}) < \rho(P_n, \mathbf{c}^*)$ , and

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) \geq \xi_n(\mathbf{c}) \sqrt{\frac{\varrho_n}{2\alpha_n}} \frac{p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2}{\sqrt{\bar{p}}} (1 + o(1)).$$

*Proof.* We broadly follow the arguments from the proofs of Lemma 2 and Lemma 3. Consider

$$\rho(P_n, \mathbf{c}) = \frac{\sum_{i < j} \varrho_n(P_{ij} - \bar{p})(\Delta_{ij} - \bar{\Delta})}{\sqrt{\binom{n}{2} \varrho_n \bar{p}(1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta}(1 - \bar{\Delta})}}.$$

Now, let  $\Delta(\mathbf{c}^*)$  and  $\Delta(\mathbf{c})$  denote the  $\Delta$  matrices corresponding to the labels  $\mathbf{c}^*$  and  $\mathbf{c}$ , respectively. Since  $\mathbf{c}^*$  and  $\mathbf{c}$  have the same core size, we have

$$1 - \bar{\Delta}(\mathbf{c}^*) = 1 - \bar{\Delta}(\mathbf{c}) = \alpha_3 = \frac{(n-k)(n-k-1)}{n(n-1)} = 1 + o(1)$$

as  $k = o(n)$  from A1, and

$$\bar{\Delta}(\mathbf{c}^*) = \bar{\Delta}(\mathbf{c}) = 1 - \alpha_3 = \frac{2nk - (k^2 + k)}{n(n-1)} = 2\alpha_n(1 + o(1)).$$

Next, we have

$$\bar{p} = \frac{1}{\binom{n}{2}} \sum_{i < j} P_{ij}.$$

Therefore, the denominator of  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$  is given by

$$\sqrt{\binom{n}{2} \varrho_n \bar{p}(1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta}(1 - \bar{\Delta})} = \binom{n}{2} \sqrt{2\varrho_n \alpha_n \bar{p}(1 - \varrho_n \bar{p})} (1 + o(1)).$$

Now consider the numerators of  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$  and note that  $\Delta_{ij}(\cdot)$  is the only term that changes between  $\rho(P_n, \mathbf{c})$  and  $\rho(P_n, \mathbf{c}^*)$ . Therefore,

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) = \frac{\sum_{i < j} \varrho_n(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}))}{\binom{n}{2} \sqrt{2\varrho_n \alpha_n \bar{p}(1 - \varrho_n \bar{p})}} (1 + o(1)).$$

Note that

$$\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}) = \begin{cases} +1 & \{c_i = 0\} \cap \{c_j = 0\} \cap \{\{c_i^* = 1\} \cup \{c_j^* = 1\}\} \\ -1 & \{c_i^* = 0\} \cap \{c_j^* = 0\} \cap \{\{c_i = 1\} \cup \{c_j = 1\}\} \\ 0 & \text{otherwise} \end{cases}.$$

Now consider the pairs  $(i, j)$  such that  $i < j$  and  $(P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) \neq 0$ . This can only happen when the pair  $(i, j)$  belongs to one of the following three sets:

- $S_1 = \{i < j\} \cap \{c_i = 0\} \cap \{c_j = 0\} \cap \{c_i^* = c_j^* = 1\}$
- $S_2 = \{i < j\} \cap \{c_i = 0\} \cap \{c_j = 0\} \cap \{\{c_i^* = 1, c_j^* = 0\} \cup \{c_i^* = 0, c_j^* = 1\}\}$
- $S_3 = \{i < j\} \cap \{c_i^* = 0\} \cap \{c_j^* = 0\} \cap \{\{c_i = 1\} \cup \{c_j = 1\}\}$

Next, to precisely characterize  $\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c})$ , we need to count the number of pairs  $(i, j)$  in the sets  $S_1$ ,  $S_2$ , and  $S_3$ . From the definition of  $\xi_n(\mathbf{c})$ , there are  $n\xi_n(\mathbf{c})$  disagreements between  $\mathbf{c}$  and  $\mathbf{c}^*$ . Let  $m$  be the number of vertices such that  $c_i^* = 1$  and  $c_i = 0$ . Since  $\mathbf{c}$  and  $\mathbf{c}^*$  have the same core size,

$$m = \frac{n\xi_n(\mathbf{c})}{2}.$$

Without loss of generality, suppose  $c_i^* = 1$  for  $i = 1, \dots, k$  and  $c_i^* = 0$  for  $i = k+1, \dots, n$ . Then there are  $m$  indices,  $1 \leq i_1 < \dots < i_m \leq k$ , such that  $c_{i_j}^* = 1$  and  $c_{i_j} = 0$  for  $i = i_1, \dots, i_m$ . Fix  $i = i_1$  and consider the pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_1$ . This means  $c_j = 0$  and  $c_j^* = 1$ , which happens for  $m$  distinct values of  $j$ . But we should only count the pairs where  $i < j$ , which means we should exclude  $j = i_1$ , giving us  $(m-1)$  pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_1$ . Next, fix  $i = i_2$ . The same argument plays out, except now we should exclude  $j = i_1$  and  $j = i_2$  to ensure that  $i < j$ . Therefore there are  $(m-2)$  pairs of the form  $(i_2, j)$  such that  $(i_2, j) \in S_1$ . Proceeding in this manner through  $i_1, \dots, i_m$ , the total number of  $(i, j)$  pairs such that  $(i, j) \in S_1$  is given by

$$(m-1) + (m-2) + \dots + 0 = \frac{m(m-1)}{2}.$$

Next, fix  $i = i_1$  and consider the pairs of the form  $(i_1, j)$  such that  $(i_1, j) \in S_2$ . Since  $c_i^* = 1$  for  $i = 1, \dots, k$  and  $c_i^* = 0$  for  $i = k+1, \dots, n$ , we do not need to consider the set  $\{c_i^* = 0, c_j^* = 1\}$  because  $i$  is always greater than  $j$  in this set. This means that  $c_j = 0$  and  $c_j^* = 0$ , the vertex  $j$  is a periphery vertex that is not misclassified. This happens for  $(n-k-m)$  distinct values of  $j$ . Proceeding in this manner through  $i_1, \dots, i_m$ , the total number of  $(i, j)$  pairs such that  $(i, j) \in S_2$  is given by

$$m(n-k-m).$$

Finally, consider the pairs of the form  $(i, j)$  such that  $(i, j) \in S_3$ . The positive and negative entries of  $\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})$  must cancel out, which means that the size of  $S_3$  is the same as the combined size of  $S_1$  and  $S_2$ . Therefore, the number of  $(i, j)$  pairs such that  $(i, j) \in S_3$  is given by

$$(n-k-1) + (n-k-2) + \dots + (n-k-m) = m \left( n - k - \frac{m+1}{2} \right).$$

Note that  $\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}) = 1$  when  $(i, j) \in S_1 \cup S_2$  and  $\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c}) = -1$  when  $(i, j) \in S_3$ . Also, the number of  $(i, j)$  pairs in  $S_1 \cup S_2$  is the same as the number of  $(i, j)$  pairs in  $S_3$ . Finally for any  $(i, j) \in S_1 \cup S_2$  we have  $P_{ij} \geq p_{12}\theta_c\theta_{p,min}$  and for any  $(i, j) \in S_3$  we have  $P_{ij} \leq p_{22}\theta_{p,max}^2$ . Therefore,

$$\begin{aligned} \sum_{i < j} (P_{ij} - \bar{p})(\Delta_{ij}(\mathbf{c}^*) - \Delta_{ij}(\mathbf{c})) &= \sum_{\{i,j\} \in S_1 \cup S_2} (P_{ij} - \bar{p}) - \sum_{\{i,j\} \in S_3} (P_{ij} - \bar{p}) \\ &= \sum_{\{i,j\} \in S_1 \cup S_2} P_{ij} - \sum_{\{i,j\} \in S_3} P_{ij} \\ &\geq m(n-k-m) (p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2) \\ &= \frac{n^2\xi_n(\mathbf{c})}{2} (p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2) (1 + o(1)). \end{aligned}$$

Putting the numerator and denominator together,

$$\begin{aligned}\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) &\geq \frac{\varrho_n(p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2)n^2\xi_n(\mathbf{c})}{2\binom{n}{2}\sqrt{2\varrho_n\alpha_n\bar{p}(1-\varrho_n\bar{p})}}(1+o(1)) \\ &= \frac{p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2}{\sqrt{\bar{p}}}\sqrt{\frac{\varrho_n}{2\alpha_n(1-\varrho_n\bar{p})}}\xi_n(\mathbf{c})(1+o(1)).\end{aligned}$$

This completes the proof.  $\square$

Comment: Notice that  $p_{12}\theta_c\theta_{p,min}$  is the minimum core-periphery edge probability, and  $p_{22}\theta_{p,max}^2$  is the largest periphery-periphery edge probability. Thus, this extra assumption ensures that all core-periphery edge probabilities are larger than all periphery-periphery edge probabilities, as sensible requirement for a CP model. Also, this condition reduces to  $p_{12} > p_{22}$  in the SBM case ( $\theta_1 = \dots = \theta_n$ ) and  $\theta_c\theta_{p,min} > \theta_{p,max}^2$  in the CL case ( $p_{12} = p_{22}$ ), recovering our previous results.

