

# A Doubled Adjacency Spectral Embedding Approach to Graph Clustering

Sinyoung Park

Department of Mathematical Sciences  
University of Bath  
sp2681@bath.ac.uk

Matthew A. Nunes

Department of Mathematical Sciences  
University of Bath  
man54@bath.ac.uk

Sandipan Roy

Department of Mathematical Sciences  
University of Bath  
sr2081@bath.ac.uk

December 15, 2025

## Abstract

Spectral clustering is a popular tool in network data analysis, with applications in a variety of scientific application areas. However, many studies have shown that spectral clustering does not perform well on certain network structures, particularly core-periphery networks. To improve clustering performance in core-periphery structures, Adjacency Spectral Embedding (ASE) has been introduced, which performs clustering via a network's adjacency matrix instead of the graph Laplacian. Despite its advantages in this setting, the optimal performance of ASE is limited to dense networks, whilst network data observed in practice is often sparse in nature. To address this limitation, we propose a new approach which we term Doubled Adjacency Spectral Embedding (DASE), motivated by the observation that the squared adjacency matrix will leverage the fewer connections in sparse structures more efficiently in clustering applications. Theoretical results establish that DASE enjoys good consistency properties when determining sparse community structure. The performance and general applicability of the proposed method is evaluated using extensive simulations on both directed and undirected networks. Our results highlight the improved clustering performance on both sparse and dense networks in the presence of core-periphery structures. We illustrate our proposed technique on real-world employment and transportation datasets.

**Keywords:** Adjacency Spectral Embedding; clustering; community detection; core-periphery networks.

## 1 Introduction

### 1.1 Background

Network data are ubiquitous in numerous disciplines, including politics (Adamic et al. 2005), neuroscience (Rubinov et al. 2010), biology (Pavlopoulos et al. 2011), power systems (Pagani et al. 2013), ecology (Wey et al. 2008), and transportation (Derudder et al. 2007). These structures provide a natural framework for analysing, interpreting, and predicting complex relationships within these domains. For example, in social network analysis, individuals are typically represented as nodes, while their interactions are modelled as directed or undirected edges. Such graph-based representations facilitate the identification of patterns and emergent behaviours, thereby enabling rigorous analysis and deeper interpretation of underlying dependence relationship. A critical aspect of network analysis lies in understanding their intrinsic structure. Networks can exhibit various structural patterns, including assortative, non-assortative, scale-free, and core–periphery configurations.

One of the most fundamental tasks in analysing networks is to find clusters or communities that group similar objects/individuals together. The mathematical approaches for identifying such groups is commonly

known in the literature as clustering or community detection techniques. Identifying communities within a network enables deeper understanding of local structures (dyads, triads etc.) as well as a holistic (global) understanding of the entire network.

Community detection has been extensively studied in the statistical learning literature, see e.g. Fjällström 1998; Fortunato 2010; Goldenberg et al. 2010 and the references therein for overviews of the literature. Several approaches have been proposed to identify communities in networks (Newman et al. 2004; Fortunato 2010; Abbe 2018), including spectral embedding (Von Luxburg 2007; Rohe et al. 2011; Rubin-Delanchy et al. 2022). Spectral clustering (SC) and its variants have developed over recent years as some of the most popular techniques for finding communities in network data, and their use is now commonplace in a number of settings, see e.g., Shi et al. 2000; Ng et al. 2001; Von Luxburg 2007. These methods predominantly perform spectral embedding using the Laplacian matrix – a representation of the connections between individuals in network data. Most of this work has been dedicated to undirected graphs, but there has been a growing body of literature on community detection in directed networks (Satuluri et al. 2011; Cucuringu et al. 2020; Wang et al. 2020).

