Individual-centered Partial Information in Social Networks

Xiao Han   XHAN011@USTC.EDU.CN [*]
International Institute of Finance, School of Management
University of Science and Technology of China
Hefei, 230026, China

Y. X. Rachel Wang   RACHEL.WANG@SYDNEY.EDU.AU
School of Mathematics and Statistics
University of Sydney
NSW, 2006, Australia

Qing Yang   YANGQ@USTC.EDU.CN
International Institute of Finance, School of Management
University of Science and Technology of China
Hefei, 230026, China

Xin Tong   XINT@MARSHALL.USC.EDU
Department of Data Sciences and Operations, Marshall School of Business
University of Southern California
CA, 90089, USA

Abstract

In statistical network analysis, we often assume either the full network is available or multiple subgraphs can be sampled to estimate various global properties of the network. However, in a real social network, people frequently make decisions based on their local view of the network alone. Here, we consider a partial information framework that characterizes the local network centered at a given individual by path length L and gives rise to a partial adjacency matrix. Under $L = 2$, we focus on the problem of (global) community detection using the popular stochastic block model (SBM) and its degree-corrected variant (DCSBM). We derive theoretical properties of the eigenvalues and eigenvectors from the signal term of the partial adjacency matrix and propose new spectral-based community detection algorithms that achieve consistency under appropriate conditions. Our analysis also allows us to propose a new centrality measure that assesses the importance of an individual's partial information in determining global community structure. Using simulated and real networks, we demonstrate the performance of our algorithms and compare our centrality measure with other popular alternatives to show it captures unique nodal information. Our results illustrate that the partial information framework enables us to compare the viewpoints of different individuals regarding the global structure.

Keywords: community detection, centrality measure, partial information

---

[*]. Han and Wang contribute equally to the work. Yang and Tong are corresponding authors.

1. Introduction

Much of the statistical network literature is focused on estimating global properties of graphs by either using the whole graphs or combining information across appropriately sampled subgraphs. However, in social networks, despite the prevalence of social media tools, most individuals still have limited understanding of the information that exists beyond their local network, e.g., friends' friends. In this way, one often needs to make important decisions, such as whether or not to share sensitive information with friends belonging to different social circles, based on their limited local view of the global network. In this paper, we adopt such an individual-centered perspective to study global structures of social networks.

To formalize an individual's local network structure, we consider an *individual-centered partial information framework* that uses path length to characterize connections visible (proximal) and hidden (distant) to the individual. Concretely, let $G = (\mathcal{V}, \mathcal{E})$ denote the global network of interest, where $\mathcal{V} = \{1, \ldots, n\}$ is the index set of all individuals (nodes) and $\mathcal{E}$ is the set of edges, assumed to be unweighted and undirected for simplicity, between individuals. We characterize an individual's partial knowledge of the network by their *knowledge depth*: an individual i has knowledge depth L if all paths (starting from i) of length up to L in the network are known to the individual. Figure 1 illustrates the knowledge depth concept with a toy example, where the left panel is the full network. Taking individual 1 as the individual of interest, the left, middle and right panels correspond to their perceived networks given knowledge depths $L = 3, 2, 1$, respectively. In choosing the value of L for our study, we first note that the case $L = 1$ leads to a too simplistic graph structure around the individual. On the other hand, the "six degrees of separation" phenomenon (Watts and Strogatz, 1998) in social networks suggests moderate values of L are likely to return the full network already. Both $L = 2$ and $L = 3$ are more interesting cases to study and coincide with most people's experience in real social networks. We focus on $L = 2$ in this paper and leave $L = 3$ for future studies.
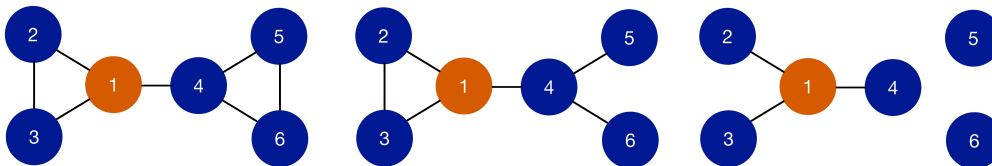

Figure 1: A toy network of 6 individuals. $L = 3$ or *full* (left); $L = 2$ (middle); $L = 1$ (right).

The structure of the partial network considered in our work is related to popular sampling schemes in social sciences, including egocentric sampling (Freeman, 1982; Wasserman et al., 1994), snowball sampling and respondent-driven sampling (RDS) (Goodman, 1961; Heckathorn, 1997; Salganik and Heckathorn, 2004), which all work by first selecting individuals as "seeds", then expanding into their neighborhoods according to certain criteria. Information across the sampled subgraphs can be combined to estimate parameters in network models and node covariates (Handcock and Gile, 2010; Rohe, 2019). Egocentric sampling and other random sampling schemes have also been used to estimate community structure (Mukherjee et al., 2021) and subgraph counts (Bhattacharyya and Bickel, 2015). Compared to these works, a major conceptual difference in our framework is that we are interested in

understanding the network structure visible to *each* node and how these structures differ for different nodes. In addition, sampling multiple seeds may be infeasible in certain networks with restricted access (e.g., a terrorist network). These observations motivate us to consider how to infer global information using *one* local network only.

In this paper, we choose community memberships to be the global network feature we aim to infer from an individual-centered partial network. Community detection is one of the most studied statistical network problems, with the stochastic block model (SBM) (Holland et al., 1983; Abbe, 2017) and its variants including the degree-corrected SBM (DCSBM) (Karrer and Newman, 2011) and the mixed membership SBM (MMSBM) (Airoldi et al., 2008) being popular generative models. There is a rich line of literature on community detection under these models; many methods are based on likelihood or spectral approaches (Bickel and Chen, 2009; Zhao et al., 2012; Rohe et al., 2011; Lei et al., 2015; Anandkumar et al., 2014; Jin, 2015; Jin et al., 2017). Much attention has also been paid to other statistical inference problems in this high-dimensional network setting, including inference for the number of communities (e.g., Bickel and Sarkar 2016; Lei 2016; Wang et al. 2017; Saldana et al. 2017; Han et al. 2023) and the membership profiles (e.g., Fan et al. 2019a). We adopt the spectral approach in this paper. Since the adjacency matrix of the partial network is significantly different from the global adjacency matrix, we first perform detailed theoretical analysis of its spectral properties before proposing novel spectral algorithms for inferring the global community memberships under the SBM and DCSBM models.

individual of interest	H	2	3	A	20	32
degrees	16	9	10	17	3	6
fraction of edges	.654	.513	.705	.641	.526	.654
fraction of individuals	.765	.676	.912	.706	.912	.971
detection accuracy	.529	.706	.941	.738	.941	.824

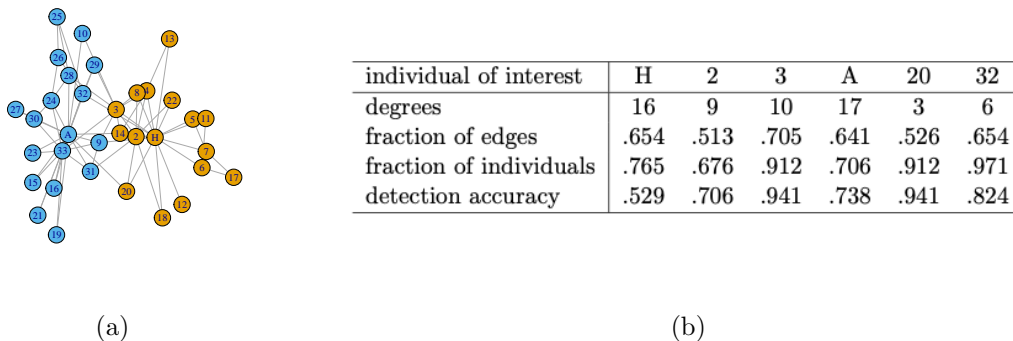(a)   (b)

Figure 2: The karate club network with nodes colored by ground truth community labels in (a) and community detection results (using Algorithm 3) for chosen individuals in (b).

Our partial information framework and new clustering algorithms enable us gain interesting insight into the relationship between global community detection and the local features perceived by a given individual. As a motivating example, take the well-known Zakary's karate club data. Figure 2 shows the network colored by ground truth community labels, and the results of community detection using our algorithm for a few chosen individuals. The network has two obvious hubs as demonstrated by their degrees: the instructor "Mr. Hi" (node "H") and the administrator "John A." (node "A"). It turns out surprisingly that these nodes perform less well than some nodes with significantly smaller degrees (nodes 3, 20, 32), suggesting their partial networks are not the most informative when it comes to understanding the affiliation structure of the whole network. In com-

parison, node 20, despite their small degree and observing a smaller fraction of edges or individuals through their $L = 2$ local network compared to nodes 3 and 32, achieves one of the highest detection accuracy rates. Noting that two of the three connections of node 20 come from hubs "H" and "A", the result implies that not just how many friends one knows, but also *whom* one knows, matters in inferring global structure. We describe our algorithm (Algorithm 3), theoretical analysis and interpretations in Section 3.3. More analysis of this network can be found in Section 5.1.

Our paper makes the following main contributions:

- We propose an individual-centered partial information framework to study social networks from *an individual*'s perspective, where the person's understanding of the whole network is characterized by their knowledge depth L.

- We address the technical challenges brought about by the partial adjacency matrix, deriving novel results on its spectral properties for $L = 2$. Different from the conventional case, the signal term of this matrix is random itself. Under some generic assumptions, we derive its rank, the explicit forms of its eigenvectors and the approximate forms of its eigenvalues, which can be of independent theoretical interest.

- Adopting the SBM as the generative model, we propose a new spectral-based algorithm for community detection and show that it achieves consistent recovery. The algorithm and theoretical guarantee can be further extended to the DCSBM, which allows us to interpret the convergence rate when individual (node) heterogeneity is reflected by their degrees and whom they are connected to.

- As a by-product of our analysis, we also propose a new centrality measure for assessing the importance of a node in determining global community structure. We demonstrate this measure captures unique information by comparing it with other popular centrality measures on simulated and real data.

Finally, we note that while this paper focuses on inferring community memberships for all nodes in the network, a natural alternative is to consider only individuals reached by the partial network of knowledge depth L (e.g., individuals 5 and 6 in Figure 1 would be excluded under $L = 1$). However, in this case, the individuals included would be random, adding another layer of complexity to the problem. We leave the theoretical analysis of this case to future work, but still demonstrate empirically our current algorithms can be applied in this situation to a real dataset in Section 5.3.

The rest of the paper is organized as follows. In Section 2, we introduce the setup of our problem and $\mathbf{B}_E$, the signal term of the partial adjacency matrix with knowledge depth $L = 2$. We analyze the spectral properties of $\mathbf{B}_E$ to motivate our detection algorithms and propose a centrality measure for each node based on the magnitude of eigenvalues. Section 3 describes our algorithms and consistency properties under the SBM and DCSBM. We demonstrate the performance of the algorithms and centrality measure using a variety of simulated and real data in Sections 4 and 5 respectively. Finally, we conclude with a discussion section. All the proofs and additional simulation results can be found in the Appendix.

2. Setup and Preliminary Results

2.1 Notations

We first introduce some notations that will be used throughout the paper. For a matrix $\mathbf{M} = (m_{ij})$, in which m_{ij} is the (i,j)-th entry, denote the ith row of $\mathbf{M}$ by $\mathbf{M}(i)$. Let $\|\mathbf{M}\|_{\max} = \max_{i,j}\{|m_{ij}|\}$ and $\|\mathbf{M}\|$ be the spectral norm of $\mathbf{M}$, which is the square root of the largest eigenvalue of $\mathbf{MM}^\top$. Moreover, we denote the Frobenius norm of $\mathbf{M}$ by $\|\mathbf{M}\|_F = \left[\mathrm{tr}(\mathbf{MM}^\top)\right]^{1/2}$. For any random matrix (or vector) $\mathbf{M}$, we use $\mathbb{E}\mathbf{M}$ to denote its expectation. We use $\|\cdot\|_2$ to denote the L_2 norm of a vector. If two positive sequences a_n and b_n satisfy $\limsup_{n\to\infty}(a_n/b_n) < \infty$, we denote $a_n \lesssim b_n$ or $b_n \gtrsim a_n$. If $a_n \lesssim b_n$ and $b_n \lesssim a_n$, we write $a_n \sim b_n$. We write $a_n \ll b_n$ or $b_n \gg a_n$ if $\lim_{n\to\infty}(a_n/b_n) = 0$. For two symmetric matrices $\mathbf{M}_n$ and $\mathbf{N}_n$, if there exists a positive constant c independent of n, $\mathbf{N}_n$ and $\mathbf{M}_n$ such that $\mathbf{N}_n - c\mathbf{M}_n$ is a semi-positive definite matrix, then we write $\mathbf{M}_n \lesssim \mathbf{N}_n$; specifically, if $c = 1$, then we write $\mathbf{M}_n \leq \mathbf{N}_n$. If there exists a positive diverging sequence c_n (i.e., $c_n \to \infty$) such that $\mathbf{N}_n - c_n\mathbf{M}_n \geq 0$ for all n large, we write $\mathbf{M}_n \ll \mathbf{N}_n$. We denote the i-th largest eigenvalue and singular value of $\mathbf{M}$ by $\lambda_i(\mathbf{M})$ and $\sigma_i(\mathbf{M})$, respectively. For any positive integer K, let $[K] = \{1, \ldots, K\}$, $[-K] = \{-K, \ldots, -1\}$, and $[\pm K] = \{-K, \ldots, -1, 1, \ldots, K\}$. Standard order notations such as $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$, $o_p(\cdot)$ are also used. We use $|\cdot|$ to denote the cardinality of a set. Throughout the paper, c and C denote constants that may vary from line to line. Events *with high probability* (w.h.p.) are defined as:

Definition 1 *We say a sequence of events $\mathcal{A}_n$ holds with high probability if for any positive constant D, there exists an $n_0(D) \in \mathbb{N}$ such that $\forall\, n \geq n_0(D)$, $\mathbb{P}(\mathcal{A}_n) \geq 1 - n^{-D}$.*

2.2 The Partial Adjacency Matrix and Signal Term

Recall that $\mathbf{A} = (a_{ij})$ is the $n \times n$ adjacency matrix of $G = (\mathcal{V}, \mathcal{E})$, the full network, in which $a_{ij} = 1$ if $(i,j) \in \mathcal{E}$ and 0 otherwise. Let $K = \mathrm{rank}(\mathbb{E}\mathbf{A})$ and assume that $\mathbf{A} = \mathbf{A}^\top$ and $\{a_{ij}\}_{1 \leq i \leq j \leq n}$ are independent Bernoulli random variables with expectation $\mathbb{E}\mathbf{A}$. We assume that K is a constant and denote the (reduced form) eigendecomposition of $\mathbb{E}\mathbf{A}$ by $\mathbb{E}\mathbf{A} = \mathbf{VDV}^\top$, where $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_K)$ with d_i being the i-th largest eigenvalue (by magnitude) of $\mathbb{E}\mathbf{A}$ and $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_K)$ being the corresponding eigenvector matrix[1].

Without loss of generality, for convenience we assume that the partial network is always *centered around individual 1* in the network. Let $\mathbf{B} = (b_{ij})$ be individual 1's perceived adjacency matrix with knowledge depth $L = 2$. An example of $\mathbf{A}$ and $\mathbf{B}$ is illustrated in the left and middle panels of Figure 1, respectively. More generally, b_{ij} takes the form $b_{ij} = a_{ij}(1 - \mathbb{I}(a_{1i} = 0)\mathbb{I}(a_{1j} = 0))$, $i, j \in [n]$, where $\mathbb{I}(\cdot)$ is an indicator function. Then it follows that

$$\mathbf{B} = -\mathbf{SAS} + \mathbf{AS} + \mathbf{SA}, \quad \text{where} \quad \mathbf{S} = \mathrm{diag}(a_{11}, \ldots, a_{1n}). \tag{1}$$

We further define

$$\mathbf{B}_E = -\mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}(\mathbb{E}\mathbf{A}). \tag{2}$$

---

1. Strictly speaking, this model allows self-loops while real networks often do not. The inclusion of self-loops does not change our conclusions, as explained in the Appendix.

We will next show that $\mathbf{B}_E$ is a signal term of $\mathbf{B}$ and analyze its spectral properties.

Unlike the case of full information network, where $\mathbb{E}\mathbf{A}$ is the signal term of $\mathbf{A}$, the partial information network has a more subtle situation. As shown in the next lemma, the obvious candidates $\mathbb{E}\mathbf{B}$ and $-(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})$ are much smaller than $\mathbf{B}_E$.

Lemma 1 *In the simplest scenario that $\mathbb{P}(a_{ij} = 1) = p_n = o(1)$, for $i, j \in [n]$, we have*

$$\| - \mathbb{E}\mathbf{S}(\mathbb{E}\mathbf{A})\mathbb{E}\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbb{E}\mathbf{S} + \mathbb{E}\mathbf{S}(\mathbb{E}\mathbf{A})\| + \|\mathbb{E}\mathbf{B}\| = o_p(\|\mathbf{B}_E\|). \tag{3}$$

It is well known that spectral clustering relies on the leading eigenvectors. In view of Lemma 1, the leading eigenvalue of $\mathbf{B}_E$ is much larger than those of $\mathbb{E}\mathbf{B}$ and $-(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})$ when $p_n = o(1)$, a typical asymptotic condition for large networks. In other words, the latter two matrices do not contribute to the leading eigenvectors, and neither can be a signal term. To see why this is the case, note that $\mathbb{E}\mathbf{B}$ is almost equal to $-(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})$ (except for differences in the diagonal entries, the first row and first column). Consider $\mathbf{SAS}$ and $(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S})$, the first terms in $\mathbf{B}$ and $\mathbb{E}\mathbf{B}$, respectively. Intuitively, the singular values of $\mathbf{SAS}$ are equivalent to those of $\mathbf{AS}$, given the fact that $\mathbf{S}^2 = \mathbf{S}$. When $p_n = o(1)$, we can see that $(\mathbb{E}\mathbf{S})^2 \ll \mathbb{E}\mathbf{S}$, indicating multiplying $\mathbb{E}\mathbf{A}$ by $\mathbb{E}\mathbf{S}$ twice would significantly deviate it from the target matrix $\mathbf{SAS}$. We refer to our proof of Lemma 1 for more details.

On the other hand, Lemma 3, to be introduced in the next section, shows that under appropriate conditions, the spectral norm $\|\mathbf{B} - \mathbf{B}_E\| \ll$ the smallest (in magnitude) non-zero eigenvalue of $\mathbf{B}_E$. Thus $\mathbf{B}_E$, which is random itself, is a signal term of $\mathbf{B}$ from the spectral point of view. Intuitively, $\mathbf{B}_E$ only differs from the conditional expectation $\mathbb{E}[\mathbf{B}|\mathbf{S}]$ in the first row and column, thus the latter can be considered as one way to interpret $\mathbf{B}_E$. As will be shown in the next section, $\mathbf{B}_E$ has an explicit low-rank structure, lending itself to explicit spectral analysis.

2.3 Spectral Properties of the Signal Term $\mathbf{B}_E$

In this section, we present a few key theoretical properties of $\mathbf{B}_E$ and establish that $\mathbf{B}_E$ is a signal term of $\mathbf{B}$. In addition to showing the order of its eigenvalues, we derive the exact forms of the eigenvectors and approximate forms of the eigenvalues. The results here depend only on generic assumptions about invertibility and the eigendecomposition of $\mathbb{E}\mathbf{A}$.

Recall $\mathbf{S} = \text{diag}(a_{11}, \ldots, a_{1n})$ and $\mathbb{E}\mathbf{A} = \mathbf{VDV}^\top$, where $\mathbf{V}$ is of dimensions $n \times K$. We have the following theorem regarding the spectral properties of $\mathbf{B}_E$.

Theorem 1 *Suppose that $\mathbf{V}^\top\mathbf{SV}$ and $\mathbf{I} - \mathbf{V}^\top\mathbf{SV}$ are invertible. Denote $\mathbf{H}(x) = \mathbf{I} - x\mathbf{DV}^\top\mathbf{SV} - x^2\mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{SV})\mathbf{DV}^\top\mathbf{SV}$. Then the determinant equation*

$$\det(\mathbf{H}(x)) = 0 \tag{4}$$

has $2K$ non-zero real solutions; we denote them by $x_{-K}, \ldots, x_{-1}, x_1, \ldots, x_K$, with $x_i \leq x_j$ for all $i < j$. Moreover, for $i \in [\pm K]$, let $\mathbf{q}_{1i}$ be an eigenvector of $\mathbf{H}(x_i)$ corresponding to the zero eigenvalue, and $\mathbf{q}_{2i} = x_i\mathbf{DV}^\top\mathbf{SV}\mathbf{q}_{1i}$. Then $\mathbf{q}_i$ defined as

$$\mathbf{q}_i = \mathbf{SV}\mathbf{q}_{1i} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2i} \tag{5}$$

is an eigenvector of $\mathbf{B}_E$ corresponding to the eigenvalue x_i^{-1}.

Conversely, if $\mathbf{q}_0 \neq \mathbf{0}$ is an eigenvector of $\mathbf{B}_E$ corresponding to a non-zero eigenvalue x_0^{-1}, then x_0 satisfies (4). Moreover, $\mathbf{q}_0$ can be decomposed (in the form of (5)) as $\mathbf{q}_0 = \mathbf{S}\mathbf{V}\mathbf{q}_{10} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{20}$, where $\mathbf{q}_{10}$ is an eigenvector of $\mathbf{H}(x_0)$ corresponding to the zero eigenvalue and $\mathbf{q}_{20} = x_0 \mathbf{D}\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{10}$.

Finally, we have rank $(\mathbf{B}_E) = 2K$.

Remark 1 *(i) Under our setting, the invertibility assumption on $\mathbf{V}^\top \mathbf{S}\mathbf{V}$ and $\mathbf{I} - \mathbf{V}^\top \mathbf{S}\mathbf{V}$ is not stringent. To see this, note that $\mathbf{V}^\top \mathbf{S}\mathbf{V} = \sum_{i=1}^{n} a_{1i} \mathbf{V}^\top(i) \mathbf{V}(i)$. Since n is large and $\mathbf{V}^\top \mathbf{S}\mathbf{V}$ is a matrix of small dimensionality $K \times K$, the invertibility of $\mathbf{V}^\top \mathbf{S}\mathbf{V}$ should be satisfied if individual 1 has enough neighbors. Similarly, the invertibility of $\mathbf{I} - \mathbf{V}^\top \mathbf{S}\mathbf{V}$ can be ensured if individual 1 does not directly connect to almost everyone else in the network, which is sensible since most real networks are sparse.*

(ii) The rank of $\mathbf{B}_E$ and the form of eigenvectors play an important role in motivating our community detection algorithms in Section 3. The two terms in Eq (5) naturally split the individuals into two subsets: those who are neighbors of individual 1 (represented by $\mathbf{S}$) and those who are not (represented by $\mathbf{I} - \mathbf{S}$). As we will show later, our algorithm first performs clustering on these two subsets separately, before merging them into K communities.

In what follows, we introduce a few conditions to perform further analysis of our partial network. Denote $p_n = \max_{i,j} \mathbb{P}(a_{ij} = 1)$.

Condition 1 $\min_{j \geq 2} \mathbb{P}(a_{1j} = 1) \sim p_n$ *and* $1 - c > p_n \gg \log n / n$ *for some constant $c > 0$.*

Condition 2 $\|\mathbf{V}\|_{\max} \leq C/\sqrt{n}$ *for some constant $C > 0$.*

Condition 3 $|d_1| \sim |d_2| \sim \ldots \sim |d_K| \sim np_n$.

Conditions 1–2 are sufficient to ensure the invertibility of $\mathbf{V}^\top \mathbf{S}\mathbf{V}$ and $\mathbf{I} - \mathbf{V}^\top \mathbf{S}\mathbf{V}$ with high probability, as we will prove in Lemma 2. We note that Conditions 1–2 can be relaxed but we adopt them in the paper for convenience and simplicity. Condition 3 is a strong condition to assume that the magnitude of the smallest non-zero eigenvalue of $\mathbb{E}\mathbf{A}$ has the same order as $\|\mathbb{E}\mathbf{A}\|_F$, which is to ensure a big enough gap between $|d_K|$ and $\|\mathbf{A} - \mathbb{E}\mathbf{A}\|$ for more straightforward analysis. This condition could also be relaxed by deeper analysis, but we leave it for future studies. We have the following invertibility lemma.

Lemma 2 *Under Conditions 1 and 2, there exists a positive constant c such that w.h.p., we have*

$$cp_n \left(1 - \sqrt[4]{\frac{\log n}{np_n}}\right) \mathbf{I} \leq \mathbf{V}^\top \mathbf{S}\mathbf{V} \leq p_n \left(1 + \sqrt[4]{\frac{\log n}{np_n}}\right) \mathbf{I} < \left(1 - \frac{c}{2}\right) \mathbf{I}. \tag{6}$$

Therefore w.h.p., $\mathbf{V}^\top \mathbf{S}\mathbf{V}$ and $\mathbf{I} - \mathbf{V}^\top \mathbf{S}\mathbf{V}$ are invertible.

Theorem 1 and Lemma 2 imply the following corollary.

Corollary 1 *Under Conditions 1 and 2, for suitably chosen* $\mathbf{q}_{1l}$, $l \in [\pm K]$, $\mathbf{q}_l$'s, *as defined in Theorem 1, satisfy w.h.p. that*

$$\mathbf{q}_i^\top \mathbf{q}_j = 0, \quad i \neq j, \ i,j \in [\pm K], \tag{7}$$

and that $\dim(\operatorname{span}\{\mathbf{q}_l, l \in [\pm K]\}) = \operatorname{rank}(\mathbf{B}_E) = 2K$.

Corollary 1 specifies the form of orthogonal eigenvectors in the eigendecomposition of $\mathbf{B}_E$. Even though the eigenvalues in Theorem 1 may have cardinality greater than 1, this does not change the eigenspace of $\mathbf{B}_E$, which is equivalent to $\mathbf{Q}\mathbf{Q}^T$. As far as spectral clustering is concerned, $\mathbf{q}_i$ do not need to be uniquely defined, and separability of eigenvectors for different nodes is typically identifiable up to an orthogonal transformation. As we develop our community detection algorithm under the SBM and DCSBM in Section 3, we will use the eigenvectors of the observed $\mathbf{B}$ and bound the deviation from their counterparts in $\mathbf{B}_E$.

Theorem 2 *Under Conditions 1-3, w.h.p., we have* $|x_i|^{-1} \sim n p_n^{3/2}$ *for* $i \in [\pm K]$. *Furthermore, if* $p_n \to 0$ *and* $\mu_i/\mu_{i+1} \geq 1 + c$ *for some positive constant* c, $1 \leq i \leq K-1$ *where* $\mu_1 \geq \ldots \geq \mu_K$ *are the eigenvalues of* $\mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V}\mathbf{D}(\mathbf{I} - \mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V})\mathbf{D}$, *it holds w.h.p. that,*

$$x_i^{-1} = \left(\lambda_i(\mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V})\right)^{\frac{1}{2}} (1 + o(\sqrt{p_n})) , \ i \in [K],$$

$$x_i^{-1} = -\left(\lambda_{K+i+1}(\mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V})\right)^{\frac{1}{2}} (1 + o(\sqrt{p_n})) , \ i \in [-K].$$

The first part of Theorem 2 shows the order of the non-zero eigenvalues of $\mathbf{B}_E$. The second part, with more stringent conditions, finds the approximate expressions for these eigenvalues. We note in advance that our community detection results do not depend on the additional conditions in Theorem 2, but the approximate expressions are of standalone interest and can be useful to study other problems under the partial information framework (e.g., in construction of centrality measure to be described in Section 3.4).

Finally, the next lemma bounds the spectral norm of $\mathbf{B} - \mathbf{B}_E$ and shows $\mathbf{B}_E$ is indeed a signal term of $\mathbf{B}$.