## Proof of Theorem 2.1

### SBM

Let  $P_n$  be a two-block SBM where  $P_n = \varrho_n P$  and  $P_{ij} = \Omega_{c_i^* c_j^*}$  where

$$\Omega = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } p_{11} > p_{12} = p_{21} > p_{22}.$$

Under A1, we have

$$\mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\sqrt{p_{22}}\alpha_n}{p_{12}-p_{22}}\sqrt{\frac{(8+\eta)\log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) \rightarrow 1 \text{ for any } \eta > 0 \text{ as } n \rightarrow \infty.$$

Furthermore, the optimal labels are weakly consistent under A1, i.e., for any  $\delta > 0$ ,

$$\mathbb{P}(\xi_n(\hat{\mathbf{c}}) > \delta) \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (12)$$

and under the additional assumption that  $\alpha_n/\sqrt{\varrho_n} = o(1/\sqrt{n\log n})$ , the labels are strongly consistent, i.e.,

$$\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) \rightarrow 1 \text{ as } n \rightarrow \infty. \quad (13)$$

*Proof:* Recall that

$$\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} T(A_n, \mathbf{c}),$$

which means that  $T(A_n, \hat{\mathbf{c}}) \geq T(A_n, \mathbf{c}^*)$ . Consider the event

$$E = \{T(A_n, \hat{\mathbf{c}}) \leq \rho(P_n, \hat{\mathbf{c}}) + \varepsilon_n\} \cap \{T(A_n, \mathbf{c}^*) \geq \rho(P_n, \mathbf{c}^*) - \varepsilon_n\}$$

where  $\varepsilon_n = \alpha_n^{1/2}\varrho_n^{1/2}\sqrt{\frac{\log(n\varrho_n\alpha_n)}{n\varrho_n\alpha_n}}$ . Then from Lemma 1, we have  $\mathbb{P}(E) \rightarrow 1$  as  $n \rightarrow \infty$ . Note that  $E$  implies that  $\rho(P_n, \mathbf{c}^*) - \rho(P_n, \hat{\mathbf{c}}) < 2\varepsilon_n$ . But, from Lemma 2, we know that

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \hat{\mathbf{c}}) = \frac{p_{12} - p_{22}}{\sqrt{p_{22}}}\sqrt{\frac{\varrho_n}{2\alpha_n(1-\varrho_n p_{22})}}\xi_n(\mathbf{c})(1+o(1)).$$

This means that  $E$  can happen only if

$$\frac{p_{12} - p_{22}}{\sqrt{p_{22}}} \sqrt{\frac{\varrho_n}{2\alpha_n(1 - \varrho_n p_{22})}} \xi_n(\mathbf{c}) < 2\varepsilon_n(1 + o(1)) \Rightarrow \xi_n(\hat{\mathbf{c}}) < \frac{\sqrt{p_{22}}\alpha_n}{p_{12} - p_{22}} \sqrt{\frac{8(1 - \varrho_n b) \log(n\varrho_n\alpha_n)}{n\varrho_n}}(1 + o(1)).$$

Therefore, for any  $\eta > 0$

$$\mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\sqrt{p_{22}}\alpha_n}{p_{12} - p_{22}} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

**Weak consistency:** Under A1, as  $n \rightarrow \infty$ ,

$$n\varrho_n\alpha_n \rightarrow \infty \Rightarrow \frac{\log(n\varrho_n\alpha_n)}{n\varrho_n\alpha_n} \rightarrow 0 \Rightarrow \frac{\sqrt{p_{22}}\alpha_n}{p_{12} - p_{22}} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}} \rightarrow 0.$$

Now fix any  $\delta > 0$ . There is some  $n_0$  such that

$$\delta > \frac{\sqrt{p_{22}}\alpha_n}{p_{12} - p_{22}} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}}$$

for all  $n > n_0$ . Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\xi_n(\hat{\mathbf{c}}) < \delta) \geq \lim_{n \rightarrow \infty} \mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\sqrt{p_{22}}\alpha_n}{p_{12} - p_{22}} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) = 1.$$

**Strong consistency:** Suppose that

$$\frac{\alpha_n}{\sqrt{\varrho_n}} = o\left(\frac{1}{\sqrt{n \log n}}\right).$$

Then

$$\frac{\sqrt{p_{22}}\alpha_n}{p_{12} - p_{22}} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}} = o\left(\sqrt{\frac{\log(n\varrho_n\alpha_n)}{n^2 \log n}}\right) = o\left(\frac{1}{n}\right).$$

Therefore,

$$\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) = \mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{1}{n}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

This completes the proof.  $\square$

## Chung-Lu

Let  $P_n$  be a Chung-Lu model where  $P_n = \varrho_n P$  and  $P_{ij} = \theta_i \theta_j$ , where  $\theta_1, \dots, \theta_n$  are the degree parameters. Recall the definition of  $\mathbf{c}^*$  from Lemma 3. Under A1 and the assumption in Lemma 3, we have

$$\mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\bar{\theta}\alpha_n}{\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)}} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) \rightarrow 1 \text{ for any } \eta > 0 \text{ as } n \rightarrow \infty.$$

Furthermore, the optimal labels are weakly consistent under A1 and the assumption in Lemma 3, i.e., for any  $\delta > 0$ ,

$$\mathbb{P}(\xi_n(\hat{\mathbf{c}}) > \delta) \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (14)$$

and under the additional assumption that  $\alpha_n/\sqrt{\varrho_n} = o(1/\sqrt{n \log n})$ , the optimal labels are strongly consistent, i.e.,

$$\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) \rightarrow 1 \text{ as } n \rightarrow \infty. \quad (15)$$

*Proof:* We broadly follow the arguments from the proofs of Theorem 2.1. Recall that

$$\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} T(A_n, \mathbf{c}),$$

which means that  $T(A_n, \hat{\mathbf{c}}) \geq T(A_n, \mathbf{c}^*)$ . Consider the event

$$E = \{T(A_n, \hat{\mathbf{c}}) \leq \rho(P_n, \hat{\mathbf{c}}) + \varepsilon_n\} \cap \{T(A_n, \mathbf{c}^*) \geq \rho(P_n, \mathbf{c}^*) - \varepsilon_n\}$$

where  $\varepsilon_n = \alpha_n \varrho_n^{1/2} \sqrt{\frac{\log(n \varrho_n \alpha_n)}{n \varrho_n \alpha_n}}$ . Then from Lemma 1, we have  $\mathbb{P}(E) \rightarrow 1$  as  $n \rightarrow \infty$ . Note that  $E$  implies that  $\rho(P_n, \mathbf{c}^*) - \rho(P_n, \hat{\mathbf{c}}) < 2\varepsilon_n$ . But, from Lemma 3, we know that

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) \geq \xi_n(\mathbf{c}) \sqrt{\frac{\varrho_n}{2\alpha_n}} \frac{\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)}}{\bar{\theta}} (1 + o(1)).$$

Let

$$\beta = \frac{\theta_{(k)}\theta_{(n)} - \theta_{(k+1)}\theta_{(k+2)}}{\bar{\theta}}.$$

Then  $E$  can happen only if

$$\beta \sqrt{\frac{\varrho_n}{2\alpha_n}} \xi_n(\mathbf{c}) < 2\varepsilon_n (1 + o(1)) \Rightarrow \xi_n(\hat{\mathbf{c}}) < \frac{\alpha_n}{\beta} \sqrt{\frac{8 \log(n \varrho_n \alpha_n)}{n \varrho_n}} (1 + o(1)).$$

Therefore, for any  $\eta > 0$

$$\mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n \varrho_n \alpha_n)}{n \varrho_n}}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

**Weak consistency:** Under A1, as  $n \rightarrow \infty$ ,

$$n \varrho_n \alpha_n \rightarrow \infty \Rightarrow \frac{\log(n \varrho_n \alpha_n)}{n \varrho_n \alpha_n} \rightarrow 0 \Rightarrow \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n \varrho_n \alpha_n)}{n \varrho_n}} \rightarrow 0.$$

Now fix any  $\delta > 0$ . There is some  $n_0$  such that

$$\delta > \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n \varrho_n \alpha_n)}{n \varrho_n}}$$

for all  $n > n_0$ . Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\xi_n(\hat{\mathbf{c}}) < \delta) \geq \lim_{n \rightarrow \infty} \mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) = 1.$$

**Strong consistency:** Suppose that

$$\frac{\alpha_n}{\sqrt{\varrho_n}} = o\left(\frac{1}{\sqrt{n \log n}}\right).$$

Then

$$\frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}} = o\left(\sqrt{\frac{\log(n\varrho_n\alpha_n)}{n^2 \log n}}\right) = o\left(\frac{1}{n}\right).$$

Therefore,

$$\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) = \mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{1}{n}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

This completes the proof.  $\square$

Let  $P_n$  be a CP-DCBM model where  $P_n = \varrho_n P$  and  $P_{ij} = \theta_i \Omega_{c_i^* c_j^*} \theta_j$ , where  $\theta_1, \dots, \theta_n$  are the degree parameters, and

$$\Omega = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } p_{11} > p_{12} = p_{21} > p_{22}.$$