Recent work has shown that Laplacian-based spectral embedding tends to perform well in assortative or non-assortative networks, but not in other settings such as core-periphery structures (Priebe et al. 2019); these networks naturally have two distinct regions: the ‘core’ within which nodes are densely interconnected, and the ‘periphery’ which exhibits sparse internal connections (Borgatti et al. 2000). Such structures appear in a number of application areas and data settings, such as co-citation networks, social interaction networks and economics (Tang et al. 2019), and have been studied more intensively in recent years, predominantly related to the identification of core structures (see e.g. Csermely et al. 2013; Rombach et al. 2014; Rombach et al. 2017; Zhang et al. 2015). To address the unsatisfactory performance of Laplacian-based spectral clustering in this setting, adjacency spectral embedding (ASE) has been proposed (Sussman et al. 2012). In particular, it has been shown that ASE provides a more meaningful interpretation of the underlying community structure than traditional spectral clustering for core-periphery networks, under a general stochastic block modelling framework (Priebe et al. 2019; Tang et al. 2018); however, ASE can struggle to detect structures when the underlying network is sparse. For the specific case of *directed* networks, recent research has proposed statistical inference methods that can detect core-periphery structures in large networks (Priebe et al. 2019; Zhang et al. 2015), and directed core-periphery structures with L-shapes (Elliott et al. 2020).

In this article we develop an alternative approach to clustering nodes suitable for both undirected and directed core-periphery network structures, based on spectral embedding of the *squared* adjacency matrix. we which term *Double Adjacency Spectral Embedding* (DASE). Our approach has similarities to the work the work of Rohe et al. 2011, who show that the spectral clustering based on the squared Laplacian achieves good consistency and misclustering properties under the stochastic block model. However our work differs in that our focus is on core-periphery networks, for which adjacency embedding is more appropriate. Combined with a clustering algorithm such as k-means, our proposed clustering algorithm DASE-CLUST provides consistent clustering of nodes in a network under the general stochastic block model, and is particularly suitable for core-periphery structures, whether directed or undirected. Our proposed approach shows improved recovery of the true community structure compared to the ASE technique of Sussman et al. 2012, particularly for challenging setups. One can show that DASE can be thought as a variant of an embedding under the weighted stochastic block model (Gallagher et al. 2024), and that DASE provides more informative embeddings than ASE in certain settings as measured by the Chernoff information (Chernoff 1952), commonly used to assess embeddings (Tang et al. 2018; Gallagher et al. 2024).

This article is structured as follows. In Section 2, we introduce the stochastic block model (SBM), upon which we build our proposed methodology, and define core-periphery networks which we focus on in this work. Our proposed doubled adjacency spectral embedding (DASE) approach is described in Section 3, whilst Section 4 establishes theoretical properties related to DASE as well as the subsequent clustering algorithm, both in the general and core-periphery network cases. We demonstrate the efficacy of the DASE embedding and our proposed clustering procedure through extensive simulation studies in Section 5, considering both directed and undirected networks. Finally, in Section 6, we consider the application of our proposed clustering methodology to two real-world datasets, namely faculty hiring data and flights observed on an airport network. In Section 7, we make some concluding remarks.

## 2 Network Model

### 2.1 Notations

We use  $\langle a, b \rangle$  to denote the inner product between two vectors  $a \in \mathbb{R}^m$  and  $b \in \mathbb{R}^n$  in the Euclidean space  $\mathbb{R}^{m \times n}$ . We use  $\text{rank}(A)$  to denote the rank of a matrix  $A$ . We use  $\|\cdot\|$  to denote the Frobenius norm of a vector or matrix. The cardinality of a set  $A$  is denoted as  $|A|$ . We denote  $\text{diag}(a_1, a_2, \dots, a_n) \in \mathbb{R}^{n \times n}$  as a diagonal matrix.

### 2.2 Set up

Let  $G = (V, E)$  denote a graph which represents a network. It has the vertex set  $V = \{1, 2, \dots, N\} = [N]$  and edge set  $E = \{(i, j) : i \in V, j \in V, i \neq j\}$ .  $N$  is the number of nodes in the network and the number of edges is given by  $|E|$ . SBM is one of the most popular probabilistic generative models for networks (Holland et al. 1983). It has an inherent block structure, where block membership is determined solely by the latent vector. We start with the following definition of SBM.