Lemma 3 *It holds w.h.p. that* $\|\mathbf{B} - \mathbf{B}_E\| \lesssim \sqrt{np_n}$. *Under Conditions 1, 2 and 3, assuming* $p_n \gg \sqrt{1/n}$, *then we have w.h.p.,* $\|\mathbf{B} - \mathbf{B}_E\| \ll \min_{i \in [2K]} \sigma_i(\mathbf{B}_E)$.

3. Community Detection under Partial Information

We propose new spectral-based algorithms for performing community detection in the partial information framework, under the commonly used SBM and DCSBM settings. Using the spectral properties derived in the previous section, we provide upper bounds on the error rates and show that the recovery is almost exact under appropriate conditions. As an application utilizing the theoretical properties of $\mathbf{B}_E$, community detection requires more specialized model assumptions. Therefore the conditions we introduce in this section are sufficient conditions implying the more general Conditions 1–3.

3.1 Stochastic Block Model

In the SBM (Holland et al., 1983), each individual belongs to exactly one of K different communities. The connection probability between two individuals depends on their community memberships. Concretely, in the SBM with K communities, $\mathbb{E}\mathbf{A}$ is given by

$$\mathbb{E}\mathbf{A} = \mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^\top, \tag{8}$$

where $\mathbf{P} = (p_{kl})$ is a symmetric $K \times K$ matrix in which p_{kl} is the connection probability between communities k and l, $\mathbf{\Pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n)^\top \in \mathbb{R}^{n \times K}$ is the matrix of community membership vectors, with individual i's membership vector $\boldsymbol{\pi}_i \in \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$, $\mathbf{e}_k \in \mathbb{R}^K$ is the standard basis vector with the k-th element being one and the other elements being zero. If individuals i, j belong to communities k and l respectively, we have $\mathbb{E}a_{ij} = \boldsymbol{\pi}_i^\top \mathbf{P}\boldsymbol{\pi}_j = p_{kl}$.

Before we present our community detection algorithm, we first need some technical results connecting the SBM with the low-rank model discussed in Section 2. Recall that $p_n = \max_{i,j} \mathbb{P}(a_{ij} = 1)$. Without loss of generality, assume that individual 1 belongs to community 1. The following condition, which is easier to interpret in the SBM setting, implies Conditions 1, 2 and 3 in Section 2. Hence we will use this condition instead in our theoretical analysis under the SBM.

Condition 4 $\min_{k \in [K]} p_{1k} \sim p_n$. $\min_{k \in [K]} \sum_{j \in [n]} \mathbb{I}(\boldsymbol{\pi}_j = \mathbf{e}_k) \geq c_0 n$ and $\sigma_K(\mathbf{P}) \geq c_1 p_n$ for some positive constants c_0 and c_1. Moreover, for some $c > 0$, $1 - c \geq p_n \gg (1/n)^{1/2}$.

Remark 2 (i) Community detection on the entire graph typically requires $np_n \gtrsim \log n$, while our lower bound on the density p_n is $1/\sqrt{n}$, which is more stringent due to significant information loss incurred by observing only a partial network. Note that the order of the largest singular value of $\mathbb{E}\mathbf{A}$ is essentially np_n, while the signal strength in our case is $x_i^{-1} \sim np_n^{3/2}$, $i \in [\pm K]$, as shown in Theorem 2.
(ii) The above condition assumes balanced community sizes to ensure the signal size in $\mathbf{B}_E$ is sufficiently large. As an example of imbalanced community sizes, consider a simple case with $\mathbf{P} = p_n\mathbf{I}_K$ and $\mathbf{\Pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n)^\top$. Here for the first $K - 1$ communities ($1 \leq k \leq K - 1$), $\boldsymbol{\pi}_i = \mathbf{e}_k$, $i \in \{n_1(k - 1) + 1, \ldots, n_1 k\}$; for the last community, $\boldsymbol{\pi}_i = \mathbf{e}_K$, $i \in \{n_1(K - 1) + 1, \ldots, n\}$. Let $n_2 = n - n_1(K - 1)$ and assume $n_2 \ll n$ so that the last community is much smaller than the others. As shown in Theorem 2, $x_K^{-1} \sim n_2 p_n^{3/2}$. Since Lemma 3 shows the order of $\|\mathbf{B} - \mathbf{B}_E\|$ is $\sqrt{np_n}$, for signal dominance over noise, one would require that $n_2 p_n^{3/2} \gg \sqrt{np_n}$, or $p_n \gg \frac{\sqrt{n}}{n_2}$, which is even more stringent than our current lower bound $1/\sqrt{n}$.

Lemma 4 Under the SBM defined in (8), Condition 4 implies $\mathrm{rank}(\mathbb{E}\mathbf{A}) = K$ and Conditions 1, 2 and 3. Moreover, there exists a $K \times K$ matrix $\mathcal{D}$ such that

$$\mathbf{V} = \mathbf{\Pi}\mathcal{D} \quad and \quad \mathcal{D}\mathcal{D}^\top \geq \frac{1}{n}\mathbf{I}. \tag{9}$$

Thus under Condition 4, Corollary 1 allows us to write down an orthonormal eigenvector matrix $\mathbf{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_K, \mathbf{q}_{-1}, \ldots, \mathbf{q}_{-K})$ such that $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ ($\|\mathbf{q}_i\|_2 = 1$). Let $\mathcal{Q}_i = (\mathbf{q}_{i,1}, \ldots, \mathbf{q}_{i,K}, \mathbf{q}_{i,-1}, \ldots, \mathbf{q}_{i,-K})$ for $i = 1, 2$, then by (5), we have

$$\mathbf{Q} = \mathbf{S}\mathbf{\Pi}\mathcal{D}\mathcal{Q}_1 + (\mathbf{I} - \mathbf{S})\mathbf{\Pi}\mathcal{D}\mathcal{Q}_2. \tag{10}$$

The next lemma bounds the distances between the rows of $\mathbf{Q}$ by grouping their row indices according to (i) their community memberships and (ii) whether they are neighbors of node 1. This indicates that both information needs to be taken into account when constructing a community detection algorithm.

Lemma 5 *Under Condition 4, for any $2K \times 2K$ orthogonal matrix $\mathbf{O}$, it holds w.h.p. that for all $i, j \in [n]$,*

$$\boldsymbol{\pi}_i \neq \boldsymbol{\pi}_j \Longrightarrow \|\mathbf{Q}(i)\mathbf{O} - \mathbf{Q}(j)\mathbf{O}\|_2 \geq \sqrt{\frac{2}{c_2 n}}\,, \tag{11}$$

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_j, a_{1i} \neq a_{1j} \Longrightarrow \|\mathbf{Q}(i)\mathbf{O} - \mathbf{Q}(j)\mathbf{O}\|_2 \geq \sqrt{\frac{2}{c_2 n}}\,, \tag{12}$$

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_j, a_{1i} = a_{1j} \Longrightarrow \|\mathbf{Q}(i)\mathbf{O} - \mathbf{Q}(j)\mathbf{O}\|_2 = 0\,. \tag{13}$$

Here c_2 is a positive constant defined in Lemma 8 in the Appendix.

In the case of SBM with the full network, it is well known that $\mathbb{E}\mathbf{A}$, the population version of $\mathbf{A}$, has exactly K distinct rows in its eigenvector matrix, each corresponding to a different community. Sufficient separations between these rows and appropriate concentration of $\mathbf{A}$ around $\mathbb{E}\mathbf{A}$ ensure that spectral clustering works by performing eigendecomposition on $\mathbf{A}$. However, such a simple approach does not work when it comes to $\mathbf{B}$ and $\mathbf{B}_E$. Indeed, (12) in Lemma 5 reveals that even if $\boldsymbol{\pi}_i = \boldsymbol{\pi}_j$, $\mathbf{Q}(i)$ and $\mathbf{Q}(j)$ can be different; (12) and (13) suggest that we should treat the nodes separately according to whether they are direct neighbors of node 1. Motivated by Lemma 5, we next propose our new community detection algorithm.

3.2 Community Detection Algorithm under the SBM

Let $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_K, \mathbf{w}_{-1}, \ldots, \mathbf{w}_{-K})$ be the collection of orthonormal eigenvectors corresponding to the $2K$ largest eigenvalues (in magnitude) of $\mathbf{B}$. Roughly speaking, our approach involves first clustering the non-zero rows of $\mathbf{SW}$ (neighbors of node 1) and $(\mathbf{I}-\mathbf{S})\mathbf{W}$ (non-neighbors) separately before merging them into K communities. As $\mathbf{W}$ is the empirical counterpart of $\mathbf{Q}$, the next lemma bounds the difference between $\mathbf{W}$ and $\mathbf{Q}$.

Lemma 6 *Under Condition 4, w.h.p. we have*

$$\|\mathbf{W} - \mathbf{Q}\mathbf{O}\|_F = O\left(\frac{1}{\sqrt{n p_n}}\right)\,, \tag{14}$$

where $\mathbf{O} = \mathbf{U}_1\mathbf{U}_2^\top$, in which $\mathbf{U}_1$ and $\mathbf{U}_2$ are from the singular value decomposition $(\mathbf{Q})^\top\mathbf{W} = \mathbf{U}_1\Sigma\mathbf{U}_2^\top$ such that Σ is the diagonal matrix with singular values.

Remark 3 *When there are repeated eigenvalues, the eigenvectors are not uniquely determined. The matrix $\mathbf{O}$ in Lemma 6 is created to handle such cases.*

Recall in (10), $\mathbf{S\Pi}\mathcal{D}\mathcal{Q}_1$ and $(\mathbf{I}-\mathbf{S})\mathbf{\Pi}\mathcal{D}\mathcal{Q}_2$ each has at most K different non-zero rows, and these rows partition the node index set $[n]$ into two parts. Hence $\mathbf{Q}$ has at most $2K$

different rows. However, rank($\mathbf{Q}$) $= 2K$ w.h.p. by Corollary 1, implying $\mathbf{Q}$ has exactly $2K$ different rows (each of the two parts in (10) getting K). Since $\mathbf{W}$ is the empirical counterpart of $\mathbf{Q}$, we will apply the k-means algorithm (with $k = K$) to the non-zero rows of the matrices $\mathbf{SW}$ and $(\mathbf{I} - \mathbf{S})\mathbf{W}$, respectively. Here $\mathbf{S}$ and $\mathbf{I} - \mathbf{S}$ separate the individuals into two groups, each returning K clusters. We note that other clustering algorithms could be used here, and we use k-means for ease of analysis. We summarize this procedure in Algorithm 1.

Algorithm 1

1: Take matrices $\mathbf{S}$ and $\mathbf{W}$ as defined respectively in equation (1) and Lemma 6.
2: Apply the k-means algorithm to the rows $\{\mathbf{W}(i) : i \in [n], a_{1i} = 1\}$ and $\{\mathbf{W}(i) : i \in [n], a_{1i} = 0\}$, respectively, to separate each group into K clusters.
3: Return (i) the $2K$ clusters and (ii) the centroid matrix $\mathbf{C} = (\mathbf{c}_1^\top, \ldots, \mathbf{c}_n^\top)^\top$ in which the length $2K$ row vector $\mathbf{c}_i$ denotes the centroid associated with individual i.

In the theoretical analysis of Algorithm 1, for simplicity we assume $\mathbf{C}$ optimizes the k-means objective for both $r = 0$ and $r = 1$, that is,

$$\{\mathbf{c}_i : a_{1i} = r\} = \operatorname*{argmin}_{\left\{\substack{\mathbf{x}_i : \mathbf{x}_i \text{ is } 2K\text{-dimensional row vector} \\ \text{with } a_{1i} = r \text{ and } |\{\mathbf{x}_i\}_{a_{1i} = r}| \leq K}\right\}} \sum_{i \in [n], a_{1i} = r} \|\mathbf{W}(i) - \mathbf{x}_i\|_2^2, \qquad (15)$$

and also assume $|\{\mathbf{c}_l, l \in [n]\}| = 2K$. When $\mathbf{c}_i = \mathbf{c}_j$ and $a_{1i} = a_{1j}$, Algorithm 1 assigns individuals i and j to the same cluster. An error bound on the clustering can be obtained by counting individuals with row in $\mathbf{Q}$ that are far (up to the permutation matrix $\mathbf{O}$) from their corresponding centroids. Similar to Rohe et al. (2011), we define $\mathcal{M}$,

$$\mathcal{M} = \left\{ i \in [n] : \|\mathbf{c}_i - \mathbf{Q}(i)\mathbf{O}\|_2 \geq \frac{1}{\sqrt{2c_2 n}} \right\}, \qquad (16)$$

where c_2 is the positive constant in Lemma 5. In Theorem 3, we will show that $|\mathcal{M}|/n$ controls the misclustering rate in our final algorithm (Algorithm 2).

Algorithm 1 returns $2K$ clusters, and we denote them by $\{\mathfrak{c}_1, \ldots, \mathfrak{c}_K\}$ and $\{\mathfrak{d}_1, \ldots, \mathfrak{d}_K\}$, for nodes in $\{\mathbf{W}(i) : i \in [n], a_{1i} = 1\}$ and $\{\mathbf{W}(i) : i \in [n], a_{1i} = 0\}$, respectively. Noting that the clustering results are only identifiable up to label permutation, our next step is to merge them into K communities. To this end, we first construct two estimates of the connection probability matrix $\mathbf{P}$ defined in (8). We then merge the clusters by finding the correct label permutation that matches these two estimates. Since for $i, j > 1$ and $i \neq j$, $\mathbb{P}(b_{ij} = 1 | a_{1i} = a_{1j} = 1) = \mathbb{P}(b_{ij} = 1 | a_{1i} = 1 - a_{1j} = 1) = \mathbb{P}(a_{ij} = 1)$, two natural estimators of $\mathbf{P}$, denoted by $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$, are defined as

$$\widehat{\mathbf{P}}_{kl}^{\mathbf{S},\mathbf{S}} = \frac{1}{|\widehat{\mathbf{S}}_{k,l}^{(1)}|} \sum_{(i,j) \in \widehat{\mathbf{S}}_{k,l}^{(1)}} b_{ij} \quad \text{and} \quad \widehat{\mathbf{P}}_{kl}^{\mathbf{S},\mathbf{I}-\mathbf{S}} = \frac{1}{|\widehat{\mathbf{S}}_{k,l}^{(2)}|} \sum_{(i,j) \in \widehat{\mathbf{S}}_{k,l}^{(2)}} b_{ij}, \qquad (17)$$

where $\widehat{\mathbf{S}}_{k,l}^{(1)} = \{(i,j) : a_{1i} = a_{1j} = 1, i \in \mathfrak{c}_k, j \in \mathfrak{c}_l\}$, and $\widehat{\mathbf{S}}_{k,l}^{(2)} = \{(i,j) : a_{1i} = 1 - a_{1j} = 1, i \in \mathfrak{c}_k, j \in \mathfrak{d}_l\}$, for $k, l \in [K]$. We aim to find a label permutation of the clusters $\{\mathfrak{c}_1, \ldots, \mathfrak{c}_K\}$ such that the difference between $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$ is minimized after permutation, which

will allow us to merge the clusters. This is equivalent to finding a permutation function $f : [K] \to [K]$ such that $\widehat{\mathbf{P}}^{\mathbf{S,S}}_{(f,f)} \approx \widehat{\mathbf{P}}^{\mathbf{S,I-S}}_{(f,*)}$, where the (i,j)-th entry of $\widehat{\mathbf{P}}^{\mathbf{S,S}}_{(f,f)}$ and $\widehat{\mathbf{P}}^{\mathbf{S,I-S}}_{(f,*)}$ are the $(f(i),f(j))$-th entry of $\widehat{\mathbf{P}}^{\mathbf{S,S}}$ and the $(f(i),j)$ entry of $\widehat{\mathbf{P}}^{\mathbf{S,I-S}}$, respectively. Algorithm 1 and this merging strategy together make up our final algorithm: Algorithm 2. As an example of the merging Step 5, if $\widehat{f}_0$ returned by Step 4 gives $\widehat{f}_0(1) = 2$, then $\mathfrak{c}_1$ and $\mathfrak{d}_2$ are merged into one community in Step 5.

Algorithm 2 Community detection under the SBM

1: Take matrices $\mathbf{S}$ and $\mathbf{W}$ as defined respectively in equation (1) and Lemma 6.
2: Apply Algorithm 1.
3: Calculate the connection probability matrix estimates $\widehat{\mathbf{P}}^{\mathbf{S,S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S,I-S}}$ defined in (17).
4: For all $K!$ permutations f, find $\widehat{\mathbf{P}}^{\mathbf{S,S}}_{(f,f)}$ and $\widehat{\mathbf{P}}^{\mathbf{S,I-S}}_{(f,*)}$ by permuting the rows and columns of $\widehat{\mathbf{P}}^{\mathbf{S,S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S,I-S}}$.
5: Find $\widehat{f}_0 = \arg\min_f \|\widehat{\mathbf{P}}^{\mathbf{S,S}}_{(f,f)} - \widehat{\mathbf{P}}^{\mathbf{S,I-S}}_{(f,*)}\|_F$.
6: Merge the $2K$ clusters into K communities using the permutation $\widehat{f}_0$.
7: Return the community memberships.

To establish the theoretical property of Algorithm 2, we need one more assumption.

Condition 5 *There exists a positive constant c such that for any $i_1 \neq i_2 \in [K]$, there exists $j_1 \in [K]$ with $|p_{i_1 j_1} - p_{i_2 j_1}| \geq c p_n$.*

As estimators of $\widehat{\mathbf{P}}$ in (17) contain noise, Condition 5 ensures there are sufficient separations between different elements of $\mathbf{P}$ to overcome the noise. Algorithm 2 essentially outputs an estimated membership matrix $\widehat{\mathbf{\Pi}} = (\widehat{\boldsymbol{\pi}}_1, \ldots, \widehat{\boldsymbol{\pi}}_n)^\top$; we define its misclustering rate by

$$d(\widehat{\mathbf{\Pi}}, \mathbf{\Pi}) = \min_{\{\mathbf{Z}:\mathbf{Z} \text{ is } K \times K \text{ permutation matrix}\}} \sum_{j=1}^{n} \mathbb{1}\left(\widehat{\boldsymbol{\pi}}_j^\top \mathbf{Z} \neq \boldsymbol{\pi}_j^\top\right)/n. \tag{18}$$

Theorem 3 *Under Conditions 4–5 and $p_n \gg 1/\sqrt{n}$, with probability tending to 1, Algorithm 2 gives,*

$$d(\widehat{\mathbf{\Pi}}, \mathbf{\Pi}) \leq \frac{|\mathcal{M}|}{n} = O\left(\frac{1}{n p_n^2}\right).$$

That is, Algorithm 2 has the almost exact recovery property (c.f. Def 4 of Abbe (2017)).

Under some stronger conditions on the density p_n and the eigenvalues of $\mathbf{B}_E$, we can extend the above result to achieve exact recovery. We refer to Section C.1 in the Appendix for detailed statements and proofs.

3.3 Extension to the Degree-corrected Stochastic Block Model

In this section, we first consider extending Algorithm 2 and the upper bound on misclustering error to the more general DCSBM, followed by a more specific conditional probability setting that enables us to interpret the convergence rate for different individuals in terms

their connection patterns, as motivated by the Karate club network example in the Introduction.

We consider the DCSBM defined as

$$\mathbb{E}(\mathbf{A}|\Theta) = \Theta\mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^\top\Theta\,, \tag{19}$$

where $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n)$ is the set of degree parameters associated with the nodes. Given Θ, the edges in $\mathbf{A}$ are still generated independently as Bernoulli random variables. Unlike the SBM where individuals within the same community have the same connection pattern, the inclusion of Θ allows for degree heterogeneity in the model and the presence of hubs (i.e., nodes with significantly higher degrees than average nodes). For ease of modeling, we assume $\theta_i \in (0, 1]$, $i \in [n]$ are i.i.d. random variables with a distribution function $F(\cdot)$. Also we assume the mean of the distribution satisfies:

Condition 6 $\mathbb{E}(\theta_i) = \theta \sim 1$.

Similar to Lemma 4, we first show that the DCSBM (19) can be written as the low-rank model in Section 2. The proof of Lemma 4 together with Lemma 4.1 of Lei and Rinaldo (2015) gives us the following proposition.

Proposition 1 *Under Conditions 4 and 6, $\mathrm{rank}(\mathbb{E}\mathbf{A}) = K$ and Conditions 1, 2 and 3 hold with high probability. Moreover, there exists a $K \times K$ orthogonal matrix $\mathcal{D}'$ such that*

$$\mathbf{V}(i) = \theta_i \left(\sum_{j \in Community\ k} \theta_j^2 \right)^{-1/2} \mathcal{D}'(k)\,, \qquad i \in Community\ k \tag{20}$$

for all $i \in [n]$.

Thus similar to (10), Theorem 1 allows us to write $\mathbf{Q}$ as

$$\mathbf{Q} = \mathbf{S}\Theta\mathbf{\Pi}\mathcal{D}\mathcal{Q}_1 + (\mathbf{I} - \mathbf{S})\Theta\mathbf{\Pi}\mathcal{D}\mathcal{Q}_2\,, \tag{21}$$

where $\mathcal{D} = \mathrm{diag}(\sum_{j \in \mathrm{Community}\ 1} \theta_j^2, \ldots, \sum_{j \in \mathrm{Community}\ K} \theta_j^2)^{-1/2}\mathcal{D}'$. Under Conditions 4 and 6, it is easy to check the key results in Sections 3.1-3.2, namely Lemmas 2, 3, 6 and the first part of Theorem 2 still hold using Proposition 1. We omit the proofs since they are almost identical.

In Algorithm 2, following Lei and Rinaldo (2015), we replace the k-means algorithm by the spherical k-median algorithm to cluster the non-zero rows of $\mathbf{S}\mathbf{W}$ and $(\mathbf{I} - \mathbf{S})\mathbf{W}$, returning centroids $\mathbf{C} = (\mathbf{c}_1^\top, \ldots, \mathbf{c}_n^\top)^\top$ defined as

$$\{\mathbf{c}_i : a_{1i} = r\} = \mathrm{argmin}_{\left\{\substack{\mathbf{x}_i : \mathbf{x}_i \text{ is } 2K\text{-dimensional row vector} \\ \text{with } a_{1i} = r \text{ and } |\{\mathbf{x}_i\}_{a_{1i}=r}| \leq K}\right\}} \sum_{i \in [n], a_{1i}=r, \mathbf{W}(i) \neq 0} \left\| \mathbf{W}(i)/\|\mathbf{W}(i)\|_2 - \mathbf{x}_i \right\|_2 \tag{22}$$

for $r = 0, 1$. Nodes i and j are assigned to the same cluster when $\mathbf{c}_i = \mathbf{c}_j$ and $a_{1i} = a_{1j}$. Similar to the SBM case, other clustering algorithms can be used here to return K clusters in each group; we adopt the spherical k-median algorithm for ease of analysis.

For the merging step in Algorithm 2, we use the same equations as (17). Roughly speaking, the matrices $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$ now estimate (up to label permutation) the quantities

$$\mathbf{X} := \mathrm{diag}(\bar{\theta}_1^{\mathbf{S}},\ldots,\bar{\theta}_K^{\mathbf{S}})\mathbf{P}\mathrm{diag}(\bar{\theta}_1^{\mathbf{S}},\ldots,\bar{\theta}_K^{\mathbf{S}}) \quad \text{and} \quad \mathbf{Y} := \mathrm{diag}(\bar{\theta}_1^{\mathbf{S}},\ldots,\bar{\theta}_K^{\mathbf{S}})\mathbf{P}\mathrm{diag}(\bar{\theta}_1^{\mathbf{I}-\mathbf{S}},\ldots,\bar{\theta}_K^{\mathbf{I}-\mathbf{S}}),$$

respectively, where $\bar{\theta}_k^{\mathbf{S}}$ is the average degree parameter for nodes $i \in$ Community k and $a_{1i} = 1$, and similarly for $\bar{\theta}_k^{\mathbf{I}-\mathbf{S}}$. To simplify our notation, define a matrix function:

Definition 2 *For any matrices* $\mathbf{X} = (x_{ij})_{1 \le i,j \le K}$ *and* $\mathbf{Y} = (y_{ij})_{1 \le i,j \le K}$, *define*

$$g(\mathbf{X},\mathbf{Y}) = \mathrm{diag}\left(\frac{\sum_{i=1}^K x_{i1}}{\sum_{i=1}^K y_{i1}},\ldots,\frac{\sum_{i=1}^K x_{iK}}{\sum_{i=1}^K y_{iK}}\right).$$

It is easy to see that if $\mathbf{X}$ and $\mathbf{Y}$ have matching community labels, $\mathbf{Y} = \mathbf{X}g(\mathbf{X},\mathbf{Y})$. Based on this observation, we aim to find a permutation f such that $\|\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S},\mathbf{S}} - \widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}g(\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S},\mathbf{S}},\widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}})\|_F$ is minimized. This leads to the following algorithm for the DCSBM case.

Algorithm 3 Community detection under the DCSBM

1: Take matrices $\mathbf{S}$ and $\mathbf{W}$ as defined respectively in equation (1) and Lemma 6.
2: Apply the spherical k-median algorithm (22) to the non-zero rows of $\mathbf{SW}$ and $(\mathbf{I}-\mathbf{S})\mathbf{W}$ respectively.
3: Calculate the matrix estimates $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$ defined in (17).
4: For all $K!$ permutations f, find $\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$ by permuting the rows and columns of $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$.
5: Find $\widehat{f}_0 = \arg\min_f \|\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S},\mathbf{S}} - \widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}g(\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S},\mathbf{S}},\widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}})\|_F$.
6: Merge the $2K$ clusters into K communities using the permutation $\widehat{f}_0$.
7: Return the community memberships.

Let $f_s = \inf\{x, F(x) > 0\}$. Using arguments similar to Theorem 3 for the SBM case and Theorem 4.2 in Lei and Rinaldo (2015), we can show that Algorithm 3 returns a misclustering rate (defined in (18)) with the following upper bound.

Theorem 4 *Under Conditions 4–6 and* $p_n \gg \frac{1}{\sqrt{n}f_s}$, *with probability tending to 1, Algorithm 3 gives*

$$d(\widehat{\mathbf{\Pi}},\mathbf{\Pi}) = O\left(\frac{1}{\sqrt{n}f_s p_n}\right).$$

The above theorem extends our result in Section 3.2 to DCSBM. However, since (within the same community) the degree parameters for individuals and their edges are generated identically under the full model, the unconditional probability setting still gives the same upper bound on misclustering rate for all individuals.

In what follows, we consider a setting that sheds more light on how the structure around a given individual affects their ability to discern global community memberships through their partial network. More specifically, we consider i) the conditional probability given a neighborhood $\mathbf{S}$ around individual 1, and ii) a two-component mixture distribution on θ_i, where one component represents the "hub" nodes with denser connections than average individuals.

Condition 7 θ_i's are i.i.d. random variables with CDF $yF_1(x) + (1-y)F_2(x)$, $y \in (0,1)$. $\mathbb{E}(\theta_i) = \theta = y\mu_1 + (1-y)\mu_2 \sim 1$, where $\mu_2 \sim 1$ and $\mu_1 \leq \mu_2$. The support of each F_j is bounded, that is, $1 > \sup\{x, F_j(x) \in (0,1)\} > \inf\{x, F_j(x) \in (0,1)\} \sim \mu_j$, $j = 1, 2$.

Remark 4 F_1 and F_2 represent degree distributions of the non-hub and hub nodes respectively. When $F_1 = F_2$ (and $\mu_1 = \mu_2$), we recover the original DCSBM. Note that $\mu_1 \ll \mu_2$ is allowed as long as y is bounded away from 1. The component F_2 with potentially a much larger mean μ_2 captures the effect of hub nodes. We also note that although for the sake of consistency, we have generated individual 1 and their edges following the same DCSBM, it is possible to relax this constraint as our arguments are conditioned on fixed $\mathbf{S}$.