Let  $\theta_c = \min\{\theta_i : c_i^* = 1\}$  be the lowest degree parameter in the core set, and  $\theta_{p,max} = \max\{\theta_i : c_i^* = 0\}$  and  $\theta_{p,min} = \min\{\theta_i : c_i^* = 0\}$  be the highest and lowest degree parameters in the periphery set, respectively. Then for any  $\mathbf{c} \neq \mathbf{c}^*$  such that the size of the core is the same for  $\mathbf{c}$  and  $\mathbf{c}^*$ , and under the assumption that

$$\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c \theta_{p,min}},$$

we have  $\rho(P_n, \mathbf{c}) < \rho(P_n, \mathbf{c}^*)$ , and

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) \geq \xi_n(\mathbf{c}) \sqrt{\frac{\varrho_n}{2\alpha_n}} \frac{p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2}{\sqrt{\bar{p}}} (1 + o(1)).$$

## DCBM

Let  $P_n$  be a CP-DCBM model where  $P_n = \varrho_n P$  and  $P_{ij} = \theta_i \Omega_{c_i^* c_j^*} \theta_j$ , where  $\theta_1, \dots, \theta_n$  are the degree parameters, and

$$\Omega = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } p_{11} > p_{12} = p_{21} > p_{22}.$$

Under A1 and the assumption in Lemma 4, we have

$$\mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\sqrt{\bar{p}}\alpha_n}{p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) \rightarrow 1 \text{ for any } \eta > 0 \text{ as } n \rightarrow \infty.$$

Furthermore, the optimal labels are weakly consistent under A1 and the assumption in Lemma 3, i.e., for any  $\delta > 0$ ,

$$\mathbb{P}(\xi_n(\hat{\mathbf{c}}) > \delta) \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (16)$$

and under the additional assumption that  $\alpha_n/\sqrt{\varrho_n} = o(1/\sqrt{n \log n})$ , the optimal labels are strongly consistent, i.e.,

$$\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) \rightarrow 1 \text{ as } n \rightarrow \infty. \quad (17)$$

*Proof:* We broadly follow the arguments from the proofs of Theorem 2.1. Recall that

$$\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} T(A_n, \mathbf{c}),$$

which means that  $T(A_n, \hat{\mathbf{c}}) \geq T(A_n, \mathbf{c}^*)$ . Consider the event

$$E = \{T(A_n, \hat{\mathbf{c}}) \leq \rho(P_n, \hat{\mathbf{c}}) + \varepsilon_n\} \cap \{T(A_n, \mathbf{c}^*) \geq \rho(P_n, \mathbf{c}^*) - \varepsilon_n\}$$

where  $\varepsilon_n = \alpha_n \varrho_n^{1/2} \sqrt{\frac{\log(n \varrho_n \alpha_n)}{n \varrho_n \alpha_n}}$ . Then from Lemma 1, we have  $\mathbb{P}(E) \rightarrow 1$  as  $n \rightarrow \infty$ . Note that  $E$  implies that  $\rho(P_n, \mathbf{c}^*) - \rho(P_n, \hat{\mathbf{c}}) < 2\varepsilon_n$ . But, from Lemma 4, we know that

$$\rho(P_n, \mathbf{c}^*) - \rho(P_n, \mathbf{c}) \geq \xi_n(\mathbf{c}) \sqrt{\frac{\varrho_n}{2\alpha_n}} \frac{p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2}{\sqrt{p}} (1 + o(1)).$$

Let

$$\beta = \frac{p_{12}\theta_c\theta_{p,min} - p_{22}\theta_{p,max}^2}{\sqrt{p}}.$$

Then  $E$  can happen only if

$$\beta \sqrt{\frac{\varrho_n}{2\alpha_n}} \xi_n(\mathbf{c}) < 2\varepsilon_n (1 + o(1)) \Rightarrow \xi_n(\hat{\mathbf{c}}) < \frac{\alpha_n}{\beta} \sqrt{\frac{8 \log(n \varrho_n \alpha_n)}{n \varrho_n}} (1 + o(1)).$$

Therefore, for any  $\eta > 0$

$$\mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n \varrho_n \alpha_n)}{n \varrho_n}}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

**Weak consistency:** Under A1, as  $n \rightarrow \infty$ ,

$$n \varrho_n \alpha_n \rightarrow \infty \Rightarrow \frac{\log(n \varrho_n \alpha_n)}{n \varrho_n \alpha_n} \rightarrow 0 \Rightarrow \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n \varrho_n \alpha_n)}{n \varrho_n}} \rightarrow 0.$$

Now fix any  $\delta > 0$ . There is some  $n_0$  such that

$$\delta > \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n \varrho_n \alpha_n)}{n \varrho_n}}$$

for all  $n > n_0$ . Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\xi_n(\hat{\mathbf{c}}) < \delta) \geq \lim_{n \rightarrow \infty} \mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}}\right) = 1.$$

**Strong consistency:** Suppose that

$$\frac{\alpha_n}{\sqrt{\varrho_n}} = o\left(\frac{1}{\sqrt{n \log n}}\right).$$

Then

$$\frac{\alpha_n}{\beta} \sqrt{\frac{(8 + \eta) \log(n\varrho_n\alpha_n)}{n\varrho_n}} = o\left(\sqrt{\frac{\log(n\varrho_n\alpha_n)}{n^2 \log n}}\right) = o\left(\frac{1}{n}\right).$$

Therefore,

$$\mathbb{P}(\hat{\mathbf{c}} = \mathbf{c}^*) = \mathbb{P}\left(\xi_n(\hat{\mathbf{c}}) < \frac{1}{n}\right) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

This completes the proof.  $\square$

## 12 Technical proofs for hypothesis tests

### Proof of Theorem 3.1

**THEOREM 3.1.** *Under  $H_0$ , i.e., when the network is generated from the ER model, the Type I error rate converges to 0, i.e., for any  $\eta > 0$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] < \eta$ , where the rejection thresholds are given by*

$$C_1 = \sqrt{\frac{\log(n\varrho_n\alpha_n)}{n}} \text{ and } C_2 = \frac{2\sqrt{2}\varrho_n \log(n)}{\sqrt{k}}. \quad (18)$$

*Under the assumption that  $\alpha_n \geq \frac{(\log(n \log n))^2}{n}$ , suppose that any of the following alternative hypotheses hold:*

- $H_1 : P$  is a CP-SBM where  $p_{11} > p_{12} = p_{21} > p_{22}$
- $H_1 : P$  is a Chung-Lu model where  $\theta_{(k)}\theta_{(n)} > \theta_{(k+1)}\theta_{(k+2)}$  and  $\theta_{(k-1)} > \theta_{(k+1)}$ .
- $H_1 : P$  is a CP-DCBM where  $\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c\theta_{p,min}}$  and  $\frac{p_{11}}{p_{12}} > \frac{\theta_{p,max}}{\theta_c}$ .

*Under each of these alternatives, the power of the test goes to one, i.e., for any  $\eta > 0$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] > 1 - \eta.$$

#### Under $H_0$ :

Under the null, i.e., when the network is generated from the ER model, we want to prove that

$$\mathbb{P}[T_1 > C_1] \rightarrow 0 \text{ and } \mathbb{P}[T_2 > C_2] \rightarrow 0.$$

First consider  $T_1(A) = T(A, \hat{\mathbf{c}})$ . Recall that from Lemma 1,

$$\mathbb{P}[T(A, \hat{\mathbf{c}}) - \rho(P_n, \hat{\mathbf{c}}) > \varepsilon_n] \rightarrow 0$$

Note that  $\rho(P_n, \mathbf{c}) = 0$  for all  $\mathbf{c}$ , which means  $\rho(P_n, \hat{\mathbf{c}}) = 0$ . Also note that  $C_1 = \varepsilon_n$ , which means

$$\mathbb{P}[T_1 > C_1] = \mathbb{P}[T_1 > \varepsilon_n] \rightarrow 0.$$

Note that we do not need to analyze the second part of the intersection test, since the first part is sufficient to prove that type-1 error converges to zero.