**Definition 2.1** (Stochastic Block Model (SBM( $B, \pi$ )) (Holland et al. 1983)). *Consider nodes  $i \in \{1, \dots, N\}$ , where  $\theta_i \in \{1, \dots, K\}$  denotes community membership, with  $K$  being the number of blocks. Given a block probability matrix  $B \in [0, 1]^{K \times K}$  and a parameter vector  $\pi \in (0, 1)^K$ , conditioned on  $\theta_i, \theta_j$  the adjacency matrix  $A$  is generated as*

$$A_{ij} | \theta_i, \theta_j \stackrel{\text{ind}}{\sim} \text{Bernoulli}(B_{\theta_i \theta_j}).$$

For undirected graphs,  $A_{ij}$  is generated only for  $i < j$  with the symmetric block matrix  $B$ . For directed graphs, all ordered pairs  $(i, j)$  are generated independently.

Here, we assume that the graph has no self-loop in both the undirected and directed graphs, i.e.,  $A_{ii} = 0$  for all  $i \in \{1, \dots, N\}$ . For directed graphs,  $A_{ij} = 1$  indicates an edge from node  $i$  to node  $j$ , whereas for undirected graphs,  $A_{ij}$  simply indicates that nodes  $i$  and  $j$  are connected. We can write

$$\begin{aligned} \mathbb{P}(A|\theta) &= \prod_{i \neq j} \mathbb{P}(A_{ij} | \theta_i, \theta_j) \\ &= \prod_{i \neq j} (B_{\theta_i \theta_j})^{A_{ij}} (1 - B_{\theta_i \theta_j})^{1 - A_{ij}} \end{aligned} \tag{1}$$

We define the node-wise probability matrix  $Q \in [0, 1]^{N \times N}$  as  $Q = ZBZ^\top$  for both undirected and directed adjacency matrices, where  $Z \in \{0, 1\}^{N \times K}$  is the membership indicator matrix. If  $Z_{ij} = 1$ , then the node  $i$  belongs to the  $j$ -th group. Let  $\pi \in (0, 1)^K$  denote the vector of group proportions, with the entries summing to 1. For example, when  $K = 2$ , the case  $\pi = (0.5, 0.5)^\top$  represents a balanced network.

We now present the Random Dot Product Graph (RDPG) model to explain our graph embedding. Further, the following definition also provides another parametrisation for SBM.

**Definition 2.2** (Random Dot Product Graphs (Young et al. 2007)). *Let  $X, Y \in \mathbb{R}^{N \times d}$ , where  $X_i \in \mathbb{R}^d$  and  $Y_i \in \mathbb{R}^d$  represent the latent position vectors, for  $i \in [1, \dots, N]$ , with latent space dimension  $d$ . The latent position matrices  $X$  and  $Y$  are random and hold*

$$\mathbb{P}(\langle X_i, Y_j \rangle \in [0, 1]) = 1, \quad \text{for all } i, j \in \{0, \dots, N\}.$$

Then, the entries of the adjacency matrix  $A_{ij}$  are independent Bernoulli random variables conditioned on  $X_i$  and  $Y_i$  with parameter  $\langle X_i, Y_j \rangle$  for all  $i \neq j$ .

$$\mathbb{P}(A|X, Y) = \prod_{i \neq j} \mathbb{P}(A_{ij} | X_i, Y_j) \tag{2}$$

For the undirected graphs  $A$ , we have

$$\mathbb{P}(A|X, Y) = \prod_{i \neq j} \mathbb{P}(A_{ij} | X_i, X_j) \tag{3}$$

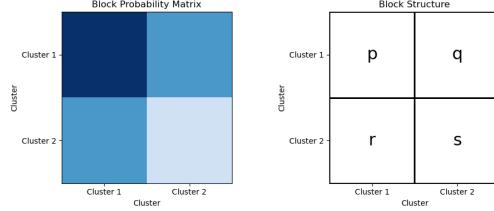


Figure 1: Connection Probability Heatmap for Core-periphery Graphs

Based on Definition 2.2, the Stochastic Block Model can be viewed as a special case of the Random Dot Product Graph (Athreya et al. 2018; Sussman et al. 2012). In particular, for directed graphs,

$$\mathbb{P}(A_{ij} = 1) = B_{\theta_i \theta_j} = \langle X_i, Y_j \rangle, \quad (4)$$

where the latent dimension  $d$  equals the rank of the block probability matrix  $B$ . In general, the number of communities are equal or greater than the latent dimension, i.e.,  $K \geq d$ .