We next analyze the misclustering rate of Algorithm 3 conditioned on $\mathbf{S}$ and θ_i's coming from high probability events. First, it is easy to see that by concentration (Lemma 17 in the Appendix) $|\{\theta_i \sim \mu_2\}| \sim n(1-y)$ with high probability. Hence we always assume θ_i's belong to this event in the following arguments. Next, for the non-zero entries of $\{a_{1i}\}$ in $\mathbf{S}$, let $n_{jk} = |\{i : a_{1i} = 1, \theta_i$ generated from F_j, and $i \in$ Community $k\}|$ for $k = 1, \ldots, K$ and $j = 1, 2$. We assume n_{1k} and n_{2k} satisfy the following condition.

Condition 8 $\min_{k,k' \in [K]} \frac{n_{2k}}{n_{2k'}} \sim 1$, $\min_{k \in [K]} \sum_{j \in [n]} \mathbb{I}(\boldsymbol{\pi}_j = \mathbf{e}_k) \geq c_0 n$ and $\sigma_K(\mathbf{P}) \geq c_1 p_n$ for some positive constants c_0 and c_1. Moreover, for some $c > 0$, $1 - c \geq p_n \min_{k \in [K]}(n_{2k}\mu_2^2 + n_{1k}\mu_1^2)/n \gg 1/n$.

We note that the last condition, $1 - c \geq p_n \min_{k \in [K]}(n_{2k}\mu_2^2 + n_{1k}\mu_1^2)/n \gg \frac{1}{n}$, can be reduced to $p_n \gg \frac{1}{\sqrt{n}}$ in Condition 4 for SBM since $\min_{k \in [K]}(n_{2k}\mu_2^2 + n_{1k}\mu_1^2)/n \sim p_n$ with high probability.

The following lemma is analogous to Lemma 2 but for the conditional setting given $\mathbf{S}$.

Lemma 7 Under Conditions 7 and 8, for any given $\mathbf{S}$ we have

$$\min_{1 \leq k \leq K} \frac{n_{1k}\mu_1^2 + n_{2k}\mu_2^2}{n} \mathbf{I} \lesssim \mathbf{V}^T \mathbf{S} \mathbf{V} \lesssim \max_{1 \leq k \leq K} \frac{n_{1k}\mu_1^2 + n_{2k}\mu_2^2}{n} \mathbf{I}.$$

The other key results in Sections 3.1-3.2 also can be checked in a similar way and adapted to a given $\mathbf{S}$, and we leave the details to the Appendix. Finally, we have the following bound on misclustering rate which depends on the neighborhood features in $\mathbf{S}$.

Theorem 5 Under Conditions 5, 7 and 8, for $\mu_1^{-2} \frac{1}{\min_{k \in [K]}(n_{2k}\mu_2^2 + n_{1k}\mu_1^2)} \ll p_n$, conditioned on $\mathbf{S}$ and Θ, with probability tending to 1, the misclustering rate of Algorithm 3 is

$$d(\widehat{\mathbf{\Pi}}, \mathbf{\Pi}) = O\left(\mu_1^{-1} \sqrt{\frac{1}{p_n \min_{k \in [K]}(n_{2k}\mu_2^2 + n_{1k}\mu_1^2)}}\right).$$

Remark 5 In this conditional setting, the bound depends on the quantity $\min_{k \in [K]} n_{2k}\mu_2^2 + n_{1k}\mu_1^2$ which describes the neighborhood structure around individual 1. When the total number of neighbors the individual knows in each community is fixed (i.e., $n_{1k} + n_{2k}$), a larger $\min_{k \in [K]} n_{2k}$ leads to a smaller bound on misclustering, suggesting knowing more powerful neighbors across all communities enables the individual to have a better understanding of the global community structure.

3.4 Centrality Measure

Our analysis above implies individuals may have different capacities for community detection in the full network using their local information. Hence a natural question is whether we can quantify this capacity of each individual with a centrality measure. We note that here, we need to switch the perspective from a local one to a global one that ranks all the individuals. As such, the centrality measure we propose below will depend on the full network $\mathbf{A}$, like many existing commonly used centrality measures such as eigenvector centrality and betweenness centrality.

Recall the explicit form of eigenvalues in Theorem 2; together with the bounds in Lemma 2, it suggests that $\lambda_{\min} = \lambda_K(\mathbf{V}^\top \mathbf{S} \mathbf{V})$ determines the gap between the smallest eigenvalue (in magnitude) and 0, and consequently how easy it is to perform spectral clustering. Therefore, $\lambda_{\min}$ captures the amount of clustering information available through the centered individual's partial network, measuring their importance in the whole network in the clustering context. In other words, $\lambda_{\min}$ can be seen as a centrality measure for the centered individual. Empirically, an estimate $\widehat{\lambda}_{\min}$ can be computed using the empirical version of $\mathbf{V}$ from $\mathbf{A}$. It is easy to check that both $\lambda_{\min}$ and $\widehat{\lambda}_{\min}$ lie between 0 and 1.

Under the conditional setting discussed in Section 3.3, $\lambda_{\min}$ behaves like $\min_{k\in[K]} \frac{n_{1k}\mu_1^2 + n_{2k}\mu_2^2}{n}$ by Lemma 7. Hence the same interpretation as in Remark 5 applies to $\lambda_{\min}$ as a measure of how "central" an individual is in the context of global community detection. That is, knowing more powerful neighbors across all communities makes an individual more "important".

Finally, we note the connection between $\lambda_{\min}$ and the well-known eigenvector centrality. It is easy to see that when $K = 1$, $\lambda_{\min} = \sum_{i:a_{1i}=1} v_i^2$, where v_i is the eigenvector centrality of node i. In this sense, $\lambda_{\min}$ is related to both degree and eigenvector centrality. In the following sections, as part of the numerical analysis, we compare $\lambda_{\min}$ with other commonly used centrality measures and show that it correlates well with clustering accuracy.

4. Simulation Studies

In this section, we consider simulating a SBM and a DCSBM with the following parameters:

- Model 1 (SBM): $\mathbf{P} = \begin{pmatrix} 3q & q \\ q & 3q \end{pmatrix}$ and each group is of size $n/2$.

- Model 2 (DCSBM): the same $\mathbf{P}$ and group proportions as the above are used. We simulate an i.i.d. uniform and a mixture distribution for θ_i, to be specified below.

Additional results for a SBM with $K = 3$ are presented in Section E.3 of the Appendix.

4.1 Results for Model 1

For Model 1, we vary the number of individuals $n \in \{300, 600, 900, 1200, 1500, 1800, 2100\}$ and the edge density $q \in \{.1, \sqrt{\log n/n}, (\log n/n)^{1/4}/2, 1/\sqrt{n}\}$. For every combination of n and q, we simulate 100 datasets; Algorithm 2 is applied to the partial network centered at node 1 in each dataset.

We first check the number of individuals and the fraction of edges observed within the partial network centered at individual 1. Tables 3 and 4 in the Appendix show that although

partial network can reach almost everyone in the network, the fraction of missing edges is quite significant ($> 50\%$ in most cases).

We calculate the mean misclustering rate for each parameter combination and report the results in Figure 3(a). Except for $q = 1/\sqrt{n}$, all the other q's lead to almost perfect clustering. As a comparison, we apply spectral clustering to $\mathbf{A}$, the full network, and plot the results in Figure 3(b). Unsurprisingly, perfect clustering is achieved for all the parameter settings since the q's we consider lie in the exact recovery regime, which requires edge density to be at least $\log n/n$ (Abbe, 2017).
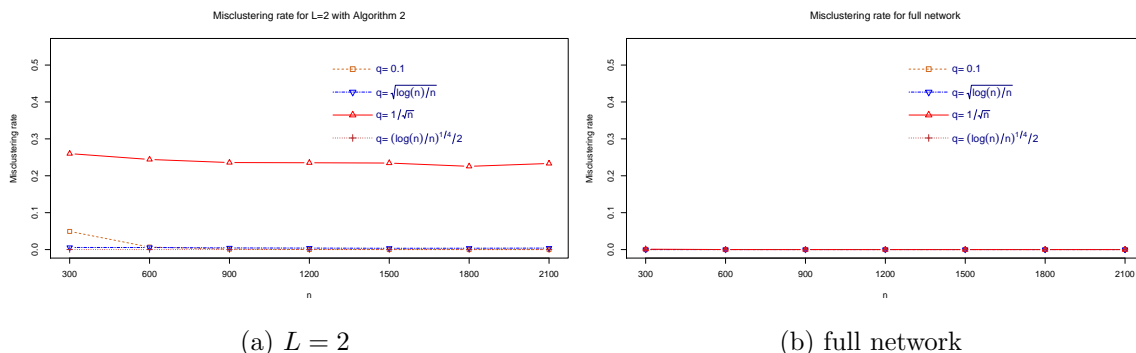

(a) $L = 2$ (b) full network

Figure 3: Misclustering rate for Model 1 ($L = 2$), (full network), averaged over 100 datasets for each combination.

In Figures 8 and 9 of the Appendix, we provide visualizations of two eigenvectors of an exemplary $\mathbf{B}$ matrix corresponding to the two positive eigenvalues (recall that $\mathbf{B}_E$ has two positive and two negative eigenvalues for $K = 2$). The nodes are separated into four clusters, based on their community memberships and whether they are neighbors of node 1, hence justifying our overall approach of clustering.

4.2 Results for Model 2

First generating θ_i from an i.i.d. Unif(.5, 1.5) distribution, we check how the clustering accuracy varies with respect to the edge density q and the number of nodes n, using the same ranges of values as in Model 1. Applying Algorithm 3 to node 1 in 100 simulated networks for each combination of q and n, Figure 4a shows the overall trends are consistent with Figure 3a, although the convergence rates are slower as expected.

Next considering the mixture setting, we take F_1, F_2 as the Unif(0.5, 0.75), Unif(0.8, 1.05) distribution respectively, with proportions $(0.85, 0.15)$. Taking $q = 0.2$, $n = 600$, we generate one instance of $\mathbf{A}$ under Model 2 and apply Algorithm 3 to all the nodes in the network. Figure 4b reports the Pearson and Spearman correlations between the clustering accuracy and various centrality measures for all the nodes. For comparison with the empirical version of $\lambda_{\min}$ (defined in Section 3.4), we choose degree centrality, fraction of edges and individuals observed in the partial network, eigenvector and betweenness centrality as alternative measures. $\widehat{\lambda}_{\min}$ exhibits the highest correlations with accuracy. Furthermore, $\widehat{\lambda}_{\min}$ is less correlated with the other four measures than those measures among themselves (Figure 11 in the Appendix), which suggests it offers unique information about node importance. We

17

note here that same as Model 1, Model 2 produces networks dense enough that all partial networks can reach almost every node in the full network. For this reason, fraction of individuals observed is not a meaningful centrality measure in this case and has almost zero correlation with all the other centrality measures.
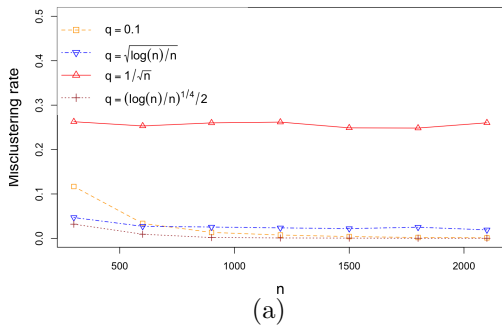
	Pearson	Spearman
degree	0.602	0.679
fraction of edges	0.624	0.696
eigen centrality	0.607	0.684
betweenness	0.581	0.700
$\widehat{\lambda}_{\min}$	0.742	0.822

(a)   (b)

Figure 4: (a) Misclustering rate for Model 2 ($L = 2$), averaged over 100 datasets for each combination, θ_i following an i.i.d. uniform distribution. (b) Correlations between clustering accuracy and various centrality measures, θ_i following a mixture distribution.

5. Real Data Analysis

5.1 Zachary's Karate Club Data

The karate club network (Zachary, 1977) mentioned in the Introduction is a well-known dataset in network analysis. The network consists of 34 individuals, 78 edges, and two ground truth communities which formed after conflicts between the instructor (node "H") and the administrator (node "A"). As shown in Figure 2b, individual 20, despite their small degree and low fraction of observed edges, outperforms the two hub nodes in detection accuracy. This trend is consistent with the empirical $\widehat{\lambda}_{\min}$ values, which are .276, .117, .120 for node 20, "H" and "A" respectively. We have also applied Algorithm 2 to the same set of individuals, and the results are presented in Table 7 in the Appendix. We note that even though the accuracy rates are different for some individuals if we assume the underlying model is SBM instead of DCSBM, the qualitative conclusions about node 20 and the two hub nodes still hold. We continue the rest of our discussion using DCSBM and Algorithm 3, since the presence of two clear hubs is indicative of degree heterogeneity.

We conjecture that the high accuracy of individual 20 is due to their direct connections to "H" and "A". In other words, knowing powerful neighbors in both communities enables this individual to have a better understanding of the global community structure. As further evidence, we delete the edge between node 20 and "A" and apply Algorithm 3 again to the same set of individuals. There is a noticeable drop in accuracy for individual 20, while the other individuals change by a small margin or are unaffected.

individual of interest	H	2	3	A	20	32
detection accuracy	.529	.706	.941	.706	.676	.824

Table 1: Results for chosen individuals after deleting the edge between "A" and 20.

5.2 Microfinance in Indian Villages

This dataset contains information about the social interactions and the diffusion of information about a microfinance program in 43 Indian villages (Banerjee et al., 2013; Cheng et al., 2021). These villages are far apart from each other and can be regarded as independent social networks. Following Banerjee et al. (2013), in each village, we take households as nodes; each edge is undirected and binary representing any of the 12 relationships collected in the survey (e.g., borrowing / lending money or material goods). We use the caste information available in the survey for households as the ground truth community labels. As shown in Figure 14 of the Appendix, many of the networks exhibit assortative structure with respect to these community labels. After a simple filtering step (described in Section F.2 of the Appendix), we analyze 39 villages, with the number of households varying between 24-155 and K between 2-4.

We apply Algorithm 3 to each household in all the villages. Each village also has several predefined leaders expected to be well-connected within the village, who served as the "injection" points for information about the microfinance program. To compare the awareness of global community structure among leaders vs. non-leaders, we compute the mean clustering accuracy among these two groups for each village. In most villages, the mean accuracy of leaders is higher than that of the non-leaders (Figure 5a), thus to an extent justifying the choice of these leaders as the injection point for spreading information.

Next we examine the usefulness of $\widehat{\lambda}_{\min}$ as a centrality measure in this dataset. Banerjee et al. (2013) proposed two centrality measures for assessing the efficiency of information spread from a node: the first is derived from their full structural model, while the second is a simpler diffusion centrality computed using only network topology. The two measures were shown to be highly correlated at village level; we include the diffusion centrality for comparison since it is much easier to compute. In Figure 5b, we calculate the correlations at village level between the mean clustering accuracy and mean centrality measures among the leaders and non-leaders. $\widehat{\lambda}_{\min}$ is the most correlated measure in most cases, although betweenness centrality performs better for leaders using the Spearman correlation.

Finally, in Figure 6, we show that $\widehat{\lambda}_{\min}$ is correlated with the village-level participation rate in the program. Here, the centrality measures are computed for all leaders in each village, who were responsible for spreading the information about the program, before an average is taken for each village. The correlation suggests the information $\widehat{\lambda}_{\min}$ captures extends beyond community detection and is related to information diffusion. Although unsurprisingly, the correlation is weaker than that of the diffusion centrality, which was designed to explain the participation rates in the original paper. The correlations with other centrality measures are shown in Figure 15 in the Appendix. While some of the other centrality measures also exhibit a positive correlation with the participation rate, $\widehat{\lambda}_{\min}$ remains one of the best fitting under a linear model.
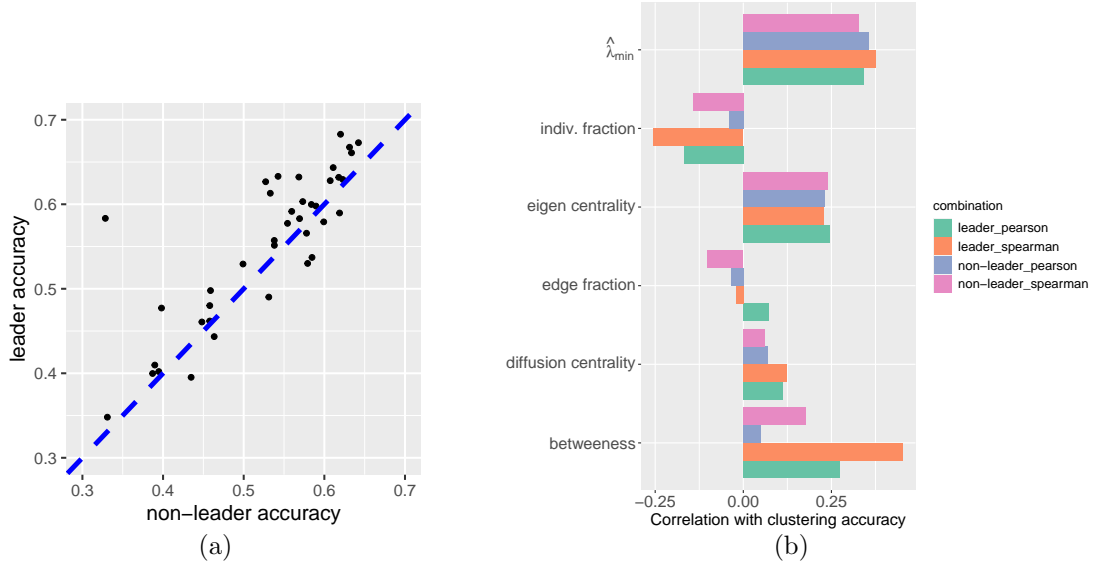
Figure 5: (a) Mean clustering accuracy among leaders vs. non-leaders for all villages. (b) Correlations (Pearson and Spearman) across all villages between the mean clustering accuracy and mean centrality measures for leaders and non-leaders.
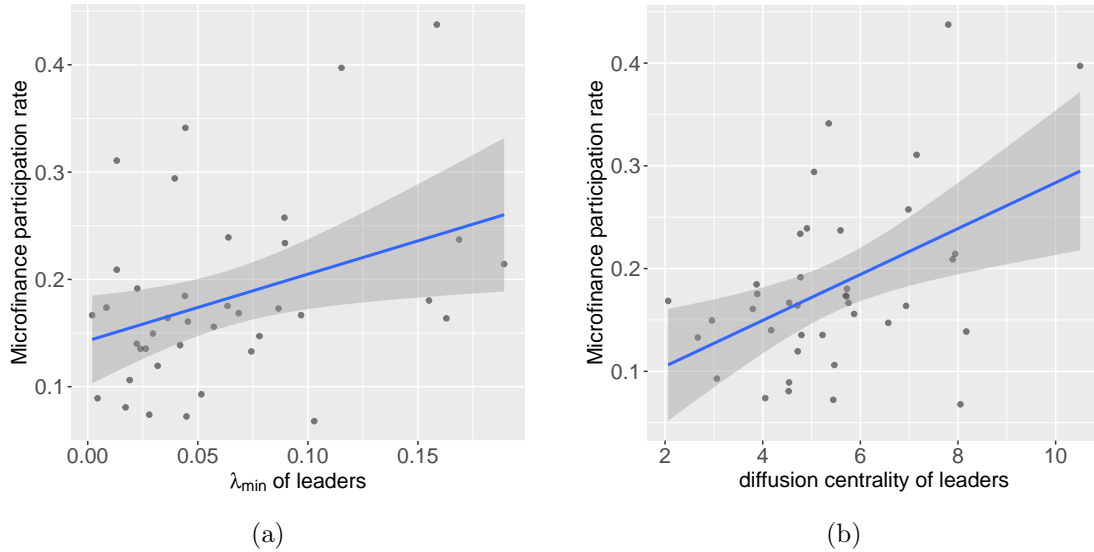

Figure 6: Program participation rate as a function of (a) $\widehat{\lambda}_{\min}$, fitted slope 0.621 with p-value 0.022; (b) diffusion centrality, fitted slope 0.022 with p-value 0.003. The centrality measures are calculated for leaders in each village and an average is taken across each village.

node of interest	1073	1074	1075	1076	1077	1078
the ratio of the edges observed	.1145	.1295	.3362	.1553	.2761	.4116
# of the nodes observed	476	485	880	715	808	793
clustering accuracy	.5798	.5505	.9261	.9021	.8936	9177

Table 2: The network information and detection results for chosen individuals (blogs) in the political blog data using Algorithm 3.

5.3 Political Blog Data

The political blog network (Adamic and Glance, 2005) records hyperlinks between web blogs observed in the run-up to the 2004 U.S. presidential election. Each blog was labeled as democratic or republican based on their political outlooks. Following most previous community detection studies using this dataset, we take the largest (weakly) connected component consisting of 1222 nodes and treat the hyperlinks as undirected.

Similar to the karate club data, we pick six individuals (blogs) and examine their network information and clustering performance in detail. Because this is a large network with relatively sparse edges, most of the partial networks with $L = 2$ can only reach a fraction of all the nodes. Therefore we apply Algorithms 2 and 3 to nodes reached by each partial network only. That is, for a given partial network, the nodes with no observed edges are removed. Table 2 summarizes the performance of Algorithm 3. The results from Algorithm 2 are similar and presented in Table 8 of the Appendix. In this setting, the accuracy rate is calculated using only nodes included in each partial network; similarly the ratio of edges is calculated out of the subnetwork after removing isolated nodes in the partial network. These observations again demonstrate the importance of identifying which nodes contain the most powerful partial information about community structure. Moreover, it shows our algorithms can be successfully applied to recover local community structure by restricting attention to nodes reached by the partial network only, even though our theoretical analysis has focused on global community memberships.

6. Discussion

Each individual in a social network only has a local understanding of the full network, which has motivated us to study the problem of inferring global community memberships using such a local view. In contrast to the popular line of works that performs statistical inference of global network properties by patching together a large number of local subnetworks, we focus on what is attainable using one subnetwork only. This perspective has also allowed us to propose a new centrality measure that assesses the importance of an individual's partial information in determining global community structure. As shown in our analysis of simulated and real networks, this measure is capable of identifying information complementary to other commonly used centrality measures.

We have considered three different types of recovery overall. On the theoretical front, we have analyzed weak recovery and exact recovery of community memberships for all nodes, both can be considered as a form of global recovery. In practice, when the network is sparse, we can apply our algorithms to only individuals reached by the partial network and compute

the accuracy rate on this subset of nodes, as we have done in the political blog data analysis in Section 5.3. This is a form of local recovery, which is more challenging to analyze since the partial network contains random individuals. A meaningful future direction is to extend our theoretical analysis and recovery guarantees to this setting. It is also plausible to consider a hierarchical form of recovery, which becomes weaker for layers of individuals further away from the center. This would be relevant if we believe the connectivity probability decreases as L increases.

Another more specific question is, can the individual of interest correctly identify their own membership? From a practical point of view, after one assigns everyone else into K communities, they can decide on their own affiliation based on side information. Alternatively, we could analyze the asymptotic properties of the eigenvectors of $\mathbf{B}$ and design a correction step accordingly. This is a point that we would like to pursue in future studies.

There are many interesting topics under the individual-centered partial information framework. We end our paper with five open problems to inspire further work. (1) A data-driven way of choosing the number of communities K: one possible solution is to study the differences between spiked eigenvalues and non-spiked eigenvalues (e.g., Fan et al. (2019b); Cai et al. (2020)), but other approaches are also possible. (2) It is possible that the individual of interest knows their edges exactly but can only identify their neighbors' edges with a high probability, then the community detection algorithm should account for the noise in these edges. (3) Extend our analysis to the more expansive $L = 3$ partial information networks. (4) In addition to community detection, we can also study other problems, such as link prediction, mixed membership profile inference, subgraph counts, under the current framework. (5) How to combine multiple, but possibly finite, individuals' partial information and collectively gain a better understanding of the full network is also an interesting problem to explore.

Acknowledgments

Appendix

The appendix contains the proofs of main results and some established results we cited in the proofs for readers' convenience. Additional results from simulation and real data are also presented.

Appendix A. Proofs of the results in Section 2

A.1 Proof of Lemma 1

To show (3), it suffices to prove

$$\| - \mathbb{E}\mathbf{S}(\mathbb{E}\mathbf{A})\mathbb{E}\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbb{E}\mathbf{S} + \mathbb{E}\mathbf{S}(\mathbb{E}\mathbf{A})\| = o_p(\|\mathbf{B}_E\|), \ \|\mathbb{E}\mathbf{B}\| = o_p(\|\mathbf{B}_E\|). \tag{23}$$

By the definition of $\mathbf{S}$, a_{ij} is independent of $\mathbf{S}$ for $i \geq 2$ and $j \geq 2$. By (2.2),

$$\mathbb{E}b_{ij} = \mathbb{E}a_{ij}(1 - (1 - \mathbb{E}a_{1i})(1 - \mathbb{E}a_{1j})), \ 2 \leq i \neq j \leq n.$$

Therefore $\mathbb{E}\mathbf{B}$ is equal to $-(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})$ except for the diagonal entries, the first row and first column. Then we have

$$\|\mathbb{E}\mathbf{B} - (-(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) + (\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A}))\| \lesssim (\sum_{j=1}^{n}(\mathbb{E}a_{1j})^2)^{1/2} + p_n \leq (\sqrt{n}+1)p_n.$$

Therefore, it suffices to prove the first inequality of (23).

First of all, we look at the matrix

$$(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}).$$

Since $(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S}) = (\mathbb{E}\mathbf{S})\mathbf{V}\mathbf{D}\mathbf{V}^\top(\mathbb{E}\mathbf{S})$, $(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S})$ has the same non-zero eigenvalues as $\mathbf{D}\mathbf{V}^\top(\mathbb{E}\mathbf{S})^2\mathbf{V}$. The counterpart of $(\mathbb{E}\mathbf{S})(\mathbb{E}\mathbf{A})(\mathbb{E}\mathbf{S})$ in $\mathbf{B}_E$ is $\mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S}$, whose non-zero eigenvalues are the same as $\mathbf{D}\mathbf{V}^\top\mathbf{S}^2\mathbf{V}$. In this case $\mathbf{V}^\top(\mathbb{E}\mathbf{S})^2\mathbf{V} = p_n^2\mathbf{I}$ and $\mathbb{E}(\mathbf{V}^\top\mathbf{S}^2\mathbf{V}) = p_n\mathbf{I}$, which means that we cannot replace $\mathbf{S}$ by $\mathbb{E}\mathbf{S}$ for $\mathbf{D}\mathbf{V}^\top\mathbf{S}^2\mathbf{V}$. Similarly, we can show that $\mathbb{E}\|(\mathbb{E}\mathbf{A})\mathbf{S}\|_F^2 = \text{tr}(\mathbb{E}\mathbf{A}\mathbb{E}(\mathbf{S}^2)\mathbb{E}\mathbf{A}) = n^2p_n^3 \gtrsim \|(\mathbb{E}\mathbf{A})\mathbb{E}\mathbf{S}\|_F^2 = n^2p_n^4$. These insights combined with Condition $p_n = o(1)$ imply that

$$\| - \mathbb{E}\mathbf{S}(\mathbb{E}\mathbf{A})\mathbb{E}\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbb{E}\mathbf{S} + \mathbb{E}\mathbf{S}(\mathbb{E}\mathbf{A})\| = o_p(\|\mathbf{B}_E\|).$$

A.2 Proof of Theorem 1

Note that $\det(y^2\mathbf{I} - y\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V} - \mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V})$ is a polynomial of y with degree $2K$. Hence the equation $\det(y^2\mathbf{I} - y\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V} - \mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V}) = 0$ has $2K$ solutions in $y \in \mathbb{C}$. Moreover, as $\mathbf{V}^\top\mathbf{S}\mathbf{V}$ and $\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V}$ are invertible, $\det(\mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V}) \neq 0$; hence $y = 0$ is NOT a solution. Let $x = y^{-1}$, then $x^{-2K} \times \det(\mathbf{H}(x)) = \det(y^2\mathbf{I} - y\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V} - \mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top\mathbf{S}\mathbf{V})$. Hence there are $2K$ non-zero solutions to $\det(\mathbf{H}(x)) = 0$ (i.e., (4)). Denote these solutions by $x_{-K}, \ldots, x_{-1}, x_1, \ldots, x_K$.