#### Under $H_1$ :

Under the alternative, we want to prove that

$$\mathbb{P}[T_1 > C_1] \rightarrow 1 \text{ and } \mathbb{P}[T_2 > C_2] \rightarrow 1.$$

First consider  $T_1$  and note that, from the definition of  $\hat{\mathbf{c}}$ ,

$$T_1(A) = T(A, \hat{\mathbf{c}}) \geq T(A, \mathbf{c}^*),$$

and that from Lemma 1,

$$\mathbb{P}[T(A, \mathbf{c}^*) > \rho(P_n, \mathbf{c}^*) - \varepsilon_n] \rightarrow 1.$$

We claim that under each of the alternatives listed in the theorem statement,

$$\rho(P_n, \mathbf{c}^*) = \mathcal{O}(\sqrt{\alpha_n \varrho_n}). \quad (19)$$

whereas, from assumption A1,

$$\varepsilon_n = \sqrt{\alpha_n \varrho_n} \sqrt{\frac{\log(n \varrho_n \alpha_n)}{n \alpha_n \varrho_n}} = o(\sqrt{\alpha_n \varrho_n}),$$

which implies that

$$\rho(P_n, \mathbf{c}^*) > 2\varepsilon_n.$$

Therefore,

$$\mathbb{P}[T_1 > C_1] \geq \mathbb{P}[T(A, \mathbf{c}^*) > \varepsilon_n] \geq \mathbb{P}[T(A, \mathbf{c}^*) > \rho(P_n, \mathbf{c}^*) - \varepsilon_n] \rightarrow 1.$$

It remains to prove that the claim in (19) is true. Recall that

$$\rho(P_n, \mathbf{c}^*) = \frac{\sum_{i < j} \varrho_n (P_{ij} - \bar{p}) (\Delta_{ij}(\mathbf{c}^*) - \bar{\Delta}(\mathbf{c}^*))}{\sqrt{\binom{n}{2} \varrho_n \bar{p} (1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta}(\mathbf{c}^*) (1 - \bar{\Delta}(\mathbf{c}^*))}} = \frac{\sum_{i < j} \varrho_n (P_{ij} - \bar{p}) \Delta_{ij}(\mathbf{c}^*)}{\sqrt{\binom{n}{2} \varrho_n \bar{p} (1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta}(\mathbf{c}^*) (1 - \bar{\Delta}(\mathbf{c}^*))}}.$$

Now consider the three alternatives listed in the theorem statement:

- $H_1 : P_n$  is a two-block CP-SBM where  $p_{11} > p_{12} = p_{21} > p_{22}$ . Then,

$$\sum_{i < j} \varrho_n (P_{ij} - \bar{p}) \Delta_{ij}(\mathbf{c}^*) \geq n^2 \alpha_n \varrho_n (p_{12} - p_{22}) = \mathcal{O}(n^2 \alpha_n \varrho_n).$$

- $H_1 : P_n$  is a Chung-Lu model where  $\theta_{(k)} \theta_{(n)} > \theta_{(k+1)} \theta_{(k+2)}$  and  $\theta_{(k-1)} > \theta_{(k+1)}$ . Then,

$$\sum_{i < j} \varrho_n (P_{ij} - \bar{p}) \Delta_{ij}(\mathbf{c}^*) \geq n^2 \alpha_n \varrho_n (\theta_{(k)} \theta_{(n)} - \theta_{(k+1)} \theta_{(k+2)}) = \mathcal{O}(n^2 \alpha_n \varrho_n).$$

- $H_1 : P_n$  is a CP-DCBM where  $\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c \theta_{p,min}}$  and  $\frac{p_{11}}{p_{12}} > \frac{\theta_{p,max}}{\theta_c}$ . Then,

$$\sum_{i < j} \varrho_n (P_{ij} - \bar{p}) \Delta_{ij}(\mathbf{c}^*) \geq n^2 \alpha_n \varrho_n (p_{12} \theta_c \theta_{p,min} - p_{22} \theta_{p,max}^2) = \mathcal{O}(n^2 \alpha_n \varrho_n).$$

Therefore, in each case, we have

$$\rho(P_n, \mathbf{c}^*) = \frac{\sum_{i < j} \varrho_n (P_{ij} - \bar{p}) \Delta_{ij}(\mathbf{c}^*)}{\sqrt{\binom{n}{2} \varrho_n \bar{p} (1 - \varrho_n \bar{p}) \binom{n}{2} \bar{\Delta}(\mathbf{c}^*) (1 - \bar{\Delta}(\mathbf{c}^*))}} = \frac{\mathcal{O}(n^2 \alpha_n \varrho_n)}{\mathcal{O}(n^2 \sqrt{\alpha_n \varrho_n})} = \mathcal{O}(\sqrt{\alpha_n \varrho_n}),$$

which proves the claim made in (19).

Next, consider  $T_2 = \hat{p}_{11} - \hat{p}_{12}$ . Define

$$\bar{p}_{11} = \frac{1}{\binom{k}{2}} \sum_{i,j:c_i^*=c_j^*=1} P_{ij}, \quad \bar{p}_{12} = \frac{1}{\binom{k}{2}} \sum_{i,j:c_i^*=1,c_j^*=0} P_{ij},$$

and note that

$$\binom{k}{2} \hat{p}_{11} = \left[ \sum_{i,j:\hat{c}_i=\hat{c}_j=1} A_{ij} - \varrho_n \sum_{i,j:\hat{c}_i=\hat{c}_j=1} P_{ij} \right] + \varrho_n \left[ \sum_{i,j:\hat{c}_i=\hat{c}_j=1} P_{ij} - \sum_{i,j:c_i^*=c_j^*=1} P_{ij} \right] + \varrho_n \binom{k}{2} \bar{p}_{11}. \quad (20)$$

Note that the third term in the above expression is  $\mathcal{O}(n^2 \alpha_n^2 \varrho_n)$ . We will show that with probability going to one, the third term on the right hand side of (20) dominates the first and second terms.

Consider the first term on the right hand side of (20). Similar to the analysis of  $\sum_{i,j:\hat{c}_i=\hat{c}_j=1} A_{ij}$  under  $H_0$ , using Bernstein's inequality and union bound, we obtain

$$\mathbb{P} \left[ \max_{\mathbf{c}} \sum_{i,j:\hat{c}_i=\hat{c}_j=1} A_{ij} - \sum_{i,j:\hat{c}_i=\hat{c}_j=1} P_{ij} > -\frac{k^{1.5}}{\sqrt{2}} \varrho_n \log n \right] \rightarrow 1.$$

Since  $\alpha_n \geq \frac{(\log(n \log n))^2}{n}$ , we have  $k^{1.5} \varrho_n \log n = o(n^2 \alpha_n^2 \varrho_n)$ , which means that the third term dominates the first term with probability going to one.

Next consider the second term on the right hand side of (20), and refer to the schematic diagram in Figure 9. The shaded part represents the  $(i, j)$  pairs where  $\hat{\mathbf{c}}$  and  $\mathbf{c}^*$  agree. The summation in the second term is over the  $(i, j)$  pairs where  $\hat{\mathbf{c}}$  and  $\mathbf{c}^*$  disagree, i.e., the unshaded parts of Figure 9. From the definition of  $\xi_n(\hat{\mathbf{c}})$ , there are  $n\xi_n(\hat{\mathbf{c}})$  disagreements between  $\hat{\mathbf{c}}$  and  $\mathbf{c}^*$ . Let  $m$  be the number of vertices such that  $c_i^* \neq \hat{c}_i$ . Since  $\hat{\mathbf{c}}$  and  $\mathbf{c}^*$  have the same core size,

$$m = \frac{n\xi_n(\hat{\mathbf{c}})}{2}.$$

Therefore, the second term is given by

$$\varrho_n \left[ \sum_{i,j:\hat{c}_i=\hat{c}_j=1} P_{ij} - \sum_{i,j:c_i^*=c_j^*=1} P_{ij} \right] = \mathcal{O}(\varrho_n m k) = \mathcal{O}(\varrho_n k n \xi_n(\hat{\mathbf{c}})) = \mathcal{O}(n^2 \alpha_n \varrho_n \xi_n(\hat{\mathbf{c}})) = o_p(n^2 \alpha_n^2 \varrho_n),$$

where the final step follows from Theorem 2.1, where we proved that

$$\xi_n(\hat{\mathbf{c}}) = o_p(\alpha_n)$$

under each of the three alternatives. Therefore, the second term is  $o_p(n^2 \alpha_n^2 \varrho_n)$  under each of the three alternatives. Since the third term is  $\mathcal{O}(n^2 \alpha_n^2 \varrho_n)$ , the third term dominates the second term with probability going to one.

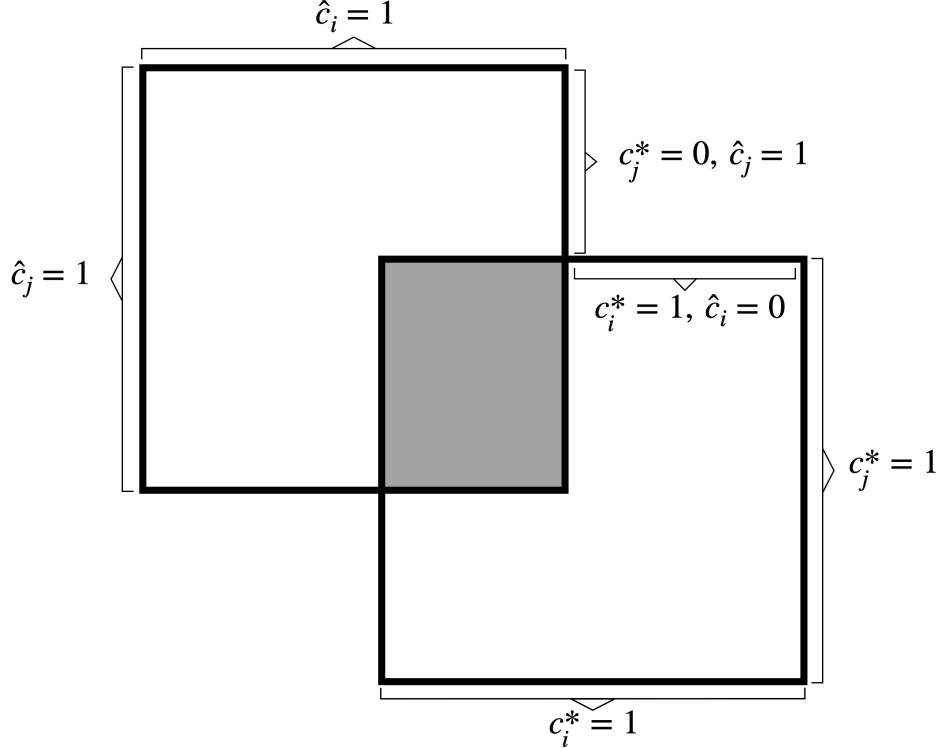


Figure 9: Schematic diagram for the second term

Summarizing the above, we conclude that with probability going to one,  $\varrho_n \binom{k}{2} \bar{p}_{11}$  dominates the expression for  $\binom{k}{2} \hat{p}_{11}$ . Therefore, with probability going to one,

$$\hat{p}_{11} - \varrho_n \bar{p}_{11} = o(\varrho_n \bar{p}_{11}) = o(\varrho_n).$$