In this paper, we focus on core-periphery structures. Several intuitive interpretations of core-periphery graphs have been summarised in (Borgatti et al. 2000; Rombach et al. 2014). A common idea in previous research is that the structure using two distinct groups, core and periphery groups, and we adopt this idea as our primary model:

**Definition 2.3** (Core-periphery graphs). *Core-periphery graphs consist of two parts: core groups and periphery groups. Connections within the core group are dense. Nodes in the core group have relatively dense connections with nodes in the periphery. In contrast, connections within the periphery group are sparse, with only a few links among the peripheral nodes.*

While Definition 2.3 focuses on a two-cluster ( $K = 2$ ) block structure, recent research has broaden this idea. For instance, (Rombach et al. 2014) introduced models with multiple core groups and (Gallagher et al. 2021) proposed a hub-and-spoke model that represents a layered structure, including an intermediate layer between core and periphery groups.

These developments highlight a fundamental challenge in defining core-periphery graphs: for  $K > 2$ , existing approaches either generalise the block structure or adopt a layered structure. For this reason, generalizing a single, consistent framework that encompasses both structural definitions for  $K > 2$  is difficult and complex. We therefore restrict our focus in this paper to the established core-periphery graphs with  $K = 2$  as defined in Definition 2.3.

For our setting, we consider an adjacency matrix  $A$  generated from an SBM with two communities where the probability matrix  $B$  is a  $2 \times 2$  matrix. The within-core and within-periphery connection probabilities are  $p$  and  $s$ , respectively, while the between group connection probabilities are  $q$  and  $r$ . We assume the ordering  $s < r, q < p$ , which implies that Cluster 1 has the highest within-cluster connectivity and thus represents the core group. Figure 1 describes the corresponding block probability structure.

### 2.3 Measuring Performance of Clustering

As discussed in detail in (Sussman et al. 2012), for a graph distributed according to the model defined in 2.1, we can use the following clustering criterion on the DASE to accurately assign nodes to blocks.

We use the following mean square error criterion for clustering the rows of  $Z$  into  $K$  blocks,

$$(\hat{\nu}, \hat{\tau}) = \arg \min_{\nu, \tau} \sum_{u=1}^N \|Z_u - \nu_{\tau(u)}\|_2^2, \quad (5)$$

where  $\hat{\nu}_i \in \mathbb{R}^K$  gives the centroid of block  $i$  and  $\tau : [N] \rightarrow [K]$  is the block assignment function.

### 3 Methodology

#### 3.1 Doubled Adjacency Spectral Embedding (DASE)

In the following, we introduce our square transformed adjacency matrix which we would use to perform our clustering to detect core and periphery communities.

**Definition 3.1** (Doubled adjacency matrix ( $\tilde{A}$ )). *Given an adjacency matrix  $A \in \{0, 1\}^{N \times N}$ , the doubled adjacency matrix  $\tilde{A} \in [0, N]^{N \times N}$  is defined as*

$$\tilde{A} := [AA]_{ij} = \sum_{k=1}^N A_{ik} A_{kj}. \quad (6)$$

The doubled adjacency matrix  $\tilde{A}$ , which is simply the square of the adjacency matrix, the number of two steps walk that one can take from one node to another for both directed and undirected graphs. For example, since  $A$  indicates the connections between nodes,  $(\tilde{A})_{12}$  indicates that the number of ways to travel from node 1 to node 2 in two steps.