Then it remains to prove that for each $i \in [\pm K]$, $\mathbf{q}_i$ is an eigenvector of $\mathbf{B}_E$ corresponding to the eigenvalue x_i^{-1}, $\mathbf{q}_i \neq \mathbf{0}$, and $x_i \in \mathbb{R}$. By the definitions of $\mathbf{q}_{1i}$ and $\mathbf{q}_{2i}$, we have

$$
\begin{aligned}
\Big(&- \mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}(\mathbb{E}\mathbf{A}) \Big)(\mathbf{S}\mathbf{V}\mathbf{q}_{1i} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2i}) \\
&= \mathbf{V}\mathbf{D}\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{1i} + x_i \mathbf{S}\mathbf{V}\mathbf{D}(\mathbf{I} - \mathbf{V}^\top \mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{1i} \\
&= x_i^{-1}(\mathbf{V}\mathbf{q}_{2i} + \mathbf{S}\mathbf{V}(\mathbf{q}_{1i} - \mathbf{q}_{2i})) = x_i^{-1}(\mathbf{S}\mathbf{V}\mathbf{q}_{1i} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2i}),
\end{aligned}
\tag{24}
$$

where the second equation follows from

$$
\mathbf{q}_{1i} - x_i \mathbf{D}\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{1i} - x_i^2 \mathbf{D}(\mathbf{I} - \mathbf{V}^\top \mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{1i} = \mathbf{0}.
\tag{25}
$$

Therefore $\mathbf{q}_i = \mathbf{S}\mathbf{V}\mathbf{q}_{1i} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2i}$ is the eigenvector of $-\mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}(\mathbb{E}\mathbf{A})$ corresponding to the eigenvalue x_i^{-1} if $\mathbf{q}_i \neq \mathbf{0}$. We prove $\mathbf{q}_i \neq \mathbf{0}$ by contradiction. Actually, if $\mathbf{q}_i = 0$, by the definition of $\mathbf{q}_i$, we have

$$
0 = \|\mathbf{q}_i\|_2^2 \geq \mathbf{q}_{1i}^\top \mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{1i} = 0.
$$

Then $(\mathbf{V}^\top \mathbf{S}\mathbf{V})^{1/2}\mathbf{q}_{1i} = \mathbf{0}$ and by (25), we have

$$
\mathbf{q}_{1i} = x_i \mathbf{D}\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{1i} + x_i^2 \mathbf{D}(\mathbf{I} - \mathbf{V}^\top \mathbf{S}\mathbf{V})\mathbf{D}\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{q}_{1i} = \mathbf{0},
$$

which contradicts with $\|\mathbf{q}_{1i}\|_2 = 1$! Finally, since $-\mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}(\mathbb{E}\mathbf{A})$ is a real symmetric matrix, its eigenvalues $x_{-K}, \dots, x_{-1}, x_1, \dots, x_K$ are real numbers. Without loss of generality, we can take $x_i \leq x_j$ for all $i < j$.

Recalling that $\mathbb{E}\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\top$ we have

$$
\mathbb{E}\mathbf{A} - (\mathbf{I} - \mathbf{S})\mathbb{E}\mathbf{A}(\mathbf{I} - \mathbf{S}) = (\mathbf{V}, (\mathbf{I} - \mathbf{S})\mathbf{V})\operatorname{diag}(\mathbf{D}, -\mathbf{D})(\mathbf{V}, (\mathbf{I} - \mathbf{S})\mathbf{V})^\top.
\tag{26}
$$

By simple algebra, the non-zero eigenvalues of $(\mathbf{V}, (\mathbf{I} - \mathbf{S})\mathbf{V})\operatorname{diag}(\mathbf{D}, -\mathbf{D})(\mathbf{V}, (\mathbf{I} - \mathbf{S})\mathbf{V})^\top$ are equal to the non-zero eigenvalues of

$$
\operatorname{diag}(\mathbf{D}, -\mathbf{D})(\mathbf{V}, (\mathbf{I} - \mathbf{S})\mathbf{V})^\top (\mathbf{V}, (\mathbf{I} - \mathbf{S})\mathbf{V}) = \begin{pmatrix} \mathbf{D} & 0 \\ 0 & -\mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} \\ \mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} & \mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} \end{pmatrix}.
$$

Since $\mathbf{V}^\top \mathbf{S}\mathbf{V}$ and $\mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V}$ are invertible , we have

$$
\begin{aligned}
\det &\left(\begin{pmatrix} \mathbf{I} & \mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} \\ \mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} & \mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} \end{pmatrix} \right) = \det(\mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} - (\mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V})^2) \\
&= \det(\mathbf{V}^\top \mathbf{S}\mathbf{V}\mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V}) \neq 0.
\end{aligned}
$$

Combining this with (26), we have

$$
\operatorname{rank}\Big(- \mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}(\mathbb{E}\mathbf{A}) \Big) = 2K.
$$

We will show that there exists $\mathbf{q}_{10}$ and $\mathbf{q}_{20}$ such that

$$
\mathbf{q}_0 = \mathbf{S}\mathbf{V}\mathbf{q}_{10} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{20}.
\tag{27}
$$

Indeed, we have

$$x_0^{-1}\mathbf{q}_0 = \mathbf{B}_E\mathbf{q}_0 = \left(-\mathbf{SVDV}^\top\mathbf{S} + \mathbf{VDV}^\top\mathbf{S} + \mathbf{SVDV}^\top\right)\mathbf{q}_0\,. \tag{28}$$

Multiplying both sides of (28) by $\mathbf{S}$, we have

$$x_0^{-1}\mathbf{Sq}_0 = \mathbf{SV}(\mathbf{DV}^\top\mathbf{q}_0)\,. \tag{29}$$

Similarly, multiplying both sides of (28) by $(\mathbf{I} - \mathbf{S})$, we have

$$x_0^{-1}(\mathbf{I} - \mathbf{S})\mathbf{q}_0 = (\mathbf{I} - \mathbf{S})\mathbf{V}(\mathbf{DV}^\top\mathbf{Sq}_0)\,. \tag{30}$$

Notice that $\mathbf{q}_0 = \mathbf{Sq}_0 + (\mathbf{I} - \mathbf{S})\mathbf{q}_0$, by (29) and (30), (27) holds by defining $\mathbf{q}_{10} = x_0\mathbf{DV}^\top\mathbf{q}_0$ and $\mathbf{q}_{20} = x_0\mathbf{DV}^\top\mathbf{Sq}_0$. Now it is ready for us to show the statement below (5) hold. Substituting $\mathbf{q}_0 = \mathbf{SVq}_{10} + (\mathbf{I} - \mathbf{S})\mathbf{Vq}_{20}$ and $\mathbb{E}\mathbf{A} = \mathbf{VDV}^\top$ into the eigenvalue definition

$$\mathbf{B}_E\mathbf{q}_0 = x_0^{-1}\mathbf{q}_0\,,$$

we have the following equality

$$\mathbf{VDV}^\top\mathbf{SVq}_{10} + \mathbf{SVD}(\mathbf{I} - \mathbf{V}^\top\mathbf{SV})\mathbf{q}_{20} = x_0^{-1}\left[\mathbf{SVq}_{10} + (\mathbf{I} - \mathbf{S})\mathbf{Vq}_{20}\right]\,. \tag{31}$$

Multiplying $\mathbf{V}^\top(\mathbf{I} - \mathbf{S})$ to both sides of (31), we have

$$\mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{VDV}^\top\mathbf{SVq}_{10} = x_0^{-1}\mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{Vq}_{20}\,.$$

This means that $\mathbf{q}_{20} = x_0\mathbf{DV}^\top\mathbf{SVq}_{10}$ if $\mathbf{V}^\top(\mathbf{I}-\mathbf{S})\mathbf{V} = \mathbf{I}-\mathbf{V}^\top\mathbf{SV}$ is invertible. Substituting $\mathbf{q}_{20} = x_0\mathbf{DV}^\top\mathbf{SVq}_{10}$ into (31) and multiply both sides of (31) by $\mathbf{V}^\top\mathbf{S}$, we see that $\mathbf{q}_{10}$ is should be an eigenvector of $\mathbf{H}(x_0)$ corresponding to the zero eigenvalue if $\mathbf{V}^\top\mathbf{SV}$ is invertible, therefore $\det(\mathbf{H}(x_0)) = 0$. Summarizing the arguments above, we finish our proof.

A.3 Proof of Lemma 2

By Condition 1, for sufficiently large n, there exists a positive constant c such that

$$\mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V} \le p_n\mathbf{I} \quad \text{and} \quad 2cp_n\mathbf{I} \le \mathbf{V}^\top(\mathbb{E}\mathbf{S} + (p_n - \mathbb{E}a_{11})\mathbf{e}_1\mathbf{e}_1^\top)\mathbf{V}\,,$$

in which $\mathbf{e}_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^n$. Condition 2 implies that

$$\|\mathbf{V}^\top(p_n - \mathbb{E}a_{11}))\mathbf{e}_1\mathbf{e}_1^\top\mathbf{V}\|_F \le \frac{p_nKC^2}{n}\,.$$

Combining the three above inequalities together, we have

$$cp_n\mathbf{I} \le \mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V} \le p_n\mathbf{I}\,. \tag{32}$$

Then we study the relation between $\mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V}$ and $\mathbf{V}^\top\mathbf{SV}$. By Lemma 17, we have for any $t > 0$,

$$\mathbb{P}\left(\left|\mathbf{v}_i^\top(\mathbf{S} - \mathbb{E}\mathbf{S})\mathbf{v}_j\right| \ge t\right) \le \exp\left(-\frac{t^2/2}{\sum_{l=1}^n \mathbf{v}_{il}^2\mathbf{v}_{jl}^2\mathrm{var}(S_{ll}) + \frac{Lt}{3}}\right)\,, \tag{33}$$

25

in which $L = \frac{C^2}{n}$ by Condition 2. It follows from Conditions 1 and 2 that,

$$\sum_{l=1}^{n} \mathbf{v}_{il}^2 \mathbf{v}_{jl}^2 \mathrm{var}(S_{ll}) \leq p_n \sum_{l=1}^{n} \mathbf{v}_{il}^2 \mathbf{v}_{jl}^2 \leq \frac{C^2 p_n}{n} \,.$$

We choose $t = 2c_1\sqrt{\log n} \cdot \sqrt{p_n \frac{C^2}{n}}$ for some constant $c_1 > 1$. With this choice of t, under Condition 1, it holds for sufficiently large n that

$$\frac{t^2/2}{\sum_{l=1}^{n} \mathbf{v}_{il}^2 \mathbf{v}_{jl}^2 \mathrm{var}(S_{ll}) + \frac{Lt}{3}} \geq c_1 \log n \,.$$

It follows from Condition 1 that $p_n \sqrt[4]{\frac{\log n}{np_n}} \gg t$. Moreover, since c_1 can any positive constant, by (33), with high probability we have

$$-cp_n\left(\sqrt[4]{\frac{\log n}{np_n}}\right)\mathbf{I} \leq \mathbf{V}^\top(\mathbf{S} - \mathbb{E}\mathbf{S})\mathbf{V} \leq p_n\left(\sqrt[4]{\frac{\log n}{np_n}}\right)\mathbf{I} \,. \tag{34}$$

This combined with (32) implies that with high probability,

$$cp_n\left(1 - \sqrt[4]{\frac{\log n}{np_n}}\right)\mathbf{I} \leq \mathbf{V}^\top\mathbf{S}\mathbf{V} \leq p_n\left(1 + \sqrt[4]{\frac{\log n}{np_n}}\right)\mathbf{I} < \left(1 - \frac{c}{2}\right)\mathbf{I} \,.$$

A.4 Proof of Corollary 1

By Lemma 2, with high probability, both $\mathbf{V}^\top\mathbf{S}\mathbf{V}$ and $\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V}$ are invertible and the inequalities (6) hold. In the rest, we restrict ourselves to this high probability event $\mathcal{A}_1$.

We will show that for suitably chosen $\mathbf{q}_l$, $\dim(\mathrm{span}\{\mathbf{q}_l, l \in [\pm K]\}) = 2K$. Concretely, for any given pair $i_0 \neq j_0$, we consider two scenarios **(I)** and **(II)**.

(I) $x_{i_0} \neq x_{j_0}$. Note that $-\mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}\mathbb{E}\mathbf{A}$ is a real symmetric matrix and $\mathbf{q}_{i_0}$ and $\mathbf{q}_{j_0}$ are eigenvectors corresponding to distinct eigenvalues, then $\mathbf{q}_{i_0}^\top\mathbf{q}_{j_0} = 0$.

(II) $x_{i_0} = x_{j_0}$. In this scenario, the argument above does not directly apply. However, we can perturb the entries of $\mathbf{A}$ and $\mathbf{S}$ and replicate the argument, and then make the perturbation vanish in the limit. Concretely, we replace a_{ij} by $\widehat{a}_{ij} = a_{ij} + e^{-m}g_{ij}$, where $m \geq n$, $g_{ij} = g_{ji}$ and g_{ij} follows i.i.d. standard guassian distribution for $i \leq j$. Then the entries of $\widehat{\mathbf{A}} = (\widehat{a}_{ij})$ are absolute continuous random variables. Then the entries of the matrix $-\widehat{\mathbf{S}}(\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + (\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + \widehat{\mathbf{S}}\mathbb{E}\widehat{\mathbf{A}}$, where $\widehat{\mathbf{S}} = \mathrm{diag}(\widehat{a}_{11}, \ldots, \widehat{a}_{1n})$, are absolute continuous random variables, and its nonzero eigenvalues are not equal almost surely (c.f. Knowles and Yin (2013)). Similar to $\mathbb{E}\mathbf{A}$, we write $\mathbb{E}\widehat{\mathbf{A}} = \mathbb{E}\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\top$. By the tail probability of standard guassian distribution, $\max_j |\widehat{a}_{1j} - a_{1j}| \leq \frac{p_n}{\log n}$ with high probability. This combined with with Lemma 2 implies that $\mathbf{V}^\top\widehat{\mathbf{S}}\mathbf{V}$ and $\mathbf{I} - \mathbf{V}^\top\widehat{\mathbf{S}}\mathbf{V}$ are invertible with high probability. Denote this high probability event by $\mathcal{A}_2$. In the following, we restrict ourselves to $\mathcal{A}_1 \cap \mathcal{A}_2$. Note that with a_{ij} replaced by $(\widehat{a}_{ij})$, counterparts of Theorem 1 holds by following exact the same proof, and we use notations $\widehat{\mathbf{q}}_{1i}$, $\widehat{\mathbf{q}}_{2i}$, and $\widehat{\mathbf{q}}_i$ accordingly. By Theorem 1, $\|\widehat{\mathbf{q}}_{1i}\|_2 \neq 0$, $i \in [\pm K]$. Without loss of generality, we assume $\|\widehat{\mathbf{q}}_{1i}\|_2 = 1$,

$i \in [\pm K]$. We denote the non-zero eigenvalues of $-\widehat{\mathbf{S}}(\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + (\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + \widehat{\mathbf{S}}\mathbb{E}\widehat{\mathbf{A}}$ by $\widehat{x}_{-j}$, $j \in \{\pm 1, \ldots, \pm K\}$ and $\widehat{x}_i^{-1} \leq \widehat{x}_j^{-1}$ for $i < j$. Moreover, as almost surely, the non-zero real eigenvalues of $-\widehat{\mathbf{S}}(\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + (\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + \widehat{\mathbf{S}}\mathbb{E}\widehat{\mathbf{A}}$ are not equal, we can just ignore the measure zero set and take $\widehat{x}_i^{-1} < \widehat{x}_j^{-1}$ for $i < j$. For the particular indexes i_0 and j_0, we have $\widehat{\mathbf{q}}_{i_0}^\top \widehat{\mathbf{q}}_{j_0} = 0$.

By Weyl's inequality, $\lim_{m \to \infty} \widehat{x}_{i_0} = x_{i_0}$ and $\lim_{m \to \infty} \widehat{x}_{j_0} = x_{j_0}$. Without loss of generality, assume that the limits $\lim_{m \to \infty} \widehat{\mathbf{q}}_{1 i_0}$ and $\lim_{m \to \infty} \widehat{\mathbf{q}}_{1 j_0}$ exist. Otherwise, because $\|\widehat{\mathbf{q}}_{1 i_0}\|_2 = \|\widehat{\mathbf{q}}_{1 j_0}\|_2 = 1 < \infty$, we can always find a subsequence of $\{m, m+1, \ldots\}$ and take the limits on this subsequence. Denote by $\mathbf{q}_{1 i_0} = \lim_{m \to \infty} \widehat{\mathbf{q}}_{1 i_0}$ and $\mathbf{q}_{1 j_0} = \lim_{m \to \infty} \widehat{\mathbf{q}}_{1 j_0}$. It can be shown easily that $\mathbf{q}_{i_0}$ and $\mathbf{q}_{j_0}$ are unit eigenvectors of $\mathbf{H}(x_{i_0})$ and $\mathbf{H}(x_{j_0})$, respectively. Let $\mathbf{q}_{i_0} = \mathbf{S}\mathbf{V}\mathbf{q}_{1 i_0} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2 i_0}$ and $\mathbf{q}_{j_0} = \mathbf{S}\mathbf{V}\mathbf{q}_{1 j_0} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2 j_0}$. By the definition of $\widehat{\mathbf{q}}_{i_0}$, we have

$$[-\widehat{\mathbf{S}}(\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + (\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + \widehat{\mathbf{S}}\mathbb{E}\widehat{\mathbf{A}}]\widehat{\mathbf{q}}_{i_0} = \widehat{x}_{i_0}^{-1}\widehat{\mathbf{q}}_{i_0}\,.$$

Then

$$[-\mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}\mathbb{E}\mathbf{A}]\mathbf{q}_{i_0} = \lim_{m \to \infty}[-\widehat{\mathbf{S}}(\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + (\mathbb{E}\widehat{\mathbf{A}})\widehat{\mathbf{S}} + \widehat{\mathbf{S}}\mathbb{E}\widehat{\mathbf{A}}]\widehat{\mathbf{q}}_{i_0}$$

$$= \lim_{m \to \infty}\left[\widehat{x}_{i_0}^{-1}\widehat{\mathbf{q}}_{i_0}\right] = x_{i_0}^{-1}\mathbf{q}_{i_0}\,.$$

Combining this with Lemma 2, $\mathbf{q}_{i_0}$ ($\mathbf{q}_{j_0}$) is not equal to $\mathbf{0}$ and it is the eigenvector of $-\mathbf{S}(\mathbb{E}\mathbf{A})\mathbf{S} + (\mathbb{E}\mathbf{A})\mathbf{S} + \mathbf{S}(\mathbb{E}\mathbf{A})$ corresponding to $x_{i_0}^{-1}$ ($x_{j_0}^{-1}$). Moreover,

$$\mathbf{q}_{i_0}^\top \mathbf{q}_{j_0} = (\lim_{m \to \infty}\widehat{\mathbf{q}}_{i_0})^\top(\lim_{m \to \infty}\widehat{\mathbf{q}}_{j_0}) = \lim_{m \to \infty}\widehat{\mathbf{q}}_{i_0}^\top\widehat{\mathbf{q}}_{j_0} = 0,\ a.s\,.$$

In the above, one should note that the limit is taken on m while n is fixed.

Therefore we can finish our proof of the first statement (7). The second statement follows from (7) directly.

A.5 Proof of Theorem 2

Because $\det(\mathbf{I} - \mathbf{A}\mathbf{B}) = \det(\mathbf{I} - \mathbf{B}\mathbf{A})$, $\det(\mathbf{H}(x)) = 0$ (i.e., (4)) is equivalent to

$$\det(\mathbf{I} - x(\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2}\mathbf{D}(\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2} - x^2(\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2}\mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}(\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2}) = 0\,. \quad (35)$$

Let $\mathbf{A}_1 = (\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2}\mathbf{D}(\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2}$, $\mathbf{A}_2 = (\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2}\mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D}(\mathbf{V}^\top\mathbf{S}\mathbf{V})^{1/2}$ and $y = 1/x$. Then (35) becomes

$$\det(y^2\mathbf{I} - y\mathbf{A}_1 - \mathbf{A}_2) = 0\,. \quad (36)$$

By Lemma 2 and Condition 3, with high probability we have

$$\|\mathbf{A}_1\| \leq |d_1| \cdot \|\mathbf{V}^\top\mathbf{S}\mathbf{V}\| \leq |d_1| \cdot p_n\left(1 + \sqrt[4]{\frac{\log n}{n p_n}}\right) \lesssim n p_n^2\,. \quad (37)$$

Also by Lemma 2 and Condition 3, it holds with high probability that

$$n^2 p_n^3 \mathbf{I} \lesssim \lambda_K(\mathbf{V}^\top\mathbf{S}\mathbf{V})\lambda_K(\mathbf{D}^2)\mathbf{I} \lesssim \mathbf{A}_2 \leq \lambda_1(\mathbf{V}^\top\mathbf{S}\mathbf{V})\lambda_1(\mathbf{D}^2)\mathbf{I} \lesssim n^2 p_n^3 \mathbf{I}\,. \quad (38)$$

Combining (37) and (38), we have with high probability,

$$\mathbf{A}_1 \lesssim np_n^2 \mathbf{I} \lesssim \frac{1}{np_n} \mathbf{A}_2 \, . \tag{39}$$

Similarly, we also have

$$\mathbf{A}_1 \gtrsim np_n^2 \mathbf{I} \, . \tag{40}$$

For y to be a solution for (36), it must hold that

$$\lambda_K(y\mathbf{A}_1 + \mathbf{A}_2)\mathbf{I} \le y^2 \mathbf{I} \le \lambda_1(y\mathbf{A}_1 + \mathbf{A}_2)\mathbf{I} \, .$$

Combining this with (38) and (39), we have

$$-|y|np_n^2 \mathbf{I} + n^2 p_n^3 \mathbf{I} \lesssim y^2 \mathbf{I} \lesssim |y|np_n^2 \mathbf{I} + n^2 p_n^3 \mathbf{I} \, .$$

Therefore we have

$$-|y_i|np_n^2 + n^2 p_n^3 \lesssim y_i^2 \lesssim |y_i|np_n^2 + n^2 p_n^3, \ \ i = \pm 1, \dots, \pm K \, . \tag{41}$$

For the right inequality that $y_i^2 \lesssim |y_i|np_n^2 + n^2 p_n^3$, by the solutions to quadratic equation in one unknown, we imply that

$$|y_i| \lesssim np_n^2 + \sqrt{n^2 p_n^4 + 4n^2 p_n^3} \lesssim np_n^{3/2} \, .$$

Similarly considering the left inequality of (41), we have

$$|y_i| \gtrsim -np_n^2 + \sqrt{n^2 p_n^4 + 4n^2 p_n^3} \gtrsim np_n^{3/2} \, .$$

It follows from the two above inequalities that $|x_i|^{-1} = |y_i| \sim np_n^{3/2}$. This conclude the proof of the first claim.

When $p_n \to 0$, by (39), we have $|y_i|\mathbf{A}_1 \lesssim p_n^{1/2}\mathbf{A}_2 \ll \mathbf{A}_2$. Then to approximately solve for y in (36), we solve z in the following determinant equation

$$\det(z^2 \mathbf{I} - \mathbf{A}_2) = 0 \, . \tag{42}$$

The left hand side of (42) is a $2K$ polynomial of z; therefore (42) has $2K$ solutions in z. A straightforward calculation shows that (42) has $2K$ solutions $\pm\sqrt{\lambda_i(\mathbf{A}_2)}$, $i \in [K]$. By Lemma 2 and an intermediate step (34) in its proof, it holds with high probability that

$$|\lambda_i(\mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V}\mathbf{D}(\mathbf{I} - \mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V})\mathbf{D}) - \lambda_i(\mathbf{A}_2)|$$

$$= |\lambda_i(\mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V}\mathbf{D}(\mathbf{I} - \mathbf{V}^\top(\mathbb{E}\mathbf{S})\mathbf{V})\mathbf{D}) - \lambda_i(\mathbf{V}^\top\mathbf{S}\mathbf{V}\mathbf{D}(\mathbf{I} - \mathbf{V}^\top\mathbf{S}\mathbf{V})\mathbf{D})| \ \lesssim n^2 p_n^3 \left(\sqrt[4]{\frac{\log n}{np_n}} \right) \, .$$

Combining this with eigenvalue separation condition μ_i, we have

$$1 + c_1 \le \frac{\lambda_{i-1}(\mathbf{A}_2)}{\lambda_i(\mathbf{A}_2)}, \quad i = 2, \dots, K \, , \tag{43}$$

28

for some positive constant c_1. For $i = 1, \ldots, K$, let $y_{1i} = \sqrt{\lambda_i(\mathbf{A}_2)} + p_n^{5/6} n^{1/2} \sqrt{\|\mathbf{A}_1\|}$ and $y_{2i} = \sqrt{\lambda_i(\mathbf{A}_2)} - p_n^{5/6} n^{1/2} \sqrt{\|\mathbf{A}_1\|}$, by (38), (39), (40), (43) and Weyl's inequality, it can be shown that with high probability

$$
\lambda_{i-1}(y_{li}\mathbf{A}_1 + \mathbf{A}_2) \geq \lambda_{i-1}(\mathbf{A}_2) - |y_{li}| \|\mathbf{A}_1\| \geq (1+c_1)\lambda_i(\mathbf{A}_2) - n^2 p_n^{7/2-1/100} > y_{li}^2
$$
$$
\sim \lambda_i(\mathbf{A}_2) - (-1)^l n^2 p_n^{10/3}, \tag{44}
$$

$$
\lambda_{i+1}(y_{li}\mathbf{A}_1 + \mathbf{A}_2) \leq \lambda_{i+1}(\mathbf{A}_2) + |y_{li}| \|\mathbf{A}_1\| \leq (1+c_1)^{-1}\lambda_i(\mathbf{A}_2) + n^2 p_n^{7/2-1/100} < y_{li}^2
$$
$$
\sim \lambda_i(\mathbf{A}_2) - (-1)^l n^2 p_n^{10/3}, l = 1, 2. \tag{45}
$$

By Weyl's inequality and (44)-(45), we have

$$
\lambda_i(y_{1i}\mathbf{A}_1 + \mathbf{A}_2) \leq \lambda_i(\mathbf{A}_2) + |y_{1i}| \|\mathbf{A}_1\| < y_{1i}^2, \quad \lambda_i(y_{2i}\mathbf{A}_1 + \mathbf{A}_2) > \lambda_i(\mathbf{A}_2) - |y_{2i}| \|\mathbf{A}_1\| > y_{2i}^2. \tag{46}
$$

It follows from (44)-(46) that

$$
\det(y_{1i}^2 \mathbf{I} - y_{1i}\mathbf{A}_1 - \mathbf{A}_2) \cdot \det(y_{2i}^2 \mathbf{I} - y_{2i}\mathbf{A}_1 - \mathbf{A}_2)
$$
$$
= \prod_{j=1}^{K} \left(y_{1i}^2 - \lambda_j(y_{1i}\mathbf{A}_1 + \mathbf{A}_2) \right) \prod_{j=1}^{K} \left(y_{2i}^2 - \lambda_j(y_{2i}\mathbf{A}_1 + \mathbf{A}_2) \right) < 0. \tag{47}
$$

Since $\det(y^2\mathbf{I} - y\mathbf{A}_1 - \mathbf{A}_2)$ is a continuous function of y, there exists one $y_i \in [y_{2i}, y_{1i}]$ satisfying (35), or equivalently (36). Moreover, by (38),(39), (40) and (43), the intervals $[y_{2i}, y_{1i}]$ are non overlapping for different i. Hence the second claim of Theorem 2 holds for $i \in [K]$. Similarly, the second claim of Theorem 2 holds for $i < 0$ by almost the same proof if we define $y_{1i} = -\sqrt{\lambda_{K+i+1}(\mathbf{A}_2)} + p_n^{5/6} n^{1/2} \sqrt{\|\mathbf{A}_1\|}$ and $y_{2i} = -\sqrt{\lambda_{K+i+1}(\mathbf{A}_2)} - p_n^{5/6} n^{1/2} \sqrt{\|\mathbf{A}_1\|}$.