Using the same approach we can show that with probability going to one,

$$\varrho_n \bar{p}_{12} - \hat{p}_{12} = o(\varrho_n).$$

We skip details given that the proof is very similar. Next, we show that  $\varrho_n (\bar{p}_{11} - \bar{p}_{12}) \geq \mathcal{O}(\varrho_n)$  under each of the three alternatives listed in the theorem statement:

- $H_1 : P_n$  is a two-block CP-SBM where  $p_{11} > p_{12} = p_{21} > p_{22}$ . Then,

$$\varrho_n (\bar{p}_{11} - \bar{p}_{12}) = \varrho_n (p_{11} - p_{12}) = \mathcal{O}(\varrho_n).$$

- $H_1 : P_n$  is a Chung-Lu model where  $\theta_{(k)} \theta_{(n)} > \theta_{(k+1)} \theta_{(k+2)}$  and  $\theta_{(k-1)} > \theta_{(k+1)}$ . Then,

$$\varrho_n (\bar{p}_{11} - \bar{p}_{12}) \geq \varrho_n (\theta_{(k)} \theta_{(k-1)} - \theta_{(k)} \theta_{(k+1)}) = \varrho_n \theta_{(k)} (\theta_{(k-1)} - \theta_{(k+1)}) = \mathcal{O}(\varrho_n).$$

- $H_1 : P_n$  is a CP-DCBM where  $\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c \theta_{p,min}}$  and  $\frac{p_{11}}{p_{12}} > \frac{\theta_{p,max}}{\theta_c}$ . Then,

$$\varrho_n (\bar{p}_{11} - \bar{p}_{12}) \geq \varrho_n (p_{11} \theta_c^2 - p_{12} \theta_c \theta_{p,max}) = \varrho_n \theta_c (p_{11} \theta_c - p_{12} \theta_{p,max}) = \mathcal{O}(\varrho_n).$$

Now, under the assumption that  $\alpha_n \geq \frac{(\log(n \log n))^2}{n}$ , we have  $C_2 = o(\varrho_n)$ , whereas  $\varrho_n(\bar{p}_{11} - \bar{p}_{12}) = \mathcal{O}(\varrho_n)$  in each case. This means

$$C_2 - \varrho_n(\bar{p}_{11} - \bar{p}_{12}) \leq -\frac{\varrho_n(\bar{p}_{11} - \bar{p}_{12})}{2},$$

and

$$\mathbb{P} \left[ \left\{ (\hat{p}_{11} - \varrho_n \bar{p}_{11}) > -\frac{\varrho_n(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \right] \rightarrow 1, \quad \mathbb{P} \left[ \left\{ (\varrho_n \bar{p}_{12} - \hat{p}_{12}) > -\frac{\varrho_n(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \right] \rightarrow 1.$$

Therefore,

$$\begin{aligned} \mathbb{P}[T_2(A) > C_2] &= \mathbb{P}[\hat{p}_{11} - \hat{p}_{12} > C_2] \\ &= \mathbb{P}[(\hat{p}_{11} - \varrho_n \bar{p}_{11}) + (\varrho_n \bar{p}_{12} - \hat{p}_{12}) + \varrho_n(\bar{p}_{11} - \bar{p}_{12}) > C_2] \\ &= \mathbb{P}[(\hat{p}_{11} - \varrho_n \bar{p}_{11}) + (\varrho_n \bar{p}_{12} - \hat{p}_{12}) > C_2 - \varrho_n(\bar{p}_{11} - \bar{p}_{12})] \\ &\geq \mathbb{P} \left[ (\hat{p}_{11} - \varrho_n \bar{p}_{11}) + (\varrho_n \bar{p}_{12} - \hat{p}_{12}) > -\frac{\varrho_n(\bar{p}_{11} - \bar{p}_{12})}{2} \right] \\ &\geq \mathbb{P} \left[ \left\{ (\hat{p}_{11} - \varrho_n \bar{p}_{11}) > -\frac{\varrho_n(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \cap \left\{ (\varrho_n \bar{p}_{12} - \hat{p}_{12}) > -\frac{\varrho_n(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \right] \\ &\rightarrow 1. \end{aligned}$$

This completes the proof.

## Proof of Theorem 3.2

**THEOREM 3.2.** *Under  $H_0$ , i.e., when the network is generated from the CL model, the Type I error rate converges to 0, i.e., for any  $\eta > 0$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] < \eta$ , where the rejection thresholds are given by*

$$C_1 = \rho(\hat{P}, \hat{c}) + \tilde{\epsilon}_n + \varepsilon'_n, \text{ where } \tilde{\epsilon}_n = \frac{\sqrt{\alpha_n} \log(n\alpha_n)}{n^{1.5} \sqrt{\varrho_n}}, \text{ and } \varepsilon'_n = \frac{\sqrt{\log(n\alpha_n \varrho_n)}}{n}. \quad (21)$$

*Under  $H_0$ , i.e., when the network is generated from the CP-DCBM, consider the three following assumptions.*

1.  $\varepsilon'_n + \tilde{\epsilon}_n = o(\alpha_n^{1.5} \varrho_n^{0.5})$ , where  $\tilde{\epsilon}_n = \frac{\sqrt{\alpha_n} \log(n\alpha_n)}{n^{1.5} \sqrt{\varrho_n}}$ ,  $\varepsilon'_n = \frac{\sqrt{\log(n\alpha_n \varrho_n)}}{n}$ .
2. There exists  $\epsilon > 0$  such that  $\frac{p_{12}^2 - p_{11}p_{22}}{n^4 \alpha_n^2 \varrho_n} (S_{10}^2 + S_{01}^2 - S_{11}S_{00}) > \epsilon$ , where  $S_{kl} = \sum_{(i,j): c_i^* = k, c_j^* = l} \theta_i \theta_j$ , for  $k, l = 0, 1$ .
3. Similar to Theorem 2.1, let  $\theta_c = \min\{\theta_i : c_i^* = 1\}$  be the lowest degree parameter in the core set, and  $\theta_{p,max} = \max\{\theta_i : c_i^* = 0\}$  and  $\theta_{p,min} = \min\{\theta_i : c_i^* = 0\}$  be the highest and lowest degree parameters in the periphery set, respectively. Then  $\frac{p_{12}}{p_{22}} > \frac{\theta_{p,max}^2}{\theta_c \theta_{p,min}}$ .

Under these three assumptions, the power of the test goes to one, i.e., for any  $\eta > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}[\{T_1(A) > C_1\} \cap \{T_2(A) > C_2\}] > 1 - \eta.$$

To prove this theorem, we first prove two results, (22) and (24), that apply to both the null and the alternative. Recall that we estimate the degree parameters as

$$\hat{\theta}_i = \frac{d_i}{\sqrt{2m}},$$

where  $2m = \sum_i \sum_j A_{ij}$  is the total degree of the network.

To make the derivations convenient to follow, we will use three “short cuts.” First, note that  $2m = \sum_i \sum_j A_{ij}$  is very tightly concentrated around its mean,  $\sum_i \sum_j P_{n,ij}$ , and the error bound for  $2m$  (or its scaled versions, such as  $\hat{p}$ ) is much smaller than error bounds for the other (appropriately scaled) random variables. Therefore, in what follows, we will ignore the randomness of  $2m$ , i.e., consider it to be approximately equal to  $\sum_i \sum_j P_{n,ij}$ . Second, for any  $i \neq j$  we will consider  $d_i$  and  $d_j$  to be independent for pedagogical convenience, although they share a common  $A_{ij}$ . This is justified because the correlation between  $d_i$  and  $d_j$  goes to zero at the rate of  $\mathcal{O}(1/n)$ . Whenever we use the first or second short cut we will flag it with the “ $\approx$ ” sign. Third, we will drop the subscript  $n$  from expressions such as  $P_{n,ij}$  for notational convenience.