Each entry of  $\tilde{A}$  is the sum of product of two Bernoulli random variables, and therefore  $\tilde{A}$  does not follow a Bernoulli distribution. Instead, a sum of independent Bernoulli variables follows a Poisson-Binomial (PB) distribution (Tang et al. 2023). Hence, the entries of the doubled adjacency matrix follow a PB distribution.

If  $A$  is a directed graph,  $A_{is}A_{sj}$  and  $A_{it}A_{tj}$ , are conditionally independent for  $s \neq t$  given any  $i$  and  $j$ , since the entries of  $A$  are conditionally independent Bernoulli random variables with parameters  $Q_{ij} = \mathbb{E}(A_{ij}) = (XY^\top)_{ij}$ . Therefore,

$$\tilde{A}_{ij} \sim \text{PB}(Q_{i1}Q_{1j}, \dots, Q_{iN}Q_{Nj}),$$

If  $A$  is an undirected graph, we separate two cases: diagonal entries and off-diagonal entries. For the diagonal entries ( $i = j$ ), then

$$\mathbb{E}(\tilde{A}_{ii}) = \tilde{Q}_{ii} = \sum_{k=1}^N Q_{ik}.$$

For off-diagonal entries ( $i \neq j$ ), then

$$\mathbb{E}(\tilde{A}_{ij}) = \tilde{Q}_{ij} = \sum_{k=1}^N Q_{ik}Q_{kj}.$$

Thus, for the undirected graphs,

$$\tilde{A}_{ij} \sim \begin{cases} \text{PB}(Q_{i1}Q_{1j}, \dots, Q_{iN}Q_{Nj}) & \text{if } i = j, \\ \text{PB}(Q_{i1}, \dots, Q_{iN}) & \text{if } i \neq j. \end{cases}$$

Beyond the interpretation of  $\tilde{A}$  as counting two-step walks, the doubled adjacency matrix can be also viewed as a Weighted Stochastic Block Models (WSBM) (Gallagher et al. 2024). Unlike a binary adjacency matrix,  $\tilde{A}$  reflects the strength of connections between nodes, and the weights follows a Poisson-Binomial distribution.

#### 3.2 Proposed Algorithm

The normalised doubled adjacency matrix ( $AA/N$ ) and the unnormalised doubled adjacency matrix ( $AA$ ) yield equivalent spectral embeddings, as normalisation simply rescales the singular values and therefore affects only the variance. Consequently, without loss of generality, we use the unnormalised version, which corresponds to a weighted SBM with Poisson–Binomial edge weights.

Our key idea in the proposed algorithm is to compute a spectral embedding of the unnormalised doubled adjacency matrix. We show in Section 4 that the spectral embedding of our unnormalised doubled adjacency

matrix provides a tighter concentration bound than unnormalised adjacency matrix in terms of Frobenius norm. We then provide a consistent cluster assignment via  $k$ -means (MacQueen 1967; Lloyd 1982) or Gaussian mixture model (McLachlan et al. 1988; Dempster et al. 1977) applied to the embedding. We describe the steps of the proposed algorithm (DASE-CLUST) below.

---

**Algorithm 1** Doubled Adjacency Spectral Embedding (DASE)-CLUST

---

**Input:** Adjacency matrix  $A \in \{0, 1\}^{N \times N}$ , number of clusters  $K \in \mathbb{N}$  and latent dimension  $d \in \mathbb{N}$   
**Output:** Clustering labels  $l \in \mathbb{R}^N$

- 1: Compute the doubled adjacency matrix  $\tilde{A}$
  - 2: Compute the singular value decomposition (SVD) of  $\tilde{A}$ , i.e.,  $\tilde{A} = \tilde{U}\tilde{\Sigma}\tilde{V}^T$ , where the singular values in  $\tilde{\Sigma}$  are ordered in decreasing magnitude.
  - 3: **if**  $A$  is directed **then**
  - 4:   Define the embedding  $S = [\tilde{U}_d\tilde{\Sigma}_d^{1/2}|\tilde{V}_d\tilde{\Sigma}_d^{1/2}] \in \mathbb{R}^{N \times 2d}$ , where  $\tilde{U}_d$  and  $\tilde{V}_d$  contain the top  $d$  left and right singular vectors, respectively.  $\tilde{\Sigma}_d$  contains the corresponding singular values.
  - 5: **else**
  - 6:   Define the embedding  $S = \tilde{U}_d\tilde{\Sigma}_d^{1/2} \in \mathbb{R}^{N \times d}$ , where  $\tilde{U}_d$  contains the top  $d$  eigenvectors corresponding to the largest  $d$  singular values in  $\tilde{\Sigma}_d$ .
  - 7: Extract the clustering labels  $l$  from either  $k$ -means or a Gaussian mixture model (with  $K$  clusters) applied to the rows of  $S$ .
- 