A.6 Proof of Lemma 3

By Theorem 5.2 of Lei and Rinaldo (2015), with high probability we have

$$
\|\mathbf{A} - \mathbb{E}\mathbf{A}\| \lesssim \sqrt{np_n}. \tag{48}
$$

Combining this with the fact that $\|\mathbf{S}\| \leq 1$, it holds with high probability that

$$
\|\mathbf{B} - \mathbf{B}_E\| \lesssim \sqrt{np_n}.
$$

(3) and Theorem 2 (first part) imply the second part of this Lemma, whose proof we omit. Therefore we complete the proof of this lemma.

A.7 The matrix A and no-self loop

In network analysis, if there is no self-loop, then the diagonal entries of the adjacency matrix are 0's. In this case we should analyze $\widetilde{\mathbf{A}} = \mathbf{A} - \text{diag}(a_{11}, \ldots, a_{nn})$ instead of $\mathbf{A}$ and therefore individual 1's perceived adjacency matrix is $\widetilde{\mathbf{B}} = -\widetilde{\mathbf{S}}\widetilde{\mathbf{A}}\widetilde{\mathbf{S}} + \widetilde{\mathbf{A}}\widetilde{\mathbf{S}} + \widetilde{\mathbf{S}}\widetilde{\mathbf{A}}$, where $\widetilde{\mathbf{S}} = \mathbf{S} - \text{diag}(a_{11}, 0, \ldots, 0)$. In this case, $\mathbb{E}\widetilde{\mathbf{A}}$ may not be a low-rank matrix, while $\mathbb{E}\mathbf{A}$ is. For instance, we look at a simple SBM with $n = 4$, where the corresponding 4×4 expected adjacency matrix with self-loop can be expressed as

$$\mathbb{E}\mathbf{A} = \begin{pmatrix} p_{11} & p_{11} & p_{12} & p_{12} \\ p_{11} & p_{11} & p_{12} & p_{12} \\ p_{21} & p_{21} & p_{22} & p_{22} \\ p_{21} & p_{21} & p_{22} & p_{22} \end{pmatrix},$$

in which $p_{12} = p_{21} < p_{11} = p_{22}$, and $p_{ij} \in (0, 1)$. Clearly rank$(\mathbb{E}\mathbf{A}) = 2$. In contrast, the expectation of the matrix $\widetilde{\mathbf{A}}$ is

$$\mathbb{E}\widetilde{\mathbf{A}} = \begin{pmatrix} 0 & p_{11} & p_{12} & p_{12} \\ p_{11} & 0 & p_{12} & p_{12} \\ p_{21} & p_{21} & 0 & p_{22} \\ p_{21} & p_{21} & p_{22} & 0 \end{pmatrix}.$$

It is easy to show that rank$(\mathbb{E}\widetilde{\mathbf{A}}) = 4$, so $\mathbb{E}\widetilde{\mathbf{A}}$ has the full rank.

There is a rich line of network literature that assumes a low-rank structure of $\mathbf{A}$, including Zhao et al. (2012), Abbe (2017), Abbe et al. (2017), and Zhang et al. (2020). Moreover, in this paper we consider the case that $\|\widetilde{\mathbf{A}} - \mathbb{E}\widetilde{\mathbf{A}}\| \to \infty$, which is a common assumption in network models. Loosely, this assumption is to avoid extremely sparse networks (e.g., we do not deal with the case that the largest degree of the nodes is bounded). By the equation $\widetilde{\mathbf{A}} = \mathbb{E}\mathbf{A} + (\widetilde{\mathbf{A}} - \mathbb{E}\mathbf{A})$ and $\|\widetilde{\mathbf{A}} - \mathbb{E}\mathbf{A}\| \sim \|\widetilde{\mathbf{A}} - \mathbb{E}\widetilde{\mathbf{A}}\| \to \infty$, $\widetilde{\mathbf{A}} - \mathbb{E}\mathbf{A}$ can be regarded as the "noise matrix" of the model. Hence the noise level (measured by spectral norm) of $\widetilde{\mathbf{A}}$ is not changed compared to $\mathbf{A}$. In other words, the signal matrix of $\widetilde{\mathbf{A}}$ is essentially $\mathbb{E}\mathbf{A}$. Therefore a major term of $\widetilde{\mathbf{B}}$ is also a major term of $\mathbf{B}$, and Theorems 1–2 can be applied too. On the other hand, by the definition of Bernoulli random variables, $\|\text{diag}(a_{11}, a_{22}, \ldots, a_{nn})\| \leq 1$. By checking the proofs carefully, the community detection results from Theorem 3 to Theorem 5 mainly rely on the order of the gap between $\|\widetilde{\mathbf{B}} - \mathbf{B}_E\|$ and the smallest non-zero eigenvalue (in magnitude) of $\mathbf{B}_E$. It is essentially the same as the gap between $\|\mathbf{B} - \mathbf{B}_E\|$ and the smallest non-zero eigenvalue (in magnitude) of $\mathbf{B}_E$, where the difference between $\mathbf{S}$ and $\widetilde{\mathbf{S}}$ can be shown to have a negligible effect on this gap. Therefore Theorems 3–5 hold for the stochastic block model without self-loop. Given the above arguments, throughout this paper, we only consider $\mathbf{A}$ (instead of $\widetilde{\mathbf{A}}$) for convenience.

Appendix B. Proofs of the results in Section 3

B.1 Proof of Lemma 4

As we have assumed that individual 1 belongs to community 1, $\mathbb{P}(a_{1l} = 1) = \boldsymbol{\pi}_1^\top \mathbf{P} \boldsymbol{\pi}_l = \mathbf{e}_1^\top \mathbf{P} \boldsymbol{\pi}_l$, for $l \in [n]$. Then in view of the definition of p_n and $\min_{k \in [K]} p_{1k} \sim p_n$, it follows

that $\min_{l \geq 2} \mathbb{P}(a_{1l} = 1) \sim p_n$. Moreover, $1 - c \geq p_n \gg (1/n)^{1/2} \gg \log n / n$. Therefore Condition 1 is validated.

By $\min_{k \in [K]} \sum_{j=1}^n \mathbf{1}(\boldsymbol{\pi}_j = \mathbf{e}_k) \geq c_0 n$, $\sigma_K(\mathbf{P}) \geq c_1 p_n$, we have

$$\text{rank}(\mathbb{E}\mathbf{A}) = \text{rank}(\mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^\top) = K \,.$$

Recall the eigen decomposition $\mathbb{E}\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\top$, in which $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_K)$. By the structure of the stochastic block model and $\text{rank}(\mathbb{E}\mathbf{A}) = K$, there are K different rows in $\mathbf{V}$ corresponding to the communities and hence there are K^2 different values in $\mathbf{V}$ at most. Indeed, let $\mathbf{v}_k(l)$ be the l-th entry of $\mathbf{v}_k$, by the definition of eigenvector, we have

$$(\mathbb{E}\mathbf{A})\mathbf{v}_k = d_k \mathbf{v}_k \,, \quad \text{for } k \in [K] \,,$$

and therefore

$$\sum_{j=1}^n \left(\boldsymbol{\pi}_l^\top \mathbf{P} \boldsymbol{\pi}_j \mathbf{v}_k(j) \right) = d_k \mathbf{v}_k(l), \quad \text{for } k \in [K] \text{ and } l \in [n] \,.$$

For any $l_1 \neq l_2$ with $\boldsymbol{\pi}_{l_1} = \boldsymbol{\pi}_{l_2}$, we have $\boldsymbol{\pi}_{l_1}^\top \mathbf{P} \boldsymbol{\pi}_j = \boldsymbol{\pi}_{l_2}^\top \mathbf{P} \boldsymbol{\pi}_j$, $j = 1, \ldots, n$ and therefore

$$\sum_{j=1}^n \left(\boldsymbol{\pi}_{l_1}^\top \mathbf{P} \boldsymbol{\pi}_j \mathbf{v}_k(j) \right) = \sum_{j=1}^n \left(\boldsymbol{\pi}_{l_2}^\top \mathbf{P} \boldsymbol{\pi}_j \mathbf{v}_k(j) \right) = d_k \mathbf{v}_k(l_1) = d_k \mathbf{v}_k(l_2) \,.$$

Notice that $\boldsymbol{\pi}_i \in \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$, $i \in [n]$. Then we conclude that $\mathbf{V}$ has at most K different rows and $\{\mathbf{v}_k(l), k \in [K], l \in [n]\}$ only has at most K^2 distinct values. Moreover, as $\boldsymbol{\pi}_{l_1} \neq \boldsymbol{\pi}_{l_2}$ means that l_1 and l_2 belong to different communities, the rows of $\mathbf{V}$ with distinct values are corresponding to different communities. Since $\text{rank}(\mathbb{E}\mathbf{A}) = \text{rank}(\mathbf{V}\mathbf{D}\mathbf{V}^\top) = K$, $\text{rank}(\mathbf{V}) = K$ and therefore $\mathbf{V}$ contains exactly K different rows.

Without loss of generality, assume that the first K rows of $\mathbf{V}$ are different and we denote this $K \times K$ matrix by $\mathbf{V}^{(K)}$. Since distinct row values are corresponding to different communities, the first K rows of $\mathbf{\Pi}$ are different. Noticing that $\boldsymbol{\pi}_i \in \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$, $i \in [n]$; without loss of generality, we assume the first K rows of $\mathbf{\Pi}$ equal to $\mathbf{I}$. Let $\mathcal{D} = \mathbf{V}^{(K)}$. Then it follows that

$$\mathbf{V} = \mathbf{\Pi}\mathcal{D} \,. \tag{49}$$

Because $\mathcal{D}^\top \mathbf{\Pi}^\top \mathbf{\Pi} \mathcal{D} = \mathbf{V}^\top \mathbf{V} = \mathbf{I}$, we have $\mathcal{D}\mathcal{D}^\top \geq \frac{1}{P}\mathbf{I}$, where $P = \max_{k \in [K]} \sum_{j=1}^n \mathbf{1}(\boldsymbol{\pi}_j = \mathbf{e}_i)$. Therefore, (9) is proved. Moreover, by the condition that $\min_{k \in [K]} \sum_{j=1}^n \mathbf{1}(\boldsymbol{\pi}_j = \mathbf{e}_k) \geq c_0 n$, we have

$$\mathcal{D}\mathcal{D}^\top \leq \frac{1}{c_0 n}\mathbf{I} \,.$$

Therefore we have

$$\|\mathcal{D}\|_{\max} \leq \frac{1}{\sqrt{c_0 n}} \,.$$

Combining this with (49), Condition 2 holds. Now we prove Condition 3. Notice that $\mathbf{\Pi}^\top \mathbf{\Pi} = \boldsymbol{\pi}_1 \boldsymbol{\pi}_1^\top + \ldots, + \boldsymbol{\pi}_n \boldsymbol{\pi}_n^\top$ is a $K \times K$ diagonal matrix whose diagonal elements are

31

$\sum_{j=1}^{n} \mathbf{1}(\boldsymbol{\pi}_j = \mathbf{e}_k)$, $k \in [K]$. By Condition 4, we have

$$d_K^2 = \sigma_K^2(\mathbb{E}\mathbf{A}) = \sigma_K^2(\boldsymbol{\Pi}\mathbf{P}\boldsymbol{\Pi}^\top) = \lambda_K(\boldsymbol{\Pi}\mathbf{P}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\mathbf{P}\boldsymbol{\Pi}^\top)$$
$$= \lambda_K(\mathbf{P}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\mathbf{P}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}) \geq \lambda_K(\boldsymbol{\Pi}^\top\boldsymbol{\Pi})\lambda_K(\mathbf{P}\boldsymbol{\Pi}^\top\boldsymbol{\Pi}\mathbf{P})$$
$$\geq \lambda_K^2(\boldsymbol{\Pi}^\top\boldsymbol{\Pi})\lambda_K(\mathbf{P}^2) = \sigma_K^2(\boldsymbol{\Pi}^\top\boldsymbol{\Pi})\sigma_K^2(\mathbf{P}) \gtrsim n^2 w_n^2.$$

This, combined with

$$\|\mathbb{E}\mathbf{A}\|_F = \left[\sum_{j=1}^{n} \sum_{i=1}^{n} (\mathbb{P}(a_{ij} = 1))^2 \right]^{1/2} \leq np_n,$$

implies that $|d_1| \sim \ldots \sim |d_K| \sim np_n$, which is Condition 3.

B.2 Proof of Lemma 5

First note that (10) can be equivalently expressed as

$$\mathbf{Q} = \left(\frac{\mathbf{S}\boldsymbol{\Pi}}{\sqrt{p_n}}, (\mathbf{I} - \mathbf{S})\boldsymbol{\Pi} \right) \mathcal{Q}, \quad \text{where} \quad \mathcal{Q} = \begin{pmatrix} \sqrt{p_n}\mathcal{D}\mathcal{Q}_1 \\ \mathcal{D}\mathcal{Q}_2 \end{pmatrix}. \tag{50}$$

We have the following lower bounds.

Lemma 8 *Under Condition 4, with high probability, there exists some positive constant c_2 such that*

$$\mathcal{Q}\mathcal{Q}^\top \geq (c_2 n)^{-1}\mathbf{I}, \quad \mathcal{D}\mathcal{Q}_2\mathcal{Q}_2^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I}, \quad \text{and} \quad p_n\mathcal{D}\mathcal{Q}_1\mathcal{Q}_1^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I}. \tag{51}$$

Proof Substituting (50) into $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$ we have,

$$\mathbf{I} = \mathbf{Q}^\top\mathbf{Q} = \mathcal{Q}^\top \text{diag}\left(\boldsymbol{\Pi}^\top\frac{\mathbf{S}}{p_n}\boldsymbol{\Pi}, \boldsymbol{\Pi}^\top(\mathbf{I} - \mathbf{S})\boldsymbol{\Pi} \right) \mathcal{Q}. \tag{52}$$

By Lemma 2 and (9) in Lemma 4, there exists a positive constant c_2 such that with high probability,

$$\text{diag}\left(\boldsymbol{\Pi}^\top\frac{\mathbf{S}}{p_n}\boldsymbol{\Pi}, \boldsymbol{\Pi}^\top(\mathbf{I} - \mathbf{S})\boldsymbol{\Pi} \right) = \text{diag}(\mathcal{D}^\top, \mathcal{D}^\top)^{-1}\text{diag}\left(\mathbf{V}^\top\frac{\mathbf{S}}{p_n}\mathbf{V}, \mathbf{V}^\top(\mathbf{I} - \mathbf{S})\mathbf{V} \right)\text{diag}(\mathcal{D}, \mathcal{D})^{-1}$$
$$\leq c_2 n\mathbf{I}. \tag{53}$$

By (52) and (53), with high prbability we have

$$\mathcal{Q}\mathcal{Q}^\top \geq (c_2 n)^{-1}\mathbf{I}.$$

Then by the definition of $\mathcal{Q}$ in (50), we have

$$\mathcal{D}\mathcal{Q}_2\mathcal{Q}_2^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I} \quad \text{and} \quad p_n\mathcal{D}\mathcal{Q}_1\mathcal{Q}_1^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I}.$$

$\blacksquare$

Proof [Proof of Lemma 5] By (10) and Corollary 1, $\mathbf{S\Pi}$ and $(\mathbf{I}-\mathbf{S})\mathbf{\Pi}$ contain K different non-zero rows in each matrix; without loss of generality, these rows can be rearranged as $2K \times 2K$ identity matrix $\mathbf{I}_{2K} = \mathrm{diag}(\mathbf{I}_K, \mathbf{I}_K)$, where the two $\mathbf{I}_K$'s correspond to the different non-zero rows of $\mathbf{S\Pi}$ and $(\mathbf{I}-\mathbf{S})\mathbf{\Pi}$ respectively. For (11), without loss of generality, assume that $\boldsymbol{\pi}_i = \mathbf{e}_1$ and $\boldsymbol{\pi}_j = \mathbf{e}_2$, $a_{1i} = a_{1j} = 1$. The other cases $a_{1i} \neq a_{1j}$ and $a_{1i} = a_{1j} = 0$ can be proved similarly.

By (51) in Lemma 8, with high probability, uniformly for j we have

$$\|\mathbf{Q}(i)\mathbf{O} - \mathbf{Q}(j)\mathbf{O}\|_2 = \left\| (\mathbf{e}_i^{(n)} - \mathbf{e}_j^{(n)})^\top \left(\frac{\mathbf{S\Pi}}{\sqrt{p_n}}, (\mathbf{I}-\mathbf{S})\mathbf{\Pi} \right) \mathcal{Q}\mathbf{O} \right\|_2$$

$$\geq \|(\mathbf{e}_1^{(2K)} - \mathbf{e}_2^{(2K)})\|_2 \times \sigma_{2K}(\mathcal{Q}\mathbf{O}) \geq \frac{\|\mathbf{e}_1^{(2K)} - \mathbf{e}_2^{(2K)}\|_2}{\sqrt{c_2 n}} = \sqrt{\frac{2}{c_2 n}}.$$

For (12), without loss of generality, assume that $\boldsymbol{\pi}_i = \boldsymbol{\pi}_j = \mathbf{e}_1$, $a_{1i} = 1 - a_{1j} = 1$. Similar to the inequality above, we have

$$\|\mathbf{Q}(i)\mathbf{O} - \mathbf{Q}(j)\mathbf{O}\|_2 = \left\| (\mathbf{e}_i^{(n)} - \mathbf{e}_j^{(n)})^\top \left(\frac{\mathbf{S\Pi}}{\sqrt{p_n}}, (\mathbf{I}-\mathbf{S})\mathbf{\Pi} \right) \mathcal{Q}\mathbf{O} \right\|_2$$

$$\geq \|(\mathbf{e}_1^{(2K)} - \mathbf{e}_{K+1}^{(2K)})\|_2 \times \sigma_{2K}(\mathcal{Q}\mathbf{O}) \geq \frac{\|\mathbf{e}_1^{(2K)} - \mathbf{e}_{K+1}^{(2K)}\|_2}{\sqrt{c_2 n}} = \sqrt{\frac{2}{c_2 n}}.$$

The implication (13) follows from the expression of $\mathbf{Q}$ in (10) and the fact that $\|\mathbf{Q}(i)\mathbf{O} - \mathbf{Q}(j)\mathbf{O}\|_2 = \|\mathbf{Q}(i) - \mathbf{Q}(j)\|_2$ for an orthogonal matrix $\mathbf{O}$. ∎

B.3 Proof of Lemma 6

By Lemma 3, with high probability we have

$$\|\mathbf{B} - \mathbf{B}_E\| \lesssim \sqrt{n p_n}.$$

Note that $\mathbf{Q}$ consists of unit eigenvectors of $\mathbf{B}_E$ and Lemma 4 validates the first statement in Theorem 2. Then by Davis-Kahan theorem in Davis and Kahan (1970) (c.f. Theorem 10 in Cai et al. (2013)), with high probability we have

$$\|\mathbf{W}\mathbf{W}^\top - \mathbf{Q}\mathbf{Q}^\top\|_F = O\left(\frac{\sqrt{n p_n}}{n p_n^{3/2}} \right) = O\left(\frac{1}{\sqrt{n} p_n} \right).$$

Moreover, it follows from the definition of $\mathbf{O}$ that

$$\|\mathbf{W} - \mathbf{Q}\mathbf{O}\|_F^2 = \mathrm{tr}\left[(\mathbf{W} - \mathbf{Q}\mathbf{U}_1\mathbf{U}_2^\top)^\top (\mathbf{W} - \mathbf{Q}\mathbf{U}_1\mathbf{U}_2^\top) \right]$$

$$= \mathrm{tr}\left(\mathbf{W}^\top\mathbf{W} + \mathbf{U}_2\mathbf{U}_1^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{U}_1\mathbf{U}_2^\top - 2\mathbf{U}_2\mathbf{U}_1^\top\mathbf{Q}^\top\mathbf{W} \right)$$

$$= 4K - 2\mathrm{tr}(\mathbf{U}_2\mathbf{U}_1^\top\mathbf{Q}^\top\mathbf{W}) \leq 4K - 2\mathrm{tr}(\mathbf{Q}^\top\mathbf{W}\mathbf{W}^\top\mathbf{Q})$$

$$= \|\mathbf{W}\mathbf{W}^\top - \mathbf{Q}\mathbf{Q}^\top\|_F^2, \tag{54}$$

where the inequality follows from $\|\Sigma\| = \|\mathbf{Q}^\top\mathbf{W}\| \leq 1$ and $\mathrm{tr}(\mathbf{U}_2\mathbf{U}_1^\top\mathbf{Q}^\top\mathbf{W}) = \mathrm{tr}(\Sigma) \geq \mathrm{tr}(\Sigma^2) = \mathrm{tr}(\mathbf{Q}^\top\mathbf{W}\mathbf{W}^\top\mathbf{Q})$. Therefore (14) is proved.

B.4 Proof of Theorem 3

To prove Theorem 3, we need a few more auxilliary results. The next lemma gives some theoretical properties of Algorithm 1.

Lemma 9 *Under Condition 4, for any $1 \times 2K$ row vector $\mathbf{c}$ and $i \in [n]$, if $\|\mathbf{c} - \mathbf{Q}(i)\mathbf{O}\|_2 < 1/\sqrt{2c_2 n}$, where c_2 is the same as in Lemma 8 and $\mathbf{O}$ is defined in Lemma 6, then with high probability, $\|\mathbf{c} - \mathbf{Q}(j)\mathbf{O}\|_2 > 1/\sqrt{2c_2 n}$ for all $j \in [n]$ such that $\boldsymbol{\pi}_j \neq \boldsymbol{\pi}_i$.*

Proof In view of Lemma 5, we have

$$\|\mathbf{c} - \mathbf{Q}(j)\mathbf{O}\|_2 \geq \|\mathbf{Q}(i)\mathbf{O} - \mathbf{Q}(j)\mathbf{O}\|_2 - \|\mathbf{c} - \mathbf{Q}(i)\mathbf{O}\|_2 > \sqrt{\frac{2}{c_2 n}} - \frac{1}{\sqrt{2c_2 n}} = \frac{1}{\sqrt{2c_2 n}} \, .$$

∎

Next we bound the set $\mathcal{M}$, which in turn will be used to give an upper bound on the final misclustering rate.

Lemma 10 *Under Condition 4, let $\{\mathbf{c}_i, i \in [n]\}$ be the centroids returned by (15), then for $\mathcal{M}$ defined in (16), it holds with high probability that $|\mathcal{M}| = O\left(1/p_n^2\right)$.*

Proof First, we show that $\mathbf{S\Pi}$ has K different nonzero rows with high probability. Concretely, we will show that

$$\mathbb{P}(\exists k \in [K] : a_{1i} = 0 \text{ for all } i \in [n] \text{ such that } \boldsymbol{\pi}_i = \mathbf{e}_k) = O(n^{-D}), \tag{55}$$

for sufficiently large n, where D is some positive constant. In fact, by Condition 4, for sufficiently large n depending on D, (55) follows from the inequality that

$$\text{L.H.S of (55)} \leq \sum_{k \in [K]} \mathbb{P}\left(a_{1i} = 0 \text{ for all } i \in [n] \text{ such that } \boldsymbol{\pi}_i = \mathbf{e}_k\right)$$

$$\leq \sum_{k \in [K]} \left[1 - \min_i \mathbb{P}(a_{1i} = 1)\right]^{\sum_{i=1}^n \mathbb{1}(\boldsymbol{\pi}_i = \mathbf{e}_k)} = O\left(e^{-\frac{\sum_{i=1}^n \mathbf{1}(\boldsymbol{\pi}_i = \mathbf{e}_k)}{\sqrt{n}}}\right) = O(n^{-D}).$$

By the decomposition equation (10) of $\mathbf{Q}$, we have $\mathbf{SQ} = \mathbf{S\Pi}\mathcal{D}\mathcal{Q}_1$. In view of Lemma 8, $\mathcal{D}\mathcal{Q}_1(\mathcal{D}\mathcal{Q}_1)^\top$ is a $K \times K$ positive definite matrix, and so $\mathbf{SQ}$ has K different non-zero rows with high probability. Then, as $\mathbf{O}$ is an orthogonal matrix, $\mathbf{SQO}$ has the same property. Similarly, we can show that $(\mathbf{I} - \mathbf{S})\mathbf{QO}$ has K different non-zero rows with high probability.

Then, as $\{\mathbf{c}_i, i \in [n]\}$ are the centroids returned by (15), we have with high probability,

$$\|\mathbf{SW} - \mathbf{SC}\|_F \leq \|\mathbf{SW} - \mathbf{SQO}\|_F \, ,$$

and

$$\|(\mathbf{I} - \mathbf{S})\mathbf{W} - (\mathbf{I} - \mathbf{S})\mathbf{C}\|_F \leq \|(\mathbf{I} - \mathbf{S})\mathbf{W} - (\mathbf{I} - \mathbf{S})\mathbf{QO}\|_F \, .$$

Then it follows

$$\|\mathbf{SC} - \mathbf{SQO}\|_F \leq \|\mathbf{SW} - \mathbf{SC}\|_F + \|\mathbf{SW} - \mathbf{SQO}\|_F \leq 2\|\mathbf{SW} - \mathbf{SQO}\|_F \, ,$$

and

$$\|(\mathbf{I} - \mathbf{S})\mathbf{C} - (\mathbf{I} - \mathbf{S})\mathbf{Q}\mathbf{O}\|_F \leq \|(\mathbf{I} - \mathbf{S})\mathbf{W} - (\mathbf{I} - \mathbf{S})\mathbf{C}\|_F + \|(\mathbf{I} - \mathbf{S})\mathbf{W} - (\mathbf{I} - \mathbf{S})\mathbf{Q}\mathbf{O}\|_F$$
$$\leq 2\|(\mathbf{I} - \mathbf{S})\mathbf{W} - (\mathbf{I} - \mathbf{S})\mathbf{Q}\mathbf{O}\|_F.$$

Combining this with Lemma 6, we conclude that with high probability

$$
\begin{aligned}
|\mathcal{M}| = \sum_{i \in \mathcal{M}} 1 &\leq 2c_0 n \sum_{i \in \mathcal{M}} \|\mathbf{c}_i - \mathbf{Q}(i)\mathbf{O}\|_2^2 \leq 2c_0 n \sum_{i \in [n]} \|\mathbf{c}_i - \mathbf{Q}(i)\mathbf{O}\|_2^2 \\
&= 2c_0 n \left(\|\mathbf{SC} - \mathbf{SQO}\|_F^2 + \|(\mathbf{I} - \mathbf{S})\mathbf{C} - (\mathbf{I} - \mathbf{S})\mathbf{QO}\|_F^2 \right) \\
&\leq 8c_0 n \left(\|\mathbf{SW} - \mathbf{SQO}\|_F^2 + \|(\mathbf{I} - \mathbf{S})\mathbf{W} - (\mathbf{I} - \mathbf{S})\mathbf{QO}\|_F^2 \right) \\
&= O\left(n(\|\mathbf{SW} - \mathbf{SQO}\|_F^2 + \|(\mathbf{I} - \mathbf{S})\mathbf{W} - (\mathbf{I} - \mathbf{S})\mathbf{QO}\|_F^2) \right) \\
&= O\left(n\|\mathbf{W} - \mathbf{QO}\|_F^2 \right) \\
&= O\left(\frac{1}{p_n^2} \right).
\end{aligned}
\tag{56}
$$

∎

Finally, we need the following guarantee for our merging strategy in Algorithm 2.