We first derive the expectation of  $\hat{P}_{ij} = \hat{\theta}_i \hat{\theta}_j$  under  $H_0$  and  $H_1$ . Under  $H_0$ , we have

$$\mathbb{E}[\hat{\theta}_i \hat{\theta}_j] = \mathbb{E}\left[\frac{d_i d_j}{2m}\right] \approx \frac{\mathbb{E}[d_i]\mathbb{E}[d_j]}{\sum_i \sum_j P_{ij}} = \frac{(\theta_i \sum_{k=1}^n \theta_k)(\theta_j \sum_{k=1}^n \theta_k)}{(\sum_{k=1}^n \theta_k)^2} = \theta_i \theta_j,$$

and under  $H_1$ , we have

$$\mathbb{E}[\hat{\theta}_i \hat{\theta}_j] = \mathbb{E}\left[\frac{d_i d_j}{2m}\right] \approx \frac{\mathbb{E}[d_i]\mathbb{E}[d_j]}{\sum_i \sum_j P_{ij}} = \frac{(\theta_i \sum_{k=1}^n p_{c_i^* c_k^*} \theta_k)(\theta_j \sum_{l=1}^n p_{c_j^* c_l^*} \theta_l)}{\sum_{k=1}^n \sum_{l=1}^n \theta_k \theta_l p_{c_k^* c_l^*}}.$$

Now, define

$$\tilde{P}_{ij} = \begin{cases} \theta_i \theta_j & \text{if } H_0 \text{ is true} \\ \theta_i \theta_j \frac{\sum_{k=1}^n \sum_{l=1}^n \theta_k \theta_l p_{c_i^* c_k^*} p_{c_j^* c_l^*}}{\sum_{k=1}^n \sum_{l=1}^n \theta_k \theta_l p_{c_k^* c_l^*}} & \text{if } H_1 \text{ is true} \end{cases}$$

for  $i, j = 1, \dots, n$ . Then we have  $\mathbb{E}[\hat{P}_{ij}] = \tilde{P}_{ij}$  under both  $H_0$  and  $H_1$ . Note that  $\tilde{P}_{ij} = P_{ij}$  for each  $i, j$  under  $H_0$  but not under  $H_1$ . However,  $\sum_{ij} \tilde{P}_{ij} = \sum_{ij} P_{ij}$  under both  $H_0$  and  $H_1$ . This is a key observation that we will leverage in what follows.

Next, we state and prove the following result: under both  $H_0$  and  $H_1$ ,

$$\mathbb{P}\left[|\rho(\hat{P}, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*)| \leq \tilde{\epsilon}_n\right] \rightarrow 0 \tag{22}$$

as  $n \rightarrow \infty$ , where

$$\tilde{\epsilon}_n = \mathcal{O}\left(\frac{\sqrt{\alpha_n} \log(n\alpha_n)}{n^{1.5} \sqrt{\varrho_n}}\right).$$

To prove (22), write

$$\rho(\hat{P}, \mathbf{c}^*) = \frac{X_1 - X_2}{X_3},$$

where

$$\begin{aligned} X_1 &= \sum_{i < j} \hat{P}_{ij} \Delta_{ij} = \sum_{i < j} \frac{d_i d_j}{2m} \Delta_{ij} \approx \frac{\sum_{i < j} d_i d_j \Delta_{ij}}{\sum_i \sum_j P_{ij}}, \\ X_2 &= \binom{n}{2} \bar{\Delta} \hat{p} \approx \sum_{i < j} \binom{n}{2} \bar{\Delta} \bar{p}, \text{ and} \\ X_3 &= \binom{n}{2} \sqrt{\bar{\Delta}(1 - \bar{\Delta})\hat{p}(1 - \hat{p})} \approx \binom{n}{2} \sqrt{\bar{\Delta}(1 - \bar{\Delta})\bar{p}(1 - \bar{p})}. \end{aligned}$$

Similarly, write

$$\rho(\tilde{P}, \mathbf{c}^*) = \frac{\mu_1 - \mu_2}{\mu_3},$$

where

$$\mu_1 = \sum_{i < j} \tilde{P}_{ij} \Delta_{ij}, \mu_2 = \sum_{i < j} \binom{n}{2} \bar{\Delta} \bar{p}, \text{ and } \mu_3 = \binom{n}{2} \sqrt{\bar{\Delta}(1 - \bar{\Delta})\bar{p}(1 - \bar{p})}.$$

Note that  $\mathbb{E}[X_1] \approx \mu_1$ ,  $X_2 \approx \mu_2$ , and  $X_3 \approx \mu_3$ . Therefore,

$$|\rho(\hat{P}, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*)| = \left| \frac{X_1 - X_2}{X_3} - \frac{\mu_1 - \mu_2}{\mu_3} \right| \approx \left| \frac{X_1 - \mathbb{E}[X_1]}{\mu_3} \right|. \quad (23)$$

Define  $Y = X_1 \sum_i \sum_j P_{ij} = \sum_{i < j} d_i d_j \Delta_{ij}$ . Now, without loss of generality, suppose that  $c_i^* = 1$  for  $i = 1, \dots, k$  and  $c_i^* = 0$  for  $i = k + 1, \dots, n$ . Then we can write

$$\begin{aligned} Y &= \sum_{i < j} d_i d_j \Delta_{ij} \\ &= d_1(d_2 + \dots + d_n) + d_2(d_3 + \dots + d_n) + \dots + d_k(d_{k+1} + \dots + d_n) \\ &= U_1 + \dots + U_k, \end{aligned}$$

where  $U_i = d_i(d_{i+1} + \dots + d_n)$  for  $i = 1, \dots, k$ . Note that while the  $U_i$ 's are dependent across  $i = 1, \dots, k$ , each  $U_i$  is a sum of independent Bernoulli random variables, since we are working under the approximation that  $d_i$  and  $d_j$  are independent for  $i \neq j$ . Therefore for each  $i = 1, \dots, k$ , using Bernstein's inequality,

$$\mathbb{P}[|U_i - \mathbb{E}(U_i)| > t] \leq 2 \exp\left(-\frac{t^2/2}{\mathbb{E}(U_i) + t/3}\right).$$

Note that  $\mathbb{E}(U_i) = \mathcal{O}(n^3 \varrho_n^2)$ . Therefore, by setting  $t = n^{1.5} \varrho_n \log(n\alpha_n)$  and using union

bound over  $i = 1, \dots, k$ , we obtain that

$$\begin{aligned} \cup_{i=1}^k \mathbb{P}[|U_i - \mathbb{E}(U_i)| > t] &\leq 2 \sum_{i=1}^k \exp\left(-\frac{t^2/2}{\mathbb{E}(U_i) + t/3}\right) \\ &= 2 \exp\left(-\frac{\mathcal{O}(n^3 \varrho_n^2 \log^2(n\alpha_n))}{\mathcal{O}(n^3 \varrho_n^2) + \mathcal{O}(n^{1.5} \varrho_n \log(n\alpha_n))} + \log k\right) \\ &= 2 \exp\left(-\frac{\mathcal{O}(n^3 \varrho_n^2 \log^2(n\alpha_n))}{\mathcal{O}(n^3 \varrho_n^2) + \mathcal{O}(n^{1.5} \varrho_n \log(n\alpha_n))} + \log(n\alpha_n)\right) \\ &\rightarrow 0, \end{aligned}$$

which implies that

$$\mathbb{P}[|Y - \mathbb{E}(Y)| > kt] \rightarrow 0 \Rightarrow \mathbb{P}\left[|X_1 - \mathbb{E}(X_1)| > \frac{kt}{\mathcal{O}(n^2 \varrho_n)}\right] \rightarrow 0.$$

Recall that from (23),

$$|\rho(\hat{P}, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*)| \approx \left| \frac{X_1 - \mathbb{E}[X_1]}{\mu_3} \right| = \left| \frac{X_1 - \mathbb{E}[X_1]}{\mathcal{O}(n^2 \sqrt{\alpha_n \varrho_n})} \right|.$$

Therefore, with probability going to 1,

$$|\rho(\hat{P}, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*)| \leq \frac{kt}{\mathcal{O}(n^2 \varrho_n) \mathcal{O}(n^2 \sqrt{\alpha_n \varrho_n})} = \frac{n^{2.5} \alpha_n \varrho_n \log(n\alpha_n)}{\mathcal{O}(n^4 \varrho_n \sqrt{\alpha_n \varrho_n})} = \mathcal{O}\left(\frac{\sqrt{\alpha_n} \log(n\alpha_n)}{n^{1.5} \sqrt{\varrho_n}}\right).$$

This completes the proof of (22).

Next, we state and prove the following result: under both  $H_0$  and  $H_1$ ,

$$\mathbb{P}[|T(A, \mathbf{c}^*) - \rho(P, \mathbf{c}^*)| \leq \varepsilon'_n] \rightarrow 0 \quad (24)$$

as  $n \rightarrow \infty$ , where

$$\varepsilon'_n = \frac{\sqrt{\log(n\alpha_n \varrho_n)}}{n}.$$

To prove (24), we will essentially mimic the proof of Lemma 1. Let  $\mathbf{c} = \mathbf{c}^*$  and  $X_n(\mathbf{c}) = \sum_{i < j} A_{ij}(\Delta_{ij} - \bar{\Delta}_n)$  and  $Y_n = \sum_{i < j} A_{ij}$ . Then

$$T(A_n, \mathbf{c}) = \frac{X_n}{\sqrt{Y_n(\binom{n}{2} - Y_n)}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}} := f(X_n, Y_n).$$

We will prove the desired result using two pieces. First, we show that  $X_n(\mathbf{c})$  and  $Y_n$  are “close” to their respective expectations,  $\mu_x(\mathbf{c})$  and  $\mu_y$ , for all  $\mathbf{c}$ . Then we will show that  $f(x, y)$  is “close” to  $f(x^*, y^*)$  when  $x \approx x^*$  and  $y \approx y^*$ .

First, Bernstein’s inequality states that if  $\mathbb{E}(X_i) = 0$  and  $|X_i| \leq M$  for  $i = 1, \dots, n$ , then

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) \leq \exp\left(-\frac{\frac{1}{2}t^2}{\sum_{i=1}^n \mathbb{E}(X_i^2) + \frac{1}{3}Mt}\right).$$

Thus,

$$\mathbb{P}(|X_n(\mathbf{c}) - \mu_x(\mathbf{c})| > t_1) \leq 2 \exp\left(-\frac{\frac{1}{2}t_1^2}{\sum_{i < j} P_{ij}(1 - P_{ij})(\Delta_{ij} - \bar{\Delta}_n)^2 + \frac{1}{3}t_1}\right) \sim \exp\left(-\frac{t_1^2}{\varrho_n \alpha_n n^2 + t_1}\right)$$

since  $P_{ij}(1 - P_{ij}) = \mathcal{O}(\varrho_n)$  and  $\sum_{i < j} (\Delta_{ij} - \bar{\Delta}_n)^2 = \mathcal{O}(\alpha_n n^2)$ . Let  $t_1 = n\sqrt{\varrho_n \alpha_n \log(n\varrho_n \alpha_n)}$ . Then

$$\mathbb{P}(|X_n - \mu_x(\mathbf{c})| > t_1) \leq 2 \exp\left(-\frac{n^2 \varrho_n \alpha_n \log(n\varrho_n \alpha_n)}{\varrho_n \alpha_n n^2}\right) \leq 2 \exp(-\log(n\varrho_n \alpha_n)) \rightarrow 0$$

since  $n\varrho_n \alpha_n \rightarrow \infty$  by A1. Additionally,

$$\mathbb{P}(|Y_n - \mu_y| > t_2) \leq 2 \exp\left(-\frac{\frac{1}{2}t_2^2}{\sum_{i < j} P_{ij}(1 - P_{ij}) + \frac{1}{3}t_2}\right) \sim 2 \exp\left(-\frac{t_2^2}{\varrho_n n^2 + t_2}\right).$$

Additionally, letting  $t_2 = n\sqrt{\varrho_n \log(n\varrho_n \alpha_n)}$ ,

$$\mathbb{P}(|Y_n - \mu_y| > n\sqrt{\varrho_n \log(n\varrho_n \alpha_n)}) \leq 2 \exp\left(-\frac{n^2 \varrho_n \log(n\varrho_n \alpha_n)}{\varrho_n n^2 + n\sqrt{\varrho_n \log(n\varrho_n \alpha_n)}}\right) \leq 2 \exp(-\log(n\varrho_n \alpha_n)) \rightarrow 0.$$

Thus, for large  $n$ ,  $X_n(\mathbf{c})$  and  $Y_n$  are “close” to their respective expectations for all  $\mathbf{c}$ .

For the second step, recall that

$$f(x, y) = \frac{x}{\sqrt{y(n^2 - y)}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}}.$$

Then we want to show that  $|f(x, y) - f(x^*, y^*)|$  is upper bounded by a small constant when  $|x - x^*| < t_1$  and  $|y - y^*| < t_2$ . Consider the first-order Taylor expansion:

$$f(x, y) - f(x^*, y^*) = (x - x^*) \frac{\partial}{\partial x} f(x, y, n) \Big|_{x=x^*, y=y^*} + (y - y^*) \frac{\partial}{\partial y} f(x, y, n) \Big|_{x=x^*, y=y^*} + R_n,$$

where

$$R_n \sim (x - x^*, y - y^*) \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial_{xy} f}{\partial x \partial y} \\ \frac{\partial_{xy} f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}.$$

We have that  $x^* = \mathcal{O}(\alpha_n \varrho_n n^2)$  and  $y^* = \mathcal{O}(\varrho_n n^2)$ . It’s also easy to see that

$$\frac{\partial}{\partial x} f(x, y) \Big|_{x=x^*, y=y^*} = \frac{1}{\sqrt{y^*(n^2 - y^*)}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}} \sim \alpha_n^{-1/2} \varrho_n^{-1/2} n^{-2},$$

and

$$\frac{\partial}{\partial y} f(x, y) \Big|_{x=x^*, y=y^*} = \frac{x^*(n^2 - 2y^*)}{2\{y^*(n^2 - y^*)\}^{3/2}} \frac{1}{\sqrt{\bar{\Delta}_n(1 - \bar{\Delta}_n)}} \sim \alpha_n^{1/2} \varrho_n^{-1/2} n^{-2}.$$

From step 1,  $|x - x^*| < t_1 = n\sqrt{\varrho_n \alpha_n \log(n\varrho_n \alpha_n)}$  and  $|y - y^*| < t_2 = n\sqrt{\varrho_n \log(n\varrho_n \alpha_n)}$  with high probability, which means that with probability going to one,

$$|f(x, y) - f(x^*, y^*)| \leq \frac{n\sqrt{\varrho_n \alpha_n \log(n\varrho_n \alpha_n)}}{n^2 \sqrt{\alpha_n \varrho_n}} + \frac{n\sqrt{\alpha_n \varrho_n \log(n\varrho_n \alpha_n)}}{\varrho_n^{1/2} n^2} \sim \frac{\sqrt{\log(n\alpha_n \varrho_n)}}{n} = \varepsilon'_n.$$

This completes the proof of (24).

Next, we will use (22) and (24) to prove the desired convergence of type-1 error and power under  $H_0$  and  $H_1$ , respectively.

**Under  $H_0$ :**

Note that since  $\alpha_n/\sqrt{\varrho_n} = o(1/\sqrt{n \log n})$ , strong consistency is achieved as per Theorem 2.1. Recall that  $C_1 = \rho(\hat{P}, \mathbf{c}^*) + \varepsilon'_n + \tilde{\epsilon}_n$ . Under  $H_0$ , we have

$$\begin{aligned} T_1 &= T(A, \hat{\mathbf{c}}) \\ &= T(A, \mathbf{c}^*) \text{ with probability going to one, since strong consistency holds} \\ &\leq \rho(P, \mathbf{c}^*) + \varepsilon'_n \text{ with probability going to one, by (24)} \\ &= \rho(\tilde{P}, \mathbf{c}^*) + \varepsilon'_n \text{ since } \tilde{P} = P \text{ under } H_0 \\ &\leq \rho(\hat{P}, \mathbf{c}^*) + \varepsilon'_n + \tilde{\epsilon}_n \text{ with probability going to one, by (22)} \\ &= \rho(\hat{P}, \hat{\mathbf{c}}) + \varepsilon'_n + \tilde{\epsilon}_n \text{ with probability going to one, since strong consistency holds} \\ &= C_1, \end{aligned}$$

which completes the proof of convergence of type-1 error. Note that we do not need to analyze the second part of the intersection test, since the first part is sufficient to prove that type-1 error converges to zero.

**Under  $H_1$ :**

Note that since  $\alpha_n/\sqrt{\varrho_n} = o(1/\sqrt{n \log n})$ , strong consistency is achieved as per Theorem 2.3. Here we will need to show that both  $\mathbb{P}[T_1 > C_1]$  and  $\mathbb{P}[T_2 > C_2]$  go to one, which will prove that the power of the intersection test goes to one.