Proposition 2 *Let $\mathfrak{C} = \{\mathfrak{c}_1, \ldots, \mathfrak{c}_K\}$ and $\mathfrak{D} = \{\mathfrak{d}_1, \ldots, \mathfrak{d}_K\}$ be two ordered sets of size K and let $f_0 : [K] \to [K]$ be the unique permutation such that $\mathfrak{c}_{f_0(i)} = \mathfrak{d}_i$ for all $i \in [K]$. Suppose that the K elements in these ordered sets have been endowed with pairwise connection probabilities. Let $\mathbf{C} = (c_{ij})$ be a $K \times K$ matrix such that the (i,j)-th entry c_{ij} is the connection probability between $\mathfrak{c}_i$ and $\mathfrak{c}_j$, and $\mathbf{D} = (d_{ij})$ be a $K \times K$ matrix such that the (i,j)-th entry d_{ij} is the connection probability between $\mathfrak{c}_i$ and $\mathfrak{d}_j$. If $\mathrm{rank}(\mathbf{C}) = K$, then for a permutation function $f : [K] \to [K]$,*

$$f = f_0 \Longleftrightarrow \mathbf{C}_{(f,f)} = \mathbf{D}_{(f,*)},$$

where the (i,j)-th entries of $\mathbf{C}_{(f,f)}$ and $\mathbf{D}_{(f,*)}$ are $c_{f(i)f(j)}$ and $d_{f(i)j}$, respectively.

Proof As $\mathfrak{c}_{f_0(i)} = \mathfrak{d}_i$ for $i \in [K]$, the matrix $\mathbf{D}$ is formed by a column permutation of $\mathbf{C}$. Therefore, we have

$$\mathrm{rank}(\mathbf{D}) = \mathrm{rank}(\mathbf{C}) = K,$$

which implies that columns of $\mathbf{D}$ are distinct. By the definition of f_0, clearly we have

$$f = f_0 \Longrightarrow \mathbf{C}_{(f_0,f_0)} = \mathbf{D}_{(f_0,*)}.$$

On the other hand, if $f \neq f_0$, there exists a $j_0 \in [K]$ such that

$$f(j_0) \neq f_0(j_0).$$

This, combined with the fact the columns of $\mathbf{D}$ are all distinct, implies

$$(c_{f(1)f(j_0)}, \ldots, c_{f(K)f(j_0)})^\top \neq (d_{f(1)j_0}, \ldots, d_{f(K)j_0})^\top.$$

Hence there exists $i_0 \in [K]$ such that

$$c_{f(i_0)f(j_0)} \neq d_{f(i_0)j_0} .$$

In other words, $\mathbf{C}_{(f,f)} \neq \mathbf{D}_{(f,*)}$. ∎

Combining the results above, we now prove Theorem 3.

Proof [Proof of Theorem 3]

For $i, j \in \mathcal{M}^c$, by Lemma 9, with high probability we have

$$\mathbf{c}_i \neq \mathbf{c}_j, a_{1i} = a_{1j} \Leftrightarrow \boldsymbol{\pi}_i \neq \boldsymbol{\pi}_j, a_{1i} = a_{1j} , \tag{57}$$

which means that in Step 2, for $i, j \in \mathcal{M}^c$, with high probability i and j are assigned to the same cluster if and only if $a_{1i} = a_{1j}$ and $\boldsymbol{\pi}_i = \boldsymbol{\pi}_j$. By (10), $\mathbf{Q}$ has $2K$ different rows

$$\mathrm{R} = \{\mathbf{e}_1^\top \mathcal{D} \mathcal{Q}_1, \mathbf{e}_1^\top \mathcal{D} \mathcal{Q}_2, \ldots, \mathbf{e}_K^\top \mathcal{D} \mathcal{Q}_1, \mathbf{e}_K^\top \mathcal{D} \mathcal{Q}_2\} .$$

Notice that, for $l = 1, 2$, respectively, $\mathbf{e}_k$ reflects the membership of individual i if $\boldsymbol{\pi}_i = \mathbf{e}_k$; hence $\{\mathbf{e}_1^\top \mathcal{D} \mathcal{Q}_l, \ldots, \mathbf{e}_K^\top \mathcal{D} \mathcal{Q}_l\} \subset \mathrm{R}$ can be regarded as the "membership" vectors for the individuals $\{i \in [n] : a_{1i} = 2 - l\}$. By (57), with high probability, in step 2 of Algorithm 2, $i \in \mathcal{M}^c \cap \{i \in n : a_{1i} = 2 - l\}$ is assigned to the cluster associated with $\mathbf{e}_k^\top \mathcal{D} \mathcal{Q}_l$ if $\boldsymbol{\pi}_i = \mathbf{e}_k$.

Therefore, with high probability we have

$$|i \in [n] : \{\text{In step 2}, i \text{ is not assigned to the cluster associated with } \mathbf{e}_k^\top \mathcal{D} \mathcal{Q}_1 \text{ or } \mathbf{e}_k^\top \mathcal{D} \mathcal{Q}_2\}| \leq |\mathcal{M}| .$$

By Lemma 10, under Condition 4, we have

$$|\mathcal{M}| = O\left(\frac{1}{p_n^2}\right) . \tag{58}$$

We say the 1st group is $\{i \in [n] : a_{1i} = 1\}$ and the 2nd group is $\{i \in [n] : a_{1i} = 0\}$. Let $\mathbf{S}_l(k)$ be the collection of individuals belonging to k-th community in the l-th group, and $\widehat{\mathbf{S}}_l(k)$ be the k-th cluster of the l-th group returned by step 2. Let $\widehat{p}_{k_1k_2}^{(1)}$ and $\widehat{p}_{k_1k_2}^{(2)}$ be the (k_1, k_2)-th entry of $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$ (defined by equation (17)) respectively. Let $\mathbf{P}^{\mathbf{S},\mathbf{S}} = (p_{k_1k_2}^{(1)})$ and $\mathbf{P}^{\mathbf{S},\mathbf{I}-\mathbf{S}} = (p_{k_1k_2}^{(2)})$ be the corresponding population versions. Note that there exists a unique permutation function f_0 such that

$$\mathbf{P}_{f_0,f_0}^{\mathbf{S},\mathbf{S}} = \mathbf{P}_{f_0,*}^{\mathbf{S},\mathbf{I}-\mathbf{S}} . \tag{59}$$

Without loss of generality, we assume $\mathbf{P}^{\mathbf{S},\mathbf{I}-\mathbf{S}} = \mathbf{P}$.

$$\begin{aligned}
\widehat{p}_{k_1k_2}^{(l)} - p_{k_1k_2}^{(l)} = \frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \Big[&\sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1k_2}^{(l)}) \\
&+ \sum_{i \in \widehat{\mathbf{S}}_1(k_1) \backslash \mathbf{S}_1(k_1), \ j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1k_2}^{(l)}) \\
&+ \sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), \ j \in \widehat{\mathbf{S}}_l(k_2) \backslash \mathbf{S}_l(k_2)} (b_{ij} - p_{k_1k_2}^{(l)}) \\
&+ \sum_{i \in \widehat{\mathbf{S}}_1(k_1) \backslash \mathbf{S}_1(k_1), \ j \in \widehat{\mathbf{S}}_l(k_2) \backslash \mathbf{S}_l(k_2)} (b_{ij} - p_{k_1k_2}^{(l)}) \Big] ,
\end{aligned} \tag{60}$$

and

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i \in \mathbf{S}_1(k_1), j \in \mathbf{S}_l(k_2)} (b_{ij} - p^{(l)}_{k_1 k_2})$$

$$= \frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \Big[\sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p^{(l)}_{k_1 k_2})$$

$$+ \sum_{i \in \mathbf{S}_1(k_1) \setminus \widehat{\mathbf{S}}_1(k_1), \ j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p^{(l)}_{k_1 k_2})$$

$$+ \sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), \ j \in \mathbf{S}_l(k_2) \setminus \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p^{(l)}_{k_1 k_2})$$

$$+ \sum_{i \in \mathbf{S}_1(k_1) \setminus \widehat{\mathbf{S}}_1(k_1), \ j \in \mathbf{S}_l(k_2) \setminus \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p^{(l)}_{k_1 k_2}) \Big]. \tag{61}$$

By Lemma 17, with high probability, we have

$$|\mathbf{S}_1(k)| = \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} a_{1i} \sim \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} \mathbb{E} a_{1i} \sim n p_n, k \in [K], \tag{62}$$

and

$$|\mathbf{S}_2(k)| = \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} (1 - a_{1i}) \sim \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} \mathbb{E}(1 - a_{1i}) \sim n, k \in [K], \tag{63}$$

Moreover, by (58), with high probability we have

$$\left| |\widehat{\mathbf{S}}_l(k)| - |\mathbf{S}_l(k)| \right| \leq |\widehat{\mathbf{S}}_l(k) \setminus \mathbf{S}_l(k)| + |\mathbf{S}_l(k) \setminus \widehat{\mathbf{S}}_l(k)| \leq |\mathcal{M}| = O\left(\frac{1}{p_n^2}\right), l = 1, 2, \ k \in [K]. \tag{64}$$

It follows from (62)–(64) that with high probability,

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} = \frac{1}{|\mathbf{S}_1(k_1)||\mathbf{S}_l(k_2)|} + O\left(\frac{1}{n^3 p_n^5}\right). \tag{65}$$

By the condition $p_n \gg (\frac{1}{n})^{1/2}$ and (65), with probability tending to 1, the first term in (60) is bounded from above by

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i \in \mathbf{S}_1(k_1), j \in \mathbf{S}_l(k_2)} (b_{ij} - p^{(l)}_{k_1 k_2}) = O\left(\frac{1}{n p_n}\right) \ll p_n. \tag{66}$$

By (56), we have

$$\sum_{a_{1i}=1, i \in \mathbf{S}_1} 1 = O(\|\mathbf{SC} - \mathbf{SQO}\|_F^2).$$

Therefore, we can see that

$$\mathbb{E} \sum_{a_{1i}=1, i \in M} 1 = O(\frac{1}{p_n}). \tag{67}$$

37

First of all, we have

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i\in\widehat{\mathbf{S}}_1(k_1)\backslash\mathbf{S}_1(k_1),\ j\in\mathbf{S}_l(k_2)} |b_{ij} - p^{(l)}_{k_1k_2}|$$

$$= \frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i\in\widehat{\mathbf{S}}_1(k_1)\backslash\mathbf{S}_1(k_1),\ j\in\mathbf{S}_l(k_2)} |1 - p^{(l)}_{k_1k_2}| I(b_{ij}=1)$$

$$+ \frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i\in\widehat{\mathbf{S}}_1(k_1)\backslash\mathbf{S}_1(k_1),\ j\in\mathbf{S}_l(k_2)} p^{(l)}_{k_1k_2} I(b_{ij}=0). \tag{68}$$

For the first term of the right hand side of (68), we have

$$\mathbb{E} \sum_{i\in\widehat{\mathbf{S}}_1(k_1)\backslash\mathbf{S}_1(k_1),\ j\in\mathbf{S}_l(k_2)} |1 - p^{(l)}_{k_1k_2}| I(b_{ij}=1) \le \sum_{a_{1i}=1\in\mathcal{M},\ j\in\mathbf{S}_l(k_2)} (\mathbb{E}I(b_{ij}=1)) = O(|S_l(k_2)|). \tag{69}$$

Moreover, considering the second term of the right hand side of (68), by (67) we have

$$\mathbb{E} \sum_{i\in\widehat{\mathbf{S}}_1(k_1)\backslash\mathbf{S}_1(k_1),\ j\in\mathbf{S}_l(k_2)} p^{(l)}_{k_1k_2} I(b_{ij}=0) = O(p_n \sum_{a_{1i}=1,i\in\mathcal{M},\ j\in\mathbf{S}_l(k_2)} (\mathbb{E}I(b_{ij}=0)))$$

$$= O(|S_l(k_2)|). \tag{70}$$

Therefore, we imply that

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i\in\widehat{\mathbf{S}}_1(k_1)\backslash\mathbf{S}_1(k_1),\ j\in\mathbf{S}_l(k_2)} |b_{ij} - p^{(l)}_{k_1k_2}| = O_p\left(\frac{1}{np_n}\right) \ll p_n. \tag{71}$$

Similar to (71), we have

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i\in\mathbf{S}_1(k_1),\ j\in\widehat{\mathbf{S}}_l(k_2)\backslash\mathbf{S}_l(k_2)} |b_{ij} - p_{k_1k_2}| = O_p\left(\frac{1}{np_n}\right) \ll p_n, \tag{72}$$

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i\in\widehat{\mathbf{S}}_1(k_1)\backslash\mathbf{S}_1(k_1),\ j\in\widehat{\mathbf{S}}_l(k_2)\backslash\mathbf{S}_l(k_2)} |b_{ij} - p_{k_1k_2}| = O_p\left(\frac{1}{np_n}\right) \ll p_n, \tag{73}$$

and

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \Big[\sum_{i\in\mathbf{S}_1(k_1)\backslash\widehat{\mathbf{S}}_1(k_1),\ j\in\mathbf{S}_l(k_2)\cap\widehat{\mathbf{S}}_l(k_2)} |b_{ij} - p^{(l)}_{k_1k_2}|$$

$$+ \sum_{i\in\mathbf{S}_1(k_1)\cap\widehat{\mathbf{S}}_1(k_1),\ j\in\mathbf{S}_l(k_2)\backslash\widehat{\mathbf{S}}_l(k_2)} |b_{ij} - p^{(l)}_{k_1k_2}| + \sum_{i\in\mathbf{S}_1(k_1)\backslash\widehat{\mathbf{S}}_1(k_1),\ j\in\mathbf{S}_l(k_2)\backslash\widehat{\mathbf{S}}_l(k_2)} |b_{ij} - p^{(l)}_{k_1k_2}| \Big]$$

$$= O_p\left(\frac{1}{np_n}\right) \ll p_n. \tag{74}$$

<div align="center">38</div>

By (61) and (74), we have

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|}\left|\sum_{i\in\mathbf{S}_1(k_1)\cap\widehat{\mathbf{S}}_1(k_1),j\in\mathbf{S}_l(k_2)\cap\widehat{\mathbf{S}}_l(k_2)}(b_{ij}-p_{k_1k_2}^{(l)})\right|\ll p_n\,. \tag{75}$$

Therefore, by (60), (66)–(73) and (75), we have $\max_{i,j\in[K]}|\widehat{p}_{k_1k_2}^{(l)}-p_{k_1k_2}^{(l)}|\ll p_n$. In view of this result, if $\widehat{f}_0=f_0$, it follows from (59) that

$$\|\widehat{\mathbf{P}}_{(f_0,f_0)}^{\mathbf{S},\mathbf{S}}-\widehat{\mathbf{P}}_{(f_0,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}\|_F\ll p_n\,.$$

Otherwise if $\widehat{f}_0\neq f_0$, by Condition 5 and Proposition 2, we have

$$\|\widehat{\mathbf{P}}_{(\widehat{f}_0,\widehat{f}_0)}^{\mathbf{S},\mathbf{S}}-\widehat{\mathbf{P}}_{(\widehat{f}_0,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}\|_F\sim p_n\,.$$

Recall that $\widehat{f}_0$ is defined by $\widehat{f}_0=\arg\min_f\|\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S},\mathbf{S}}-\widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}\|_F$. Hence, with probability tending to 1, we have $\widehat{f}_0=f_0$. Therefore, by Algorithm 2 with probability tending to 1, the set R can be merged into

$$\widetilde{\mathrm{R}}=\left\{\{\mathbf{e}_1^\top\mathcal{D}\mathcal{Q}_1,\mathbf{e}_1^\top\mathcal{D}\mathcal{Q}_2\},\ldots,\{\mathbf{e}_K^\top\mathcal{D}\mathcal{Q}_1,\mathbf{e}_K^\top\mathcal{D}\mathcal{Q}_2\}\right\}\,.$$

Notice that the $2K$ clusters are merged into K communities According to $\widetilde{\mathrm{R}}$. Then it follows from (57) that with probability tending to 1, for $i,j\in\mathcal{M}^c$,

$$\text{Individuals } i \text{ and } j \text{ are assigned to the same community}\Leftrightarrow\boldsymbol{\pi}_i=\boldsymbol{\pi}_j\,.$$

Furthermore, it holds with probability tending to 1 that

$$1-\text{Misclustering rate}\geq\frac{|\{i\in\mathcal{M}^c:i\text{ is assigned to group }k\text{ such that }\boldsymbol{\pi}_i=\mathbf{e}_k\}|}{n}$$

$$=\frac{|\mathcal{M}^c|}{n}=1-\frac{|\mathcal{M}|}{n}\,,$$

in which the first equality follows from Lemma 3. By Lemma 10, we have with probability tending to 1, $|\mathcal{M}|=O(1/p_n^2)$. Then $p_n\gg(1/n)^{1/2}$ implies that $|\mathcal{M}|/n=o(1)$. ∎

B.5 Proof of Proposition 1

Actually, (20) follows from Lemma 4.1 of Lei and Rinaldo (2015) directly. Therefore we only need to prove the first statement of Proposition 1. Conditions 4 and 6, together with Lemma 17 imply that

$$n\mathbf{I}\lesssim\boldsymbol{\Pi}^\top\mathbb{E}\Theta^2\boldsymbol{\Pi}\lesssim n\mathbf{I},\ \|\boldsymbol{\Pi}^\top\left(\Theta^2-\mathbb{E}\Theta^2\right)\boldsymbol{\Pi}\|=o(n)\,, \tag{76}$$

holds with high probability. By (76), the remaining proof of Proposition 1 is almost the same as the proof of Lemma 4 and thus we omit it.

B.6 Proof of Theorem 4

Similar to the way we prove Theorem 3, we first need some auxilliary results regarding the theoretical properties of the clustering step. The next lemma is analogous to Lemma 8.

Lemma 11 *Under Conditions 4 and 6, with high probability, there exists some positive constant c_2 such that*

$$\mathcal{Q}\mathcal{Q}^\top \geq (c_2 n)^{-1}\mathbf{I}, \quad \mathcal{D}\mathcal{Q}_2\mathcal{Q}_2^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I}, \quad and \quad p_n\mathcal{D}\mathcal{Q}_1\mathcal{Q}_1^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I}. \quad (77)$$

Proof By checking the proof of Lemma 8 carefully, we can see that the order of the matrices in (77) mainly depends on the order of $\mathbf{V}^\top\mathbf{S}\mathbf{V}$. Actually, by almost the same proof as Lemma 2, the conclusion of Lemma 2 holds by Proposition 1 and Condition 6. Therefore our proof is finished. ∎

Let

$$\widehat{\mathbf{E}}_\mathbf{S} = \min_{\mathbf{E}\in\{K\times K \text{ permutation matrices}\}} \|\mathbf{S}\widehat{\mathbf{\Pi}}\mathbf{E} - \mathbf{S}\mathbf{\Pi}\|_0,$$

$$\widehat{\mathbf{E}}_{\mathbf{I}-\mathbf{S}} = \min_{\mathbf{E}\in\{K\times K \text{ permutation matrices}\}} \|(\mathbf{I}-\mathbf{S})\widehat{\mathbf{\Pi}}\mathbf{E} - (\mathbf{I}-\mathbf{S})\mathbf{\Pi}\|_0$$

and $M = \{i, a_{1i} = 1, \widehat{\mathbf{\Pi}}(i)\widehat{\mathbf{E}}_\mathbf{S} \neq \mathbf{\Pi}(i)\} \cup \{i, a_{1i} = 0, \widehat{\mathbf{\Pi}}(i)\widehat{\mathbf{E}}_{\mathbf{I}-\mathbf{S}} \neq \mathbf{\Pi}(i)\}$. Similar to Theorem 4.2 of Lei and Rinaldo (2015) we have the following result.

Lemma 12 *Under Conditions 4 and 6, if $p_n \gg \frac{1}{\sqrt{n}f_s}$, with probability tending to 1 we have*

$$|M| \lesssim \frac{\sqrt{n}}{f_s p_n}.$$

Proof Let $\mathbf{Q}'$ be the row normalized version of $\mathbf{Q}$ and $\mathbf{Q}'_{\mathcal{I}_1}$ be the sub matrix of $\mathbf{Q}'$ corresponding to the non zero rows of $\mathbf{W}$. Then by the inequality in front of the proof of Theorem 4.2 in Lei and Rinaldo (2015), with high probability we have

$$\begin{aligned}
\|\mathbf{W}' - \mathbf{Q}'_{\mathcal{I}_1}\mathbf{O}\|_{2,1} &\leq 2\sum_{i=1}^n \frac{\|\mathbf{W}(i) - \mathbf{Q}_{\mathcal{I}_1}(i)\mathbf{O}\|}{\|\mathbf{Q}_{\mathcal{I}_1}(i)\|} \\
&\leq 2\sqrt{\sum_{i=1}^n \|\mathbf{W}(i) - \mathbf{Q}_{\mathcal{I}_1}(i)\mathbf{O}\|^2 \sum_{i=1}^n \|\mathbf{Q}_{\mathcal{I}_1}(i)\|^{-2}} \\
&\lesssim n f_s^{-1}\|\mathbf{W} - \mathbf{Q}\mathbf{O}\|_F.
\end{aligned} \quad (78)$$

Similar to the proof of Lemma 10, the misclustered nodes can be bounded by considering $\mathbf{S}\mathbf{W}$ and $(\mathbf{I}-\mathbf{S})\mathbf{W}$ separately. The rest of the proof is essentially the same as the proof of Theorem 4.2 in Lei and Rinaldo (2015); the key points are (78), Lemma 11 and Lemma 6 in this setting. Therefore we omit the proof. ∎

The next result justifies our merging strategy in Algorithm 3.

Proposition 3 *Let $\mathfrak{C} = \{\mathfrak{c}_1, \ldots, \mathfrak{c}_K\}$ and $\mathfrak{D} = \{\mathfrak{d}_1, \ldots, \mathfrak{d}_K\}$ be two ordered sets of size K and let $f_0 : [K] \to [K]$ be the unique permutation such that $\mathfrak{c}_{f_0(i)} = \mathfrak{d}_i$ for all $i \in [K]$. Suppose that the K elements in these ordered sets have been endowed with pairwise connection probabilities. Let $\mathbf{C} = (c_{ij})$ be a $K \times K$ matrix such that the (i,j)-th entry c_{ij} is the connection probability between $\mathfrak{c}_i$ and $\mathfrak{c}_j$, and $\widetilde{\mathbf{D}} = (d_{ij}/e_i)$ be a $K \times K$ matrix such that the (i,j)-th entry d_{ij} is the connection probability between $\mathfrak{c}_i$ and $\mathfrak{d}_j$. Let $\mathbf{X} = diag(x_1, \ldots, x_K)$ be a $K \times K$ diagonal matrix, where the entries of $\mathbf{X}$ are bounded below and above. If $\mathrm{rank}(\mathbf{C}) = K$, then for a permutation function $f : [K] \to [K]$,*

$$f = f_0, x_i = e_{f(i)}, i = 1, \ldots, K \iff \mathbf{C}_{(f,f)} = \widetilde{\mathbf{D}}_{(f,*)}\mathbf{X},$$

where the (i,j)-th entries of $\mathbf{C}_{(f,f)}$ and $\widetilde{\mathbf{D}}_{(f,*)}$ are $c_{f(i)f(j)}$ and $d_{f(i)j}/e_{f(i)}$, respectively.

Proof By similar argument as the proof of Proposition 2, we can also imply the columns (rows) of $\mathbf{D}$ ($\widetilde{\mathbf{D}}$) are distinct. Therefore, by the definition of $\mathbf{D}$, if

$$\mathbf{C}_{(f,f)} \neq \mathbf{D}_{(f,*)}.$$

It is not hard to see that there does not exist a diagonal matrix $\mathbf{X}$ such that

$$\mathbf{C}_{(f,f)} = \widetilde{\mathbf{D}}_{(f,*)}\mathbf{X}.$$

Hence, our proof is reduced to the proof of Proposition 2 and thus we omit the remaining part. ∎

Now to prove Theorem 4, it suffices for us to prove the following result.

Theorem 6 *Under Conditions 4–6 and $p_n \gg \frac{1}{\sqrt{n}f_s}$, it holds with probability tending to 1, uniformly for $i, j \in \mathrm{M}^c$, that*

individuals i and j are assigned to the same community by Algorithm 3 $\Leftrightarrow$ $\boldsymbol{\pi}_i = \boldsymbol{\pi}_j$.

Proof By Lemma 12, we have

$$|\mathrm{M}| = O\left(\frac{\sqrt{n}}{f_s p_n}\right). \tag{79}$$

Similar to the proof of Theorem 3, we say the 1st group is $\{i \in [n] : a_{1i} = 1\}$ and the 2nd group is $\{i \in [n] : a_{1i} = 0\}$. Let $\mathbf{S}_l(k)$ be the collection of individuals belonging to k-th community in the l-th group, and $\widehat{\mathbf{S}}_l(k)$ be the k-th cluster of the l-th group returned by step 2. Let $\widehat{p}^{(1)}_{k_1 k_2}$ and $\widehat{p}^{(2)}_{k_1 k_2}$ be the (k_1, k_2)-th entry of $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{I}-\mathbf{S}}$ (defined by equation (17)) respectively. Let $\mathbf{P}^{\mathbf{S},\mathbf{S}} = (p^{(1)}_{k_1 k_2})$ and $\mathbf{P}^{\mathbf{S},\mathbf{I}-\mathbf{S}} = (p^{(2)}_{k_1 k_2})$ be the corresponding population versions. Note that there exists an unique permutation function f_0 and the corresponding diagonal matrix g such that

$$\mathbf{P}^{\mathbf{S},\mathbf{S}}_{f_0,f_0} = \mathbf{P}^{\mathbf{S},\mathbf{I}-\mathbf{S}}_{f_0,*} g(\mathbf{P}^{\mathbf{S},\mathbf{S}}_{(f_0,f_0)}, \mathbf{P}^{\mathbf{S},\mathbf{I}-\mathbf{S}}_{(f_0,*)}). \tag{80}$$

Without loss of generality, we assume $\mathbf{P}^{\mathbf{S},\mathbf{I}-\mathbf{S}} = \mathbf{P}$.

$$
\begin{aligned}
\widehat{p}_{k_1 k_2}^{(l)} - p_{k_1 k_2}^{(l)} = & \frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \Bigg[\sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \\
& + \sum_{i \in \widehat{\mathbf{S}}_1(k_1) \setminus \mathbf{S}_1(k_1),\ j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \\
& + \sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1),\ j \in \widehat{\mathbf{S}}_l(k_2) \setminus \mathbf{S}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \\
& + \sum_{i \in \widehat{\mathbf{S}}_1(k_1) \setminus \mathbf{S}_1(k_1),\ j \in \widehat{\mathbf{S}}_l(k_2) \setminus \mathbf{S}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \Bigg],
\end{aligned}
\tag{81}
$$

and

$$
\begin{aligned}
& \frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i \in \mathbf{S}_1(k_1), j \in \mathbf{S}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \\
& = \frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \Bigg[\sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \\
& + \sum_{i \in \mathbf{S}_1(k_1) \setminus \widehat{\mathbf{S}}_1(k_1),\ j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \\
& + \sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1),\ j \in \mathbf{S}_l(k_2) \setminus \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \\
& + \sum_{i \in \mathbf{S}_1(k_1) \setminus \widehat{\mathbf{S}}_1(k_1),\ j \in \mathbf{S}_l(k_2) \setminus \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \Bigg].
\end{aligned}
\tag{82}
$$

By Lemma 17, with high probability, we have

$$
|\mathbf{S}_1(k)| = \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} a_{1i} \sim \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} \mathbb{E} a_{1i} \sim n p_n, k \in [K],
\tag{83}
$$

and

$$
|\mathbf{S}_2(k)| = \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} (1 - a_{1i}) \sim \sum_{i \in [n]: \boldsymbol{\pi}_i = \mathbf{e}_k} \mathbb{E}(1 - a_{1i}) \sim n, k \in [K],
\tag{84}
$$

Moreover, by (79), with high probability we have

$$
\left| |\widehat{\mathbf{S}}_l(k)| - |\mathbf{S}_l(k)| \right| \le |\widehat{\mathbf{S}}_l(k) \setminus \mathbf{S}_l(k)| + |\mathbf{S}_l(k) \setminus \widehat{\mathbf{S}}_l(k)| \le |M| = O\left(\frac{\sqrt{n}}{p_n f_s} \right), l = 1, 2,\ k \in [K].
\tag{85}
$$

It follows from (83)–(85) that with high probability,

$$
\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} = \frac{1}{|\mathbf{S}_1(k_1)||\mathbf{S}_l(k_2)|} + O\left(\frac{\sqrt{n}}{f_s n^3 p_n^4} \right).
\tag{86}
$$

By $p_n \gg 1/(\sqrt{n} f_s)$ and (86) with probability tending to 1, we have

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i \in \mathbf{S}_1(k_1), j \in \mathbf{S}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) = O\left(\frac{1}{np_n}\right) \ll p_n. \tag{87}$$

Similar to (69)–(70), By (79), the condition that $p_n \gg 1/(\sqrt{n} f_s)$, we have

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i \in \widehat{\mathbf{S}}_1(k_1) \backslash \mathbf{S}_1(k_1), \ j \in \mathbf{S}_l(k_2)} |b_{ij} - p_{k_1 k_2}^{(l)}|$$
$$= O_p\left(\frac{1}{\sqrt{n} f_s}\right) \ll p_n. \tag{88}$$

Similar to (88), with probability tending to 1 we have

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i \in \mathbf{S}_1(k_1), \ j \in \widehat{\mathbf{S}}_l(k_2) \backslash \mathbf{S}_l(k_2)} |b_{ij} - p_{k_1 k_2}| = O\left(\frac{1}{\sqrt{n} f_s}\right) \ll p_n, \tag{89}$$

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \sum_{i \in \widehat{\mathbf{S}}_1(k_1) \backslash \mathbf{S}_1(k_1), \ j \in \widehat{\mathbf{S}}_l(k_2) \backslash \mathbf{S}_l(k_2)} |b_{ij} - p_{k_1 k_2}| = O\left(\frac{1}{\sqrt{n} f_s}\right) \ll p_n. \tag{90}$$

and

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \Big[\sum_{i \in \mathbf{S}_1(k_1) \backslash \widehat{\mathbf{S}}_1(k_1), \ j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} |b_{ij} - p_{k_1 k_2}^{(l)}|$$
$$+ \sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), \ j \in \mathbf{S}_l(k_2) \backslash \widehat{\mathbf{S}}_l(k_2)} |b_{ij} - p_{k_1 k_2}^{(l)}| + \sum_{i \in \mathbf{S}_1(k_1) \backslash \widehat{\mathbf{S}}_1(k_1), \ j \in \mathbf{S}_l(k_2) \backslash \widehat{\mathbf{S}}_l(k_2)} |b_{ij} - p_{k_1 k_2}^{(l)}| \Big]$$
$$= O\left(\frac{1}{\sqrt{n} f_s}\right) \ll p_n. \tag{91}$$

By (82) and (91), we have

$$\frac{1}{|\widehat{\mathbf{S}}_1(k_1)||\widehat{\mathbf{S}}_l(k_2)|} \left| \sum_{i \in \mathbf{S}_1(k_1) \cap \widehat{\mathbf{S}}_1(k_1), j \in \mathbf{S}_l(k_2) \cap \widehat{\mathbf{S}}_l(k_2)} (b_{ij} - p_{k_1 k_2}^{(l)}) \right| \ll p_n. \tag{92}$$

Considering the estimation of g, by Lemma 17, with probability tending to 1 we have

$$\|g(\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S},\mathbf{S}}, \widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S},\mathbf{I-S}}) - g(\mathbf{P}_{(f,f)}^{\mathbf{S},\mathbf{S}}, \mathbf{P}_{(f,*)}^{\mathbf{S},\mathbf{I-S}})\| = O(\frac{\sqrt{\log n}}{np_n}). \tag{93}$$

Therefore, by almost the same arguments as the proof of Theorem 3, our conclusion holds. ∎

B.7 Proof of Lemma 7

By Lemma 4.1 of Lei and Rinaldo (2015) and Condition 7, $\mathbf{V}(i) \sim \frac{\theta_i}{\sqrt{n}}\mathbf{H}_k$, i belongs to community k, where $\{\mathbf{H}_k, k = 1, \ldots, K\}$ are orthonormal vectors. Therefore, we imply that

$$\mathbf{V}^T\mathbf{S}\mathbf{V} = \sum_{\mathbf{S}_{ii}=1} \mathbf{V}(i)\mathbf{V}^T(i) \sim \sum_{\mathbf{S}_{ii}=1, i \text{ belongs to community } k} \frac{\theta_i^2}{n}\mathbf{H}_k\mathbf{H}_k^\top.$$

Combining this with the assumption that there are n_{1k} θ_i's generated from F_1 and n_{2k} θ_i's from F_2, it is easy to see that our conclusion holds.

B.8 Proof of of Theorem 5

To prove Theorem 5, we first need to show results analogous to the unconditional case hold for the conditonal setting. First note that by Lemma 17, we have that

$$\max_{k \in [K]}(n_{1k} + n_{2k}) = O(np_n), \tag{94}$$

with high probability. Using this result and by almost the same proof as the first part of Theorem 2 and the concentration inequality (94), the non-zero eigenvalues of $\mathbf{B}_E$ satisfy the following.

Theorem 7 *Under Conditions 7 and 8, for a given $\mathbf{S}$, we have*

$$|x_i|^{-1} \sim np_n \min_{k \in [K]} \sqrt{\frac{n_{2k}\mu_2^2 + n_{1k}\mu_1^2}{n}}$$

for $i \in [\pm K]$.

By Theorem 7, we can conclude that Lemma 3 holds for a given $\mathbf{S}$ under Conditions 7 and 8, in the following sense.

Lemma 13 *Under Conditions 7 and 8. For fixed $\mathbf{S}$, it holds with high probability that*

$$\|\mathbf{B} - \mathbf{B}_E\| = O(\sqrt{np_n}).$$

Moreover, with high probability we have $\|\mathbf{B} - \mathbf{B}_E\| \ll \min_{i \in [2K]} \sigma_i(\mathbf{B}_E)$.

Proof The first statement of Lemma 13 follows the same argument as the proof of Lemma 3 by noticing that $\|\mathbf{S}\| \leq 1$ for fixed $\mathbf{S}$. The second statement follows from Theorem 7 directly and thus we omit the proof. ∎

Furthermore, Theorem 7 and almost the same arguments as in Lemma 6 give another analogous result below (and we omit the proof).

Lemma 14 *Under Conditions 7 and 8, conditioned on $\mathbf{S}$, w.h.p. we have*

$$\|\mathbf{W} - \mathbf{Q}\mathbf{O}\|_F = O\left(\frac{1}{\sqrt{\min_{k \in [K]}(n_{2k}\mu_2^2 + n_{1k}\mu_1^2)p_n}}\right), \tag{95}$$

where $\mathbf{O}$ is the same as defined in Lemma 6.

The next lemma is analogous to Lemma 11. We omit the proof as it is almost identical.

Lemma 15 *Under Conditions 7 and 8, for fixed* $\mathbf{S}$, *with high probability, there exists some positive constant* c_2 *such that*

$$\mathcal{D}\mathcal{Q}_2\mathcal{Q}_2^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I}, \quad and \quad \frac{\min_k n_{2k}}{n}\mathcal{D}\mathcal{Q}_1\mathcal{Q}_1^\top\mathcal{D}^\top \geq (c_2 n)^{-1}\mathbf{I}. \tag{96}$$

Then corresponding to Lemma 12, we also have the following result, whose proof we again omit due to its similarity.

Lemma 16 *Under Conditions 7 and 8, if* $p_n \gg \frac{1}{\min_{k\in[K]}(n_{2k}\mu_2^2+n_{1k}\mu_1^2)\mu_1^2}$, *then conditioned on* $\mathbf{S}$ *and* Θ, *with high probability we have*

$$|\mathrm{M}| \lesssim \mu_1^{-1}n\sqrt{\frac{1}{p_n\min_{k\in[K]}(n_{2k}\mu_2^2+n_{1k}\mu_1^2)}}.$$

Now to prove Theorem 5, it suffices to prove the following, the proof of which is omitted due to its similarity to Theorem 6.

Theorem 8 *Under Conditions 5, 7 and 8, the same conclusion as Theorem 6 holds, conditioned on* $\mathbf{S}$ *and* Θ, *if*

$$\mu_1^{-2}\frac{1}{\min_{k\in[K]}(n_{2k}\mu_2^2+n_{1k}\mu_1^2)} \ll p_n.$$

B.9 Concentration inequalities

The following two concentration inequalities are used throughout the paper.

Lemma 17 (Bernstein inequality) *Suppose that* $\{y_i\}_{i=1}^n$ *are independent bernoulli random variables, then for any non-random series* $\{a_i\}_{i=1}^n$ *such that* $|a_i| \leq L$ *for some positive constant* L, *we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^n a_i(y_i - \mathbb{E}y_i)\right| \geq t\right) \leq \exp\left(-\frac{t^2/2}{\sum_{i=1}^n(a_i^2\mathbb{E}(y_i - \mathbb{E}y_i)^2) + \frac{Lt}{3}}\right), \quad t > 0. \tag{97}$$

Lemma 18 (Matrix Bernstein inequality, Theorem 6.2 of Tropp (2012)) *Consider a finite sequence* $\{\mathbf{X}_k\}$ *of independent, random, self-adjoint* $d \times d$ *matrices. Assume that*

$$\mathbb{E}\mathbf{X}_k = \mathbf{0} \quad and \quad \mathbb{E}\left(\mathbf{X}_k^p\right) \leq \frac{p!}{2}\cdot R^{p-2}\mathbf{A}_k^2 \quad for\ p = 2, 3, 4, \ldots$$

Compute the variance parameter

$$\sigma^2 := \left\|\sum_k \mathbf{A}_k^2\right\|.$$

Then the following chain of inequalities holds for all $t \geq 0$

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum_k \mathbf{X}_k\right) \geq t\right\} \leq d\cdot\exp\left(\frac{-t^2/2}{\sigma^2 + Rt}\right) \leq \begin{cases} d\cdot\exp\left(-t^2/4\sigma^2\right) & for\ t \leq \sigma^2/R, \\ d\cdot\exp(-t/4R) & for\ t \geq \sigma^2/R. \end{cases}$$

Appendix C. Extensions of theoretical results

C.1 Exact Recovery

We prove the exact recovery result for stochastic block model under stronger condition by applying similar approach as Su et al. (2019). By checking the proof of Theorem 2.3 in Su et al. (2019), the crucial step is to prove the almost sure convergence for the entries of spiked eigenvectors. By Wu (2021) and the Borel-Cantelli Lemma, noticing that the eigenvector bound $O_p(x_K^2\sqrt{np_n})$ is calculated by second moment of the small order terms, we have the following Lemma.

Condition 9 *Condition 4 holds. Moreover, we assume that $p_n \geq (n/\log n)^{-2/7}$ and $|x_i/x_j - 1| \geq c$ for some positive constant c.*

Lemma 19 *Under Condition 9, we have the following expansion:*

$$\mathbf{e}_i^\top \mathbf{w}_k = \mathbf{e}_i^\top \mathbf{q}_k + o_{a.s.}(\frac{1}{n^{1/2}}), \tag{98}$$

uniformly for all $1 \leq i \leq n$.

Proof The proof of Lemma 19 is essentially a modification of Theorem 2.2.1 of Wu (2021). Noticing that the proof of Theorem 2.2.1 in Wu (2021) essentially rely on the upper bound of $Var(\mathbf{x}^\top(\mathbf{B} - \mathbf{B}_E)^l \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are unit vectors(maybe depending on $\mathbf{S}$). Following almost the same steps as Theorem 2.2.1 of Wu (2021), we calculate the fourth moment of $\mathbf{e}_i^\top(\mathbf{B} - \mathbf{B}_E)^2 - \mathbb{E}(\mathbf{B} - \mathbf{B}_E)^2 \mathbf{q}_k$ and have the following result

$$\mathbf{e}_i^\top \left((\mathbf{B} - \mathbf{B}_E)^2 - \mathbb{E}(\mathbf{B} - \mathbf{B}_E)^2\right) \mathbf{q}_k = O_{L_4}(\sqrt{np_n}).$$

Therefore, we have an improved version of Theorem 2.2.1 in Wu (2021) such that

$$\mathbf{e}_i^\top \mathbf{w}_k = \mathbf{e}_i^\top \mathbf{q}_k + x_i \mathbf{e}_i^\top(\mathbf{B} - \mathbf{B}_E)\mathbf{q}_k + O_{L_4}(|x_k|^2\sqrt{np_n}) + O_{L_2}(|x_k|^3 np_n) + O_{a.s.}(\frac{1}{n}), \tag{99}$$

where $X_n = O_{L_{2k}}(s_n)$ means that $\mathbb{E}X_n^{2k} = O(s_n^{2k})$, $k \in \mathbb{N}$, the last term $O_{a.s}(\frac{1}{n})$ holds uniformly for all $1 \leq i \leq n$. By Theorem 2, we have $|x_i|^{-1} \sim np_n^{3/2}$. By the condition that $p_n \geq (n/\log n)^{-2/7}$, we have $|x_i|^2\sqrt{np_n} = o(n^{-1/2-1/4-1/36})$ and $|x_i|^3 np_n = O(n^{-1/2-1/2}\log^{-1} n)$. Combining this with the Borel-Cantelli Lemma, we imply that (99) holds uniformly for i and k, such that

$$\mathbf{e}_i^\top \mathbf{w}_k = \mathbf{e}_i^\top \mathbf{q}_k + x_k \mathbf{e}_i^\top(\mathbf{B} - \mathbf{B}_E)\mathbf{q}_k + o_{a.s.}(\frac{1}{\sqrt{n}}). \tag{100}$$

Recalling that $\mathbf{q}_k = \mathbf{SV}\mathbf{q}_{1k} + (\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2k}$, by Corollary 2.6.1 of Wu (2021) we have

$$\mathbf{q}_{1k} = O_{a.s.}(1/\sqrt{p_n}), \ \mathbf{q}_{2k} = O_{a.s.}(\sqrt{p_n}).$$

It follows from Lemma 17 that

$$\mathbb{P}(\mathbf{e}_i^\top(\mathbf{B} - \mathbf{B}_E)\mathbf{SV}\mathbf{q}_{1k} > 4\sqrt{p_n}\log n) \leq \frac{1}{n^4}.$$

Similarly we have

$$\mathbb{P}(\mathbf{e}_i^\top (\mathbf{B} - \mathbf{B}_E)(\mathbf{I} - \mathbf{S})\mathbf{V}\mathbf{q}_{2k} > 4\sqrt{p_n} \log n) \leq \frac{1}{n^4}.$$

Therefore we imply that $x_k \mathbf{e}_i^\top (\mathbf{B} - \mathbf{B}_E)\mathbf{q}_k = o_{a.s.}(1/\sqrt{n})$ uniformly for all $1 \leq i \leq n$. ∎

According to Lemma 19, the deviation $(\mathbf{e}_i^\top \mathbf{w}_k - \mathbf{e}_i^\top \mathbf{q}_k)$ can be bounded by controlling the order of $x_k \mathbf{e}_i^\top (\mathbf{B} - \mathbf{B}_E)\mathbf{q}_k$. In fact, given $\mathbf{S}$, $\mathbf{q}_k$ is a non-random vector, and it is not hard to see that $x_k \mathbf{e}_i^\top (\mathbf{B} - \mathbf{B}_E)\mathbf{q}_k = o_{a.s.}(\frac{1}{n^{1/2}})$ by Lemma 17. Combining this with Lemma 5 in the main paper, we immediately obtain the following result bounding the distances between different types of rows in $\mathbf{Q}$.

Lemma 20 *Under Condition 9, we imply the following results:*

- *There exists a positive constant C such that $\sqrt{n}|\mathbf{e}_i^\top \mathbf{q}_k| \leq C$.*

- *There exists deterministic sequences $c_{2n} = o(1)$ and $c_{1n} \sim 1$ such that $\sqrt{n}|\mathbf{e}_i^\top (\mathbf{w}_k - \mathbf{q}_k)| \leq c_{2n}$, a.s. $\sqrt{n}\|\mathbf{e}_i^\top \mathbf{Q} - \mathbf{e}_j \mathbf{Q}\| \geq c_{1n}$, $\pi_i \neq \pi_j$ or $a_{1i} \neq a_{1j}$.*

The validity of Assumption 4 in Su et al. (2019) is implied by Lemma 20 above. Combining this with Condition 9, it is easy to check that Lemma 2.2 in Su et al. (2019) holds. Following the same argument as in the proof of Theorem 2.3 in Su et al. (2019), we establish the exact recovery result for clustering the two groups of nodes $\{i \in [n], a_{1i} = 1\}$ and $\{i \in [n], a_{1i} = 0\}$ using Algorithm 1. Finally, we can apply the rest of arguments in the proof of Theorem 3 together with the exact recovery result and show that

$$\|\widehat{\mathbf{P}}_{(f_0, f_0)}^{\mathbf{S},\mathbf{S}} - \widehat{\mathbf{P}}_{(f_0, *)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}\|_F = o_{a.s.}(p_n),$$

and

$$\|\widehat{\mathbf{P}}_{(\widehat{f}_0, \widehat{f}_0)}^{\mathbf{S},\mathbf{S}} - \widehat{\mathbf{P}}_{(\widehat{f}_0, *)}^{\mathbf{S},\mathbf{I}-\mathbf{S}}\|_F \sim p_n, \quad a.s.$$

Therefore Theorem 3 holds almost surely and the proof of exact recovery is completed.

Finally, we make the following remarks about the exact recovery result above. (i) Since Condition 19 is significantly more stringent than Condition 4 in the main paper, we have chosen to extend our theoretical results to the SBM setting only for illustrative purposes. (ii) While the lower bound on p_n in Condition 19 may not be optimal, we note that the arguments already involve novel theoretical results from Wu (2021), which is work currently in submission by the authors. We will leave refining this bound as one possible direction for future work. (iii) The distinct eigenvalue assumption in Condition 19 is made in our proof of Lemma 19, as we refer to Wu (2021) for establishing exact recovery, where this assumption is required. In other words, if the eigenvector expansion can be extended to allow eigenvalue multiplicity, exact recovery without this assumption can be proved. Moreover, considering the treatment of eigenvalue multiplicity in establishing asymptotic expansions of eigenvectors in Han et al. (2023), we believe that both the results in Wu (2021) and our Lemma 19 still hold without assuming distinct eigenvalues. We plan to adopt a similar approach as presented in the proof of Han et al. (2023) to improve the eigenvector expansion of Wu (2021). Investigating this extension would require a significant amount of work and thus we will leave it for future research.

C.2 Extension to K growing with n

It is possible to extend our results to the case of $K \to \infty$ using the current approaches. However, as mentioned earlier, observing only a partial network has already resulted in a significant amount of signal loss, and an increasing K would only make the conditions more stringent. For clarity reasons, we choose to focus on $K = O(1)$ as this paper is the first attempt at investigating the proposed partial information framework. To get a sense of how conditions on K and p_n change for the main theorems, we give the following outline of key points.

- Theorem 1, which describes the eigenvalues and eigenvectors of $\mathbf{B}_E$, only requires $\mathbf{V}^\top \mathbf{S} \mathbf{V}$ and $\mathbf{I} - \mathbf{V}^\top \mathbf{S} \mathbf{V}$ to be invertible. Since Theorem 2 further establishes the order of the eigenvalues and thus requires a stronger condition on K, we will first consider Theorem 2 for $K \to \infty$. Theorem 2 relies on Lemma 2; inspecting the proof of this lemma, a crucial step involves establishing the entrywise inequality $|\mathbf{v}_i^\top (\mathbf{S} - \mathbb{E}\mathbf{S})\mathbf{v}_j| \leq C p_n \sqrt{\frac{\log n}{n p_n}}$ with high probability, for some positive constant C. It follows then $\|\mathbf{V}^\top (\mathbf{S} - \mathbb{E}\mathbf{S})\mathbf{V}\|_F \leq K C p_n \sqrt{\frac{\log n}{n p_n}}$, and requiring $K \ll \sqrt{\frac{n p_n}{\log n}}$ would lead to the key result in Eq (34) for proving the lemma.

- In Condition 3, given that $\sum_{i=1}^K d_i^2 \leq n^2 p_n^2$, the condition should be revised as $|d_1| \sim |d_2| \sim \ldots \sim |d_K| \sim n p_n / \sqrt{K}$. Consequently, the order of $|x_i|^{-1}$ now becomes $n p_n^{3/2} / \sqrt{K}$ in Theorem 2.

- For community detection, now we have $\sigma_{2K}(\mathbf{B}_E) = O(\frac{n p_n^{3/2}}{\sqrt{K}})$ for the signal part. The order of the noise $\|\mathbf{B} - \mathbf{B}_E\|$ remains $O_p(\sqrt{n p_n})$. Thus for the signal to dominate over noise, we would further require $p_n \gg \sqrt{K}/\sqrt{n}$.

Appendix D. Additional algorithm

Both Algorithms 2 and 3 use combinatorial optimization to resolve the label permutation problem arising from matching the $2K$ clusters from Algorithm 1, which scales exponentially in K. we propose the following heuristic algorithm as a faster alternative to exact combinatorial optimization. With a slight abuse of notation, we use $\widehat{\boldsymbol{\Pi}}^{\mathbf{S}}$ to denote the submatrix of $\widehat{\boldsymbol{\Pi}}$, consisting of row i with $\{i : a_{1i} = 1\}$; similarly $\widehat{\boldsymbol{\Pi}}^{\mathbf{I}-\mathbf{S}}$ consists of row i with $\{i : a_{1i} = 0\}$. $\mathbf{B}^{\mathbf{I}-\mathbf{S},\mathbf{S}}$ is similarly defined as a submatrix of $\mathbf{B}$.

We note that on the population submatrix, $(\mathbb{E}\mathbf{B})^{\mathbf{S},\mathbf{I}-\mathbf{S}} = \boldsymbol{\Pi}^{\mathbf{S}} \mathbf{P} (\boldsymbol{\Pi}^{\mathbf{I}-\mathbf{S}})^T$. Plugging in the estimated $\widehat{\boldsymbol{\Pi}}^{\mathbf{S}}$ (from Algorithm 1) and $\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}}$, and replacing $\mathbb{E}\mathbf{B}$ by $\mathbf{B}$, we can estimate $\boldsymbol{\Pi}^{\mathbf{I}-\mathbf{S}}$ using the least squares equation,

$$\widehat{\boldsymbol{\Pi}}_{\text{reg}}^{\mathbf{I}-\mathbf{S}} = \mathbf{B}^{\mathbf{I}-\mathbf{S},\mathbf{S}} \widehat{\boldsymbol{\Pi}}^{\mathbf{S}} \left((\widehat{\boldsymbol{\Pi}}^{\mathbf{S}})^T \widehat{\boldsymbol{\Pi}}^{\mathbf{S}} \right)^{-1} (\widehat{\mathbf{P}}^{\mathbf{S},\mathbf{S}})^{-1}. \tag{101}$$

Taking $\operatorname{argmax} \widehat{\boldsymbol{\Pi}}_{\text{reg}}^{\mathbf{I}-\mathbf{S}}(i, :)$ for each row i gives one version of estimated labels for the non-neighbor nodes, which are consistent with $\widehat{\boldsymbol{\Pi}}^{\mathbf{S}}$ in terms of group labeling. Recalling that

Algorithm 1 gives another version of estimated labels $\widehat{\boldsymbol{\Pi}}_1^{\mathbf{I-S}}$, we resolve the label permutation between $\widehat{\boldsymbol{\Pi}}_1^{\mathbf{I-S}}$ and $\widehat{\boldsymbol{\Pi}}_{\mathrm{reg}}^{\mathbf{I-S}}$ through a majority vote. Any remaining unresolved labels will go through the same combinatorial optimization as in the original Algorithm 2 (in a significantly reduced space). The details of this heuristic algorithm are presented in Algorithm 4. For DCSBM, the same algorithm can be used as a faster alternative to exact combinatorial optimization for the label matching step.

Empirically, we observe that combining this algorithm with Algorithm 1 gives almost identical accuracy rates on all simulated and real data. Figure 7 compares the running time of Algorithm 1 and Algorithm 4, the two main components of our community detection algorithm, as K increases. For each datapoint, 10 networks are generated under Model 1 with 200 nodes in each block. The running time of Algorithm 4 (the matching step) is negligible compared to Algorithm 1 (the main spectral clustering part).

Finally, we make two remarks about Algorithm 4. (i) Eq (101) itself in general is not a good estimate of $\boldsymbol{\Pi}^{\mathbf{I-S}}$ because, as shown in Lemma 3, $\mathbb{E}\mathbf{B}$ is not the signal term of $\mathbf{B}$. Thus, we only use this estimate to guide the search for the optimal label permutation. (ii) We have used $\widehat{\boldsymbol{\Pi}}^{\mathbf{S}}$ as the "reference" set in Eq (101) instead of $\widehat{\boldsymbol{\Pi}}_1^{\mathbf{I-S}}$, since the neighbor nodes have more observed edges than the non-neighbor nodes and should contain more information.

Algorithm 4 Community detection under the SBM (faster alternative for matching)

Input: $\widehat{\boldsymbol{\Pi}}^{\mathbf{S}}$, $\widehat{\boldsymbol{\Pi}}_1^{\mathbf{I-S}}$ and its corresponding clusters $\{\mathfrak{d}_1, \ldots, \mathfrak{d}_K\}$ (from Algorithm 1); $\widehat{\mathbf{P}}^{\mathbf{S,S}}$ and $\widehat{\mathbf{P}}^{\mathbf{S,I-S}}$ from (17).
Output: optimal permutation $\widehat{f}_0$.
Initialize: $f = \mathbf{0}_{1 \times K}$; $\tau_k^* = 0$, $k \in [K]$.