First, consider  $T_1$  and recall that  $C_1 = \rho(\hat{P}, \mathbf{c}^*) + \varepsilon'_n + \tilde{\epsilon}_n$ . Also recall that under  $H_1$ ,  $\mathbb{E}[\hat{P}_{ij}] = \hat{P}_{ij} \neq P_{ij}$ . The proof hinges on this key observation, and we start by carefully quantifying the difference between  $P_{ij}$  and  $\hat{P}_{ij}$ . To this end, note that  $\sum_{ij} \hat{P}_{ij} = \sum_{ij} P_{ij}$ , which means

$$\rho(P, \mathbf{c}^*) - \rho(\hat{P}, \mathbf{c}^*) = \frac{\sum_{i < j} (P_{ij} - \hat{P}_{ij}) \Delta_{ij}}{\frac{1}{2} n(n-1) \{\bar{P}(1-\bar{P})\bar{\Delta}(1-\bar{\Delta})\}^{1/2}}.$$

Recall that  $S_{kl} = \sum_{(i,j): c_i^* = k, c_j^* = l} \theta_i \theta_j$ , for  $k, l = 1, 2$ , and consider the numerator of the above

expression.

$$\begin{aligned}
\sum_{i < j} \left( P_{ij} - \tilde{P}_{ij} \right) \Delta_{ij} &= \sum_{i < j} \left( \theta_i \theta_j p_{c_i^* c_j^*} - \theta_i \theta_j \frac{\sum_{k=1}^n \sum_{l=1}^n \theta_k \theta_l p_{c_i^* c_k^*} p_{c_j^* c_l^*}}{\sum_{k=1}^n \sum_{l=1}^n \theta_k \theta_l p_{c_k^* c_l^*}} \right) \Delta_{ij} \\
&= \frac{1}{2} \sum_{i,j} \theta_i \theta_j p_{c_i^* c_j^*} \Delta_{ij} \left( 1 - \frac{\sum_{k=1}^n \sum_{l=1}^n \theta_k \theta_l p_{c_i^* c_k^*} p_{c_j^* c_l^*}}{\sum_{k=1}^n \sum_{l=1}^n \theta_k \theta_l p_{c_k^* c_l^*} p_{c_i^* c_j^*}} \right) \\
&= \frac{1}{2} \sum_{i,j} \theta_i \theta_j p_{c_i^* c_j^*} \Delta_{ij} \left( \frac{\sum_{k,l=1}^n \theta_k \theta_l (p_{c_k^* c_l^*} p_{c_i^* c_j^*} - p_{c_i^* c_k^*} p_{c_j^* c_l^*})}{\sum_{k,l=1}^n \theta_k \theta_l p_{c_k^* c_l^*} p_{c_i^* c_j^*}} \right) \\
&= \frac{1}{2 \sum_{k,l=1}^n \theta_k \theta_l p_{c_k^* c_l^*}} \sum_{i,j} \theta_i \theta_j \Delta_{ij} \left( \sum_{k,l=1}^n \theta_k \theta_l (p_{c_k^* c_l^*} p_{c_i^* c_j^*} - p_{c_i^* c_k^*} p_{c_j^* c_l^*}) \right) \\
&= \frac{1}{2n^2 \bar{p}} \sum_{i,j} \theta_i \theta_j \Delta_{ij} \left( \sum_{k,l=1}^n \theta_k \theta_l (p_{c_k^* c_l^*} p_{c_i^* c_j^*} - p_{c_i^* c_k^*} p_{c_j^* c_l^*}) \right) \\
&= \frac{p_{12}^2 - p_{11}p_{22}}{2n^2 \bar{p}} (S_{12}^2 + S_{21}^2 - S_{11}S_{22}) \\
&> \epsilon n^2 \alpha_n^2 \varrho_n
\end{aligned}$$

from the condition in the theorem statement. This implies that

$$\rho(P, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*) = \frac{\sum_{i < j} \left( P_{ij} - \tilde{P}_{ij} \right) \Delta_{ij}}{\binom{n}{2} \{ \bar{P}(1 - \bar{P}) \bar{\Delta}(1 - \bar{\Delta}) \}^{1/2}} > \frac{\epsilon n^2 \alpha_n^2 \varrho_n}{\mathcal{O}(n^2 \sqrt{\alpha_n \varrho_n})} > \mathcal{O}(\alpha_n^{1.5} \varrho_n^{0.5}) >> \varepsilon'_n + \tilde{\varepsilon}_n. \quad (25)$$

where the last two steps follow from the assumptions in the theorem. Therefore, under  $H_1$ ,

$$\begin{aligned}
T_1 &= T(A, \hat{\mathbf{c}}) \\
&\geq T(A, \mathbf{c}^*) \text{ by definition of } \hat{\mathbf{c}} \\
&\geq \rho(P, \mathbf{c}^*) - \varepsilon'_n \text{ with probability going to one, by (24)} \\
&= \left( \rho(\hat{P}, \mathbf{c}^*) + \tilde{\varepsilon}_n + \varepsilon'_n \right) + \left( \rho(P, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*) - (2\tilde{\varepsilon}_n + \varepsilon'_n) \right) + \left( \rho(\tilde{P}, \mathbf{c}^*) - \rho(\hat{P}, \mathbf{c}^*) \right) \\
&\geq \left( \rho(\hat{P}, \mathbf{c}^*) + \tilde{\varepsilon}_n + \varepsilon'_n \right) + \left( \rho(P, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*) - (2\tilde{\varepsilon}_n + \varepsilon'_n) \right) - \tilde{\varepsilon}_n \text{ with prob } \rightarrow 1, \text{ by (22)} \\
&= \left( \rho(\hat{P}, \mathbf{c}^*) + \tilde{\varepsilon}_n + \varepsilon'_n \right) + \left( \rho(P, \mathbf{c}^*) - \rho(\tilde{P}, \mathbf{c}^*) - 2(\tilde{\varepsilon}_n + \varepsilon'_n) \right) \\
&> \rho(\hat{P}, \mathbf{c}^*) + \tilde{\varepsilon}_n + \varepsilon'_n \text{ by (25)} \\
&= \rho(\hat{P}, \hat{\mathbf{c}}) + \tilde{\varepsilon}_n + \varepsilon'_n \text{ with probability going to one, since strong consistency holds} \\
&= C_1,
\end{aligned}$$

which completes the proof of convergence of power for the first part of the intersection test.

For the second part of the intersection test, we will essentially mimic the second part of the proof of Theorem 3.1. Consider  $T_2 = \hat{p}_{11} - \hat{p}_{12}$ , and note that when  $\hat{\mathbf{c}} = \mathbf{c}^*$ ,

$$\binom{k}{2} \hat{p}_{11} = \sum_{i,j: c_i^* = c_j^* = 1} (A_{ij} - P_{ij}) + \sum_{i,j: c_i^* = c_j^* = 1} P_{ij}.$$

For the first term, using Bernstein's inequality we obtain

$$\mathbb{P} \left[ \left| \sum_{i,j:c_i^*=c_j^*=1} (A_{ij} - P_{ij}) \right| > n\alpha_n \sqrt{\varrho_n \log(n\alpha_n \varrho_n)} \right] \rightarrow 1.$$

Note that  $\sum_{i,j:c_i^*=c_j^*=1} P_{ij} = \mathcal{O}(n^2 \alpha_n^2 \varrho_n) \gg n\alpha_n \sqrt{\varrho_n \log(n\alpha_n \varrho_n)}$ , which means that the second term dominates the first term with probability going to one. Therefore, with probability going to one,

$$\hat{p}_{11} - \bar{p}_{11} = o(\bar{p}_{11}) = o(\varrho_n),$$

where

$$\bar{p}_{11} = \frac{1}{\binom{k}{2}} \sum_{i,j:c_i^*=c_j^*=1} P_{ij}.$$

Using the same approach we can show that with probability going to one,

$$\bar{p}_{12} - \hat{p}_{12} = o(\bar{p}_{12}) = o(\varrho_n),$$

where

$$\bar{p}_{11} = \frac{1}{k(n-k)} \sum_{i,j:c_i^*=1,c_j^*=0} P_{ij}.$$

We skip details given that the proof is very similar. Finally, note that under  $H_1$  and the assumption that  $\alpha_n \geq \frac{(\log(n \log n))^2}{n}$ , we have  $C_2 = o(\varrho_n)$ , whereas  $(\bar{p}_{11} - \bar{p}_{12}) = \mathcal{O}(\varrho_n)$  from the third assumption in the theorem. This means

$$C_2 - (\bar{p}_{11} - \bar{p}_{12}) \leq -\frac{(\bar{p}_{11} - \bar{p}_{12})}{2},$$

and

$$\mathbb{P} \left[ \left\{ (\hat{p}_{11} - \bar{p}_{11}) > -\frac{(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \right] \rightarrow 1, \quad \mathbb{P} \left[ \left\{ (\bar{p}_{12} - \hat{p}_{12}) > -\frac{(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \right] \rightarrow 1.$$

Therefore,

$$\begin{aligned} \mathbb{P}[T_2(A) > C_2] &= \mathbb{P}[\hat{p}_{11} - \hat{p}_{12} > C_2] \\ &= \mathbb{P}[(\hat{p}_{11} - \bar{p}_{11}) + (\bar{p}_{12} - \hat{p}_{12}) + (\bar{p}_{11} - \bar{p}_{12}) > C_2] \\ &= \mathbb{P}[(\hat{p}_{11} - \bar{p}_{11}) + (\bar{p}_{12} - \hat{p}_{12}) > C_2 - (\bar{p}_{11} - \bar{p}_{12})] \\ &\geq \mathbb{P} \left[ (\hat{p}_{11} - \bar{p}_{11}) + (\bar{p}_{12} - \hat{p}_{12}) > -\frac{(\bar{p}_{11} - \bar{p}_{12})}{2} \right] \\ &\geq \mathbb{P} \left[ \left\{ (\hat{p}_{11} - \bar{p}_{11}) > -\frac{(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \cap \left\{ (\bar{p}_{12} - \hat{p}_{12}) > -\frac{(\bar{p}_{11} - \bar{p}_{12})}{4} \right\} \right] \\ &\rightarrow 1. \end{aligned}$$

This completes the proof.