1: Compute $\widehat{\boldsymbol{\Pi}}_{\mathrm{reg}}^{\mathbf{I-S}}$ defined in (101). Define estimated membership matrix $\widehat{\boldsymbol{\Pi}}_2^{\mathbf{I-S}}$ as
 $\widehat{\boldsymbol{\Pi}}_2^{\mathbf{I-S}}(i, \operatorname{argmax} \widehat{\boldsymbol{\Pi}}_{\mathrm{reg}}^{\mathbf{I-S}}(i,:)) = 1$
2: **for** $k = 1, \ldots, K$ **do**
3:   $\tau_k \leftarrow \frac{\mathbf{1}^T \widehat{\boldsymbol{\Pi}}_2^{\mathbf{I-S}}(i \in \mathfrak{d}_k,:)}{|\mathfrak{d}_k|}$ $\triangleright$ determine label mapping from $\widehat{\boldsymbol{\Pi}}_1^{\mathbf{I-S}}$ to $\widehat{\boldsymbol{\Pi}}_2^{\mathbf{I-S}}$ by majority
  vote
4:   $\tau_k'^* \leftarrow \max \tau_k(:)$
5:   **if** $\tau_k'^* > \tau_k^*$ **then**
6:     $\tau_k^* \leftarrow \tau_k'^*$
7:     $f(k) \leftarrow \operatorname{argmax} \tau_k(:)$
8:   **end if**
9: **end for**
10: **if** number of zero elements in f=0 **then**
11:   return $\widehat{f}_0 \leftarrow f$;
12: **else** $\triangleright$ any unresolved labels go through combinatorial optimization
13:   extract $\mathcal{K} \leftarrow \{k \in [K] : f(k) = 0\}$
14:   for all permutations $\mathcal{S}_{\mathcal{K}}$ of the labels in $\mathcal{K}$, $\widehat{f}_0 \leftarrow \operatorname{argmin}_{f \in \mathcal{S}_{\mathcal{K}}} \|\widehat{\mathbf{P}}_{(f,f)}^{\mathbf{S,S}} - \widehat{\mathbf{P}}_{(f,*)}^{\mathbf{S,I-S}}\|_F$.
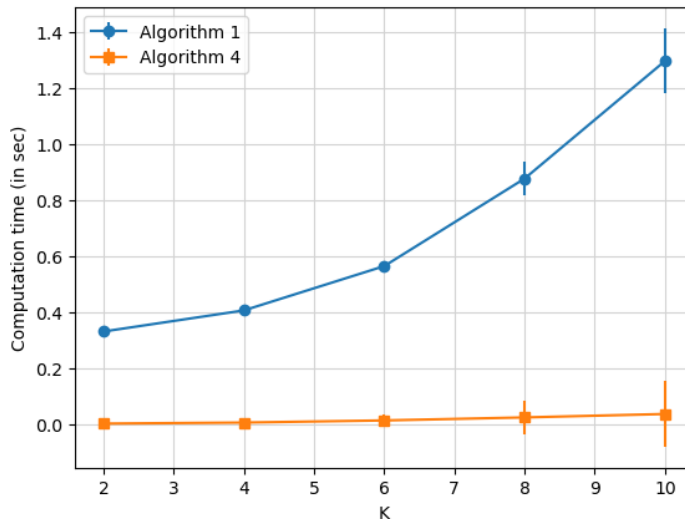15: **end if**

Figure 7: Running time of Algorithm 1 and Algorithm 4 (the two main components of our community detection algorithm) under SBM as K increases.

$q \setminus n$	300	600	900	1200	1500	1800	2100
.1	.3590	.3560	.3587	.3567	.3560	.3575	.3559
$\sqrt{\log n / n}$	.4696	.3645	.3147	.2836	.2613	.2402	.2237
$(\log n / n)^{1/4}/2$	.5890	.5280	.4902	.4693	.4511	.4368	.4266
$1/\sqrt{n}$	.2226	.1580	.1290	.1117	.1002	.0925	.0839

Table 3: The ratio of the edges observed by individual 1 out of the full networks for Model 1, averaged over 100 datasets for each (n, q) combination.

Appendix E. Additional results from simulation

E.1 Additional tables and figures for Model 1

We plot Figure 4a to provide some visualization support of Algorithm 2. This is a scatter plot in which the axes are the two eigenvectors corresponding to the positive eigenvalues of **B**. Recall that we have shown in previous sections that **B** for $K = 2$ has two positive

$q \setminus n$	300	600	900	1200	1500	1800	2100
.1	1	1	1	1	1	1	1
$\sqrt{\log n / n}$	1	1	1	1	1	1	1
$(\log n / n)^{1/4}/2$	1	1	1	1	1	1	1
$1/\sqrt{n}$	.9766	.9732	.9726	.9724	.9726	.9726	.9711

Table 4: The fraction of the individuals within individual 1's knowledge depth for Model 1, averaged over 100 datasets for each (n, q) combination.

eigenvalues and two negative ones. The outlier point (a blue point) close to the vertical axis between .01 and .02 represents individual 1. The blue points are the individuals not adjacent to individual 1, while the red points represent those who are adjacent to 1. Then on the same dataset, we color the points by their true community memberships in Figure 9. Comparing Figure 8 with Figure 9, one can see that it makes sense to develop a strategy to first apply k-means respectively to the two groups of individuals separated by whether they are adjacent to individual 1 and then merge the corresponding clusters across groups.

In addition, to demonstrate the need for our algorithms, we apply spectral clustering to **B** directly and plot the misclustering rates in Figure 10. For all choices of n and q, this approach fails to produce satisfactory clustering accuracy.
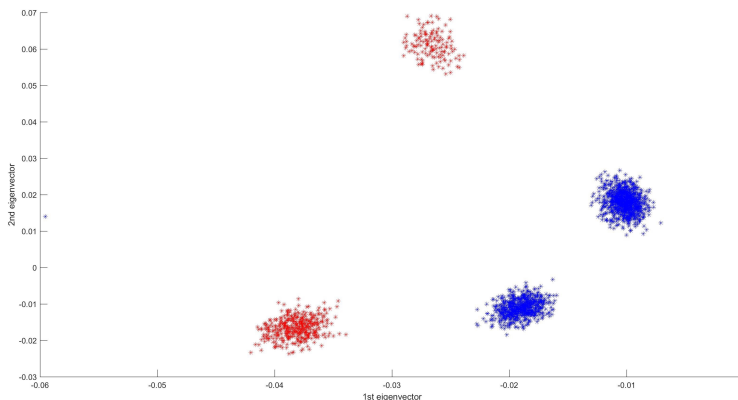

Figure 8: Scatter plot of the two eigenvectors corresponding to the positive eigenvalues of **B** for one dataset from Model 1 when $q = (\log n/n)^{1/4}/2$ and $n = 2100$. The blue points are not adjacent to individual 1, while the red points are adjacent to individual 1.

E.2 Additional figure for Model 2

Figure 11 computes the Pearson and Spearman correlations between five types of centrality measures using a randomly generated network from Model 2. $\widehat{\lambda}_{\min}$ is less correlated with the other four measures than those measures among themselves.

E.3 Additional model for SBM

Model 3 ($K = 3$): $\mathbf{P} = \begin{pmatrix} 3q & 1.5q & q \\ 1.5q & 3q & 1.5q \\ q & 1.5q & 3q \end{pmatrix}$ and each group is of size $n/3$.
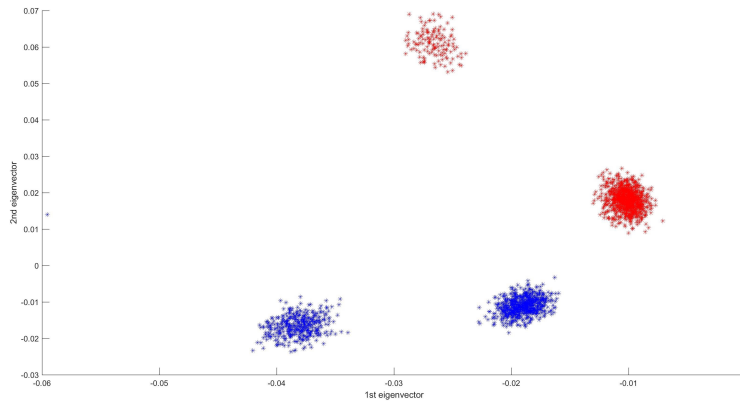
51

Figure 9: Scatter plot of the first two eigenvectors corresponding to positive eigenvalues of $\mathbf{B}$ for one dataset from Model 1 when $q = (\log n/n)^{1/4}/2$ and $n = 2100$. The blue points belong to community 1 and the red points belong to community 2.
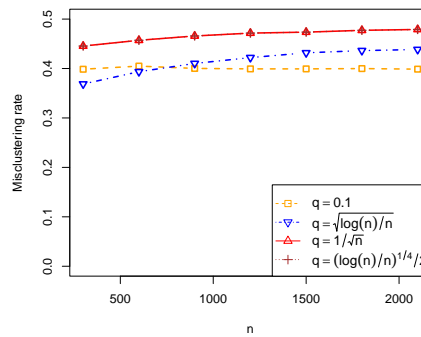

Figure 10: Misclustering rate for Model 1 ($L = 2$) obtained by applying spectral clustering directly to $\mathbf{B}$, averaged over 100 datasets for each combination.
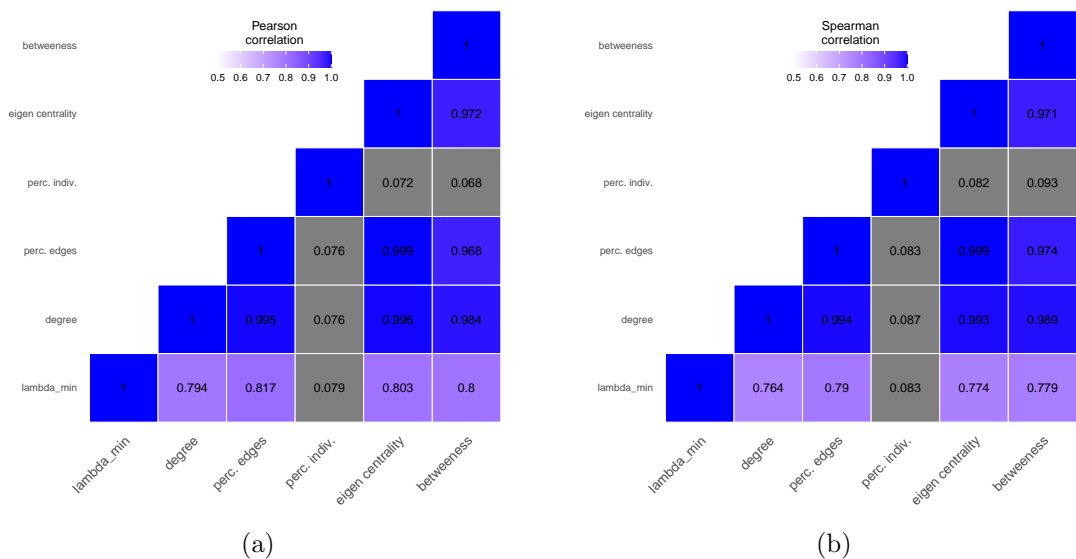
Figure 11: Pearson (a) and Spearman (b) correlations between centrality measures under Model 2.

Similar to their counterparts for Model 1, Tables 5 and 6 show that in Model 3 ($K = 3$), although individual 1 can observe at least one edge of (almost) every other individual in the network, the proportion of total missing edges in her perspective is large. The visualization of Model 3 data in 3-D, similar to that of Model 1 in 2-D, to support Algorithm 2 is attached in Section E.3 of the Appendix. We report the misclustering results in Figure 12. This figure indicates that $K = 3$ is a more challenging situation compared to $K = 2$. Algorithm 2 for $q = 1/\sqrt{n}$ with Model 3 works worse than with Model 1. Also note that $q = \sqrt{\log n/n}$ in Model 1 delivers almost perfect clustering results, but the trend in Figure 12 suggests that even as n goes to infinity, the misclustering rate does not seem to go down to zero. The rate $p_n \sim q = \sqrt{\log n/n}$ is smaller than the rate in the theoretic Condition 5; but at this rate, Algorithm 2 works well for Model 1 while its performance is acceptable for Model 3. As a comparison, we report in Figure 13 the simulation results based on the adjacency matrix $\mathbf{A}$ with the usual spectral clustering algorithm. For larger n in each combination, the misclustering rate is very close to 0, which is theoretically guaranteed by a few works in the literature (c.f. Abbe (2017)). Although we did not work on the boundary condition, we conjecture that the boundary condition under the new partial information framework for almost exact recovery is at least of order $\sqrt{\log n/n}$.

Appendix F. Additional results from real data analysis

F.1 Karate club data

Table 7 shows the community detection accuracy for selected individuals in the karate club network, using Algorithm 2 and plain spectral clustering. The latter in general performs worse than our algorithm.

$q \setminus n$	300	600	900	1200	1500	1800	2100
.1	.3298	.3302	.3309	.3314	.3286	.3308	.3294
$\sqrt{\log n / n}$	.4387	.3407	.2932	.2619	.2361	.2221	.2069
$(\log n/n)^{1/4}/2$	.5531	.4941	.4623	.4413	.4199	.4064	.3948
$1/\sqrt{n}$	.2018	.1441	.1204	.1047	.0930	.0849	.0797

Table 5: The ratio of the edges observed by individual 1 out of the full networks for Model 3, averaged over 100 datasets for each (n, q) combination.

$q \setminus n$	300	600	900	1200	1500	1800	2100
.1	.9999	1	1	1	1	1	1
$\sqrt{\log n / n}$	1	1	1	1	1	1	1
$(\log n/n)^{1/4}/2$	1	1	1	1	1	1	1
$1/\sqrt{n}$	.9653	.9632	.9619	.9642	.9632	.9656	.9648

Table 6: The fraction of the individuals within individual 1's knowledge depth for Model 3, averaged over 100 datasets for each (n, q) combination.
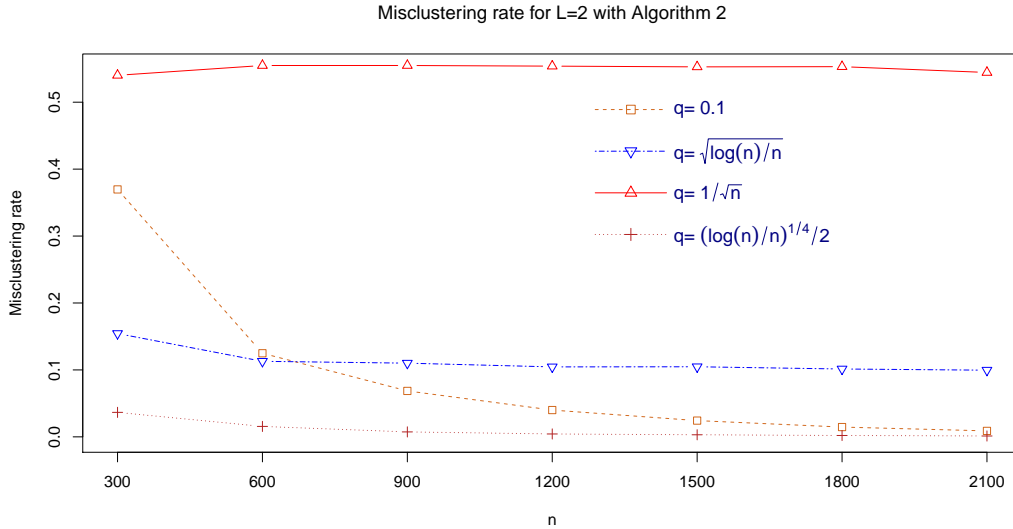

Figure 12: Misclustering rate for Model 3 ($L = 2$), averaged over 100 datasets for each combination.

individual of interest	H	2	3	A	20	32
accuracy (Algorithm 2)	.676	.559	.676	.706	.971	1
accuracy (spectral clustering)	.706	.529	.529	.529	.559	.500

Table 7: The network information and detection results for chosen individuals using Algorithm 2 and plain spectral clustering.

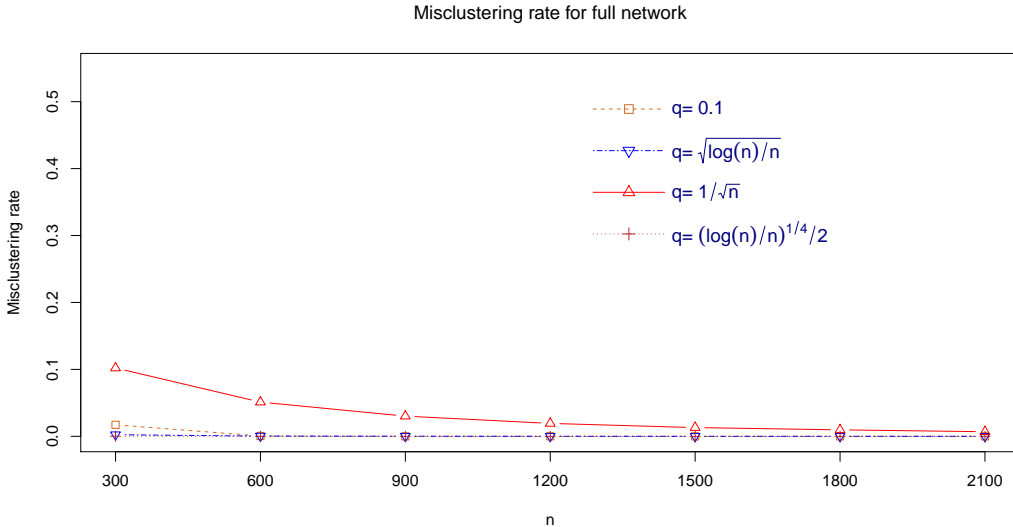Misclustering rate for full network

Figure 13: Misclustering rate for Model 3 (full network), averaged over 100 datasets for each combination.

F.2 Indian villages

Figure 14 shows two examples of social networks from the dataset, with nodes (households) colored by their caste.

Data preprocessing. For each of the 43 villages, the dataset contains households sampled for individual surveys, which contain their meta information such as caste, religion, language and occupation. Within each village, we first remove households whose caste information is "NA" or inconsistent among the surveyed household members (both constituting only a small number of the total). To avoid imbalanced cluster sizes, in each village we also remove castes with fewer than 10 households. After filtering out these households, we include villages with $K \geq 2$. The process results in 39 villages to be included in the analysis.

Both the eigenvector and betweenness centrality are normalized to make them comparable across different villages.

F.3 Political blogs

node of interest	1073	1074	1075	1076	1077	1078
the ratio of the edges observed	.1145	.1295	.3362	.1553	.2761	.4116
# of the nodes observed	476	485	880	715	808	793
clustering accuracy	.5819	.5505	.9090	.9035	.9331	.8601

Table 8: The network information and detection results for individuals (blogs) in the political blog data using Algorithm 2.
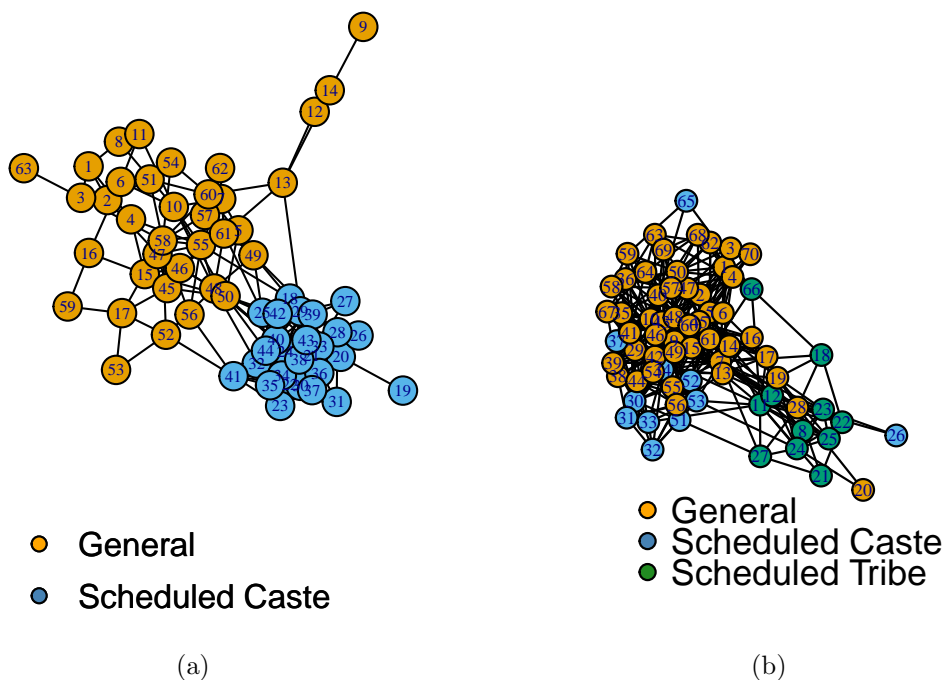
Figure 14: Social networks in villages 2 (a) and 49 (b), colored by caste information.

References

E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *https://arxiv.org/pdf/1709.09565.pdf*, 2017.

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, pages 36–43, 2005.

E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

Animashree Anandkumar, Rong Ge, Daniel Hsu, and Sham M Kakade. A tensor approach to learning mixed membership community models. *Journal of Machine Learning Research*, 15(1):2239–2312, 2014.

Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.

Sharmodeep Bhattacharyya and Peter J Bickel. Subsampling bootstrap of count features of networks. *The Annals of Statistics*, 43(6):2384–2411, 2015.
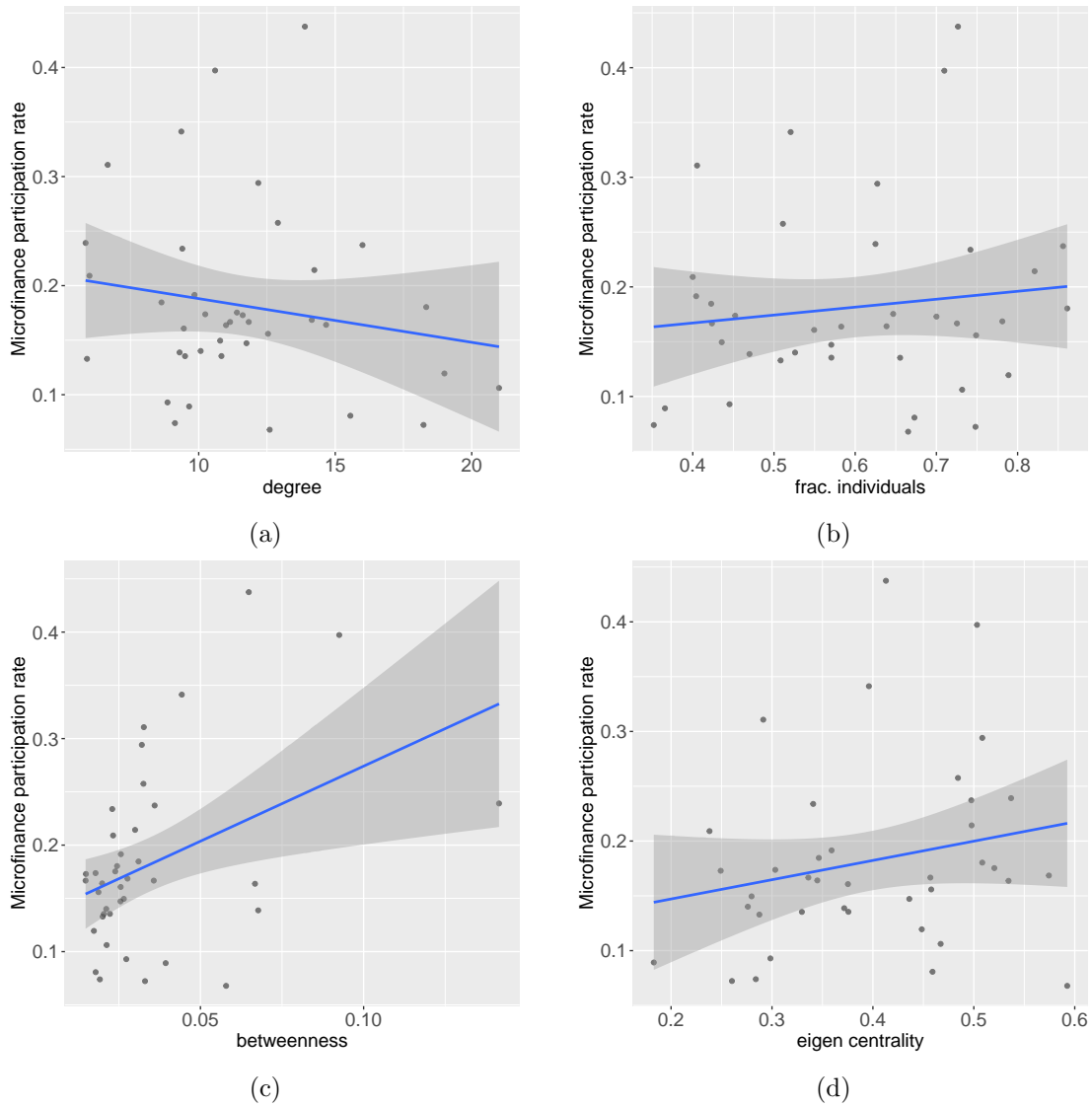
Figure 15: Program participation rate as a function of (a) node degree; (b) fraction of individuals reached by the partial network; (c) betweenness centrality; (d) eigenvector centrality. The centrality measures are calculated for leaders in each village and an average is taken across each village.

P. J. Bickel and P. Sarkar. Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B*, 78:253–273, 2016.

Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50): 21068–21073, 2009.

T Tony Cai, Zongming Ma, Yihong Wu, et al. Sparse pca: Optimal rates and adaptive estimation. *The Annals of Statistics*, 41(6):3074–3110, 2013.

T Tony Cai, Xiao Han, Guangming Pan, et al. Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Annals of Statistics*, 48 (3):1255–1280, 2020.

Chen Cheng, Wei Huang, and Yiqing Xing. A theory of multiplexity: Sustaining cooperation with multiple relations. *Available at SSRN: https://ssrn.com/abstract=3811181 or http://dx.doi.org/10.2139/ssrn.3811181*, 2021.

Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Jianqing Fan, Yingying Fan, Xiao Han, and Jinchi Lv. Simple: Statistical inference on membership profiles in large networks. *https://arxiv.org/pdf/1910.01734.pdf*, 2019a.

Jianqing Fan, Jianhua Guo, and Shurong Zheng. Estimating number of factors by adjusted eigenvalues thresholding. *https://arxiv.org/pdf/1909.10710.pdf*, 2019b.

Linton C Freeman. Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291–304, 1982.

Leo A Goodman. Snowball sampling. *The annals of mathematical statistics*, pages 148–170, 1961.

Xiao Han, Qing Yang, and Yingying Fan. Universal rank inference via residual subsampling with application to large networks. *The Annals of Statistics*, 51(3):1109–1133, 2023.

Mark S Handcock and Krista J Gile. Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1):5, 2010.

Douglas D Heckathorn. Respondent-driven sampling: a new approach to the study of hidden populations. *Social problems*, 44(2):174–199, 1997.

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

J. Jin. Fast community detection by SCORE. *The Annals of Statistics.*, 43:57–89, 2015.

Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Estimating network memberships by simplex vertex hunting. *https://arxiv.org/pdf/1708.07852.pdf*, 2017.

B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83:016107, 2011.

A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66:1663–1749, 2013.

J. Lei. A goodness-of-fit test for stochastic block models. *The Annals of Statistics.*, 44: 401–424, 2016.

Jing Lei and Alessandro Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015. ISSN 00905364. URL `http://www.jstor.org/stable/43556513`.

Jing Lei, Alessandro Rinaldo, et al. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

Soumendu Sundar Mukherjee, Purnamrita Sarkar, and Peter J. Bickel. Two provably consistent divide-and-conquer clustering algorithms for large networks. *Proceedings of the National Academy of Sciences*, 118(44):e2100482118, 2021.

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39:1878–1915, 2011.

Karl Rohe. A critical threshold for design effects in network sampling. *The Annals of Statistics*, 47(1):556–582, 2019.

D Franco Saldana, Yi Yu, and Yang Feng. How many communities are there? *Journal of Computational and Graphical Statistics*, 26(1):171–181, 2017.

Matthew J. Salganik and Douglas D. Heckathorn. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology*, 34(1):193–240, 2004. doi: 10.1111/j.0081-1750.2004.00152.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0081-1750.2004.00152.x`.

Liangjun Su, Wuyi Wang, and Yichong Zhang. Strong consistency of spectral clustering for stochastic block models. *IEEE Transactions on Information Theory*, 66(1):324–338, 2019.

J. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12:389–434, 2012.

YX Rachel Wang, Peter J Bickel, et al. Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2):500–528, 2017.

Stanley Wasserman, Katherine Faust, et al. Social network analysis: Methods and applications. 1994.

Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

Yusheng Wu. *Thesis dissertation*. University of Southern California, 2021.

Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Detecting overlapping communities in networks using spectral methods. *SIAM Journal on Mathematics of Data Science*, 2(2):265–283, 2020.

Yunpeng Zhao, Elizaveta Levina, and Ji Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292, 2012. doi: 10.1214/12-AOS1036. URL `https://doi.org/10.1214/12-AOS1036`.