As the fourth (or the sixth) step of Algorithm 1, we need to compute the  $d$  largest singular values and their corresponding singular vectors. However, calculating all  $N$  singular values and vectors is computationally expensive. To improve the computational efficiency, we can directly compute only the  $d$  largest singular values and vectors using svds from the `scipy.sparse.linalg` package (Virtanen et al. 2020) in Python.

## 4 Main Theoretical Results

### 4.1 Notations

We use the following notation for the remainder of this paper. Let  $d = \text{rank}(B)$ , so that  $d \leq K$ .  $Q = \mathbb{E}(A) = ZBZ^\top$ , where block membership indicator matrix  $Z \in \{0, 1\}^{N \times K}$ .  $Z^\top Z = D_n = \text{diag}(n_1, \dots, n_d) \in \mathbb{R}^{K \times K}$ , where  $n_i$  is the size of group  $i$ . We denote  $\tilde{B} = BD_nB$ . Let the singular value decomposition of  $\tilde{B}$  be

$$\tilde{B} = \tilde{L}\tilde{\Lambda}\tilde{R}^\top,$$

where  $\tilde{L}, \tilde{R} \in \mathbb{R}^{K \times d}$  and  $\tilde{\Lambda} \in \mathbb{R}^{d \times d}$  is diagonal. We assume  $\lambda_d(\tilde{B}) = \tilde{b}N$  for some constant  $\tilde{b} > 0$ . Suppose  $\tilde{\pi} = \min_{i \in \{1, \dots, K\}} \pi_i$ , where  $\pi_i = n_i/N \in (0, 1)$  denotes the proportion of group  $i$ . Define  $\tau^{(N)} : [N] \rightarrow [K]$  to be the random block membership function and  $n_i = \#\{u : \tau(u) = i\}$  be the size of block  $i$ .

For ASE comparison, let the singular value decomposition of  $B$  be

$$B = L\Lambda R^T,$$

where  $L, R \in \mathbb{R}^{K \times d}$  and  $\Lambda \in \mathbb{R}^{d \times d}$  is diagonal. We assume  $\lambda_d(B) = bN$  for some constant  $b > 0$ . Here,  $\tilde{\beta}, \tilde{b}, \beta, b$  and  $\tilde{\pi}$  are not dependent on  $N$ .

The recovery of cluster memberships for each node is dependent on the quality of the embedding  $S$  in 1. If the embedding encodes the distribution of the nodes in different clusters distinctively it renders the clustering algorithm to be consistent in finding the communities.

The main contribution of this paper is the following consistency result in terms of the estimation of the community membership for each node based on the block assignment function which assigns blocks based on  $S$ . We present consistency results for both general directed networks and directed core-periphery networks. As far as our knowledge, this is the first work which provides a consistent community detection algorithm which has a faster rate of convergence compared to existing works for general networks and first novel theoretical result of rate of mis-classification in core-periphery networks albeit with  $K = 2$ .

**A1.** We assume a sequence of random adjacency matrices  $A^{(N)}$  with node set  $[N]$  for  $N \in \{1, 2, \dots\}$ . The edges are distributed according to a stochastic block model (see 2.1) with parameters  $B$  and  $\pi$ .  $B \in [0, 1]^{K \times K}$  and let  $\pi \in (0, 1)^K$  be a vector with positive entries summing to unity.

**A2.** Assume  $\text{rank}(B) = d$  and  $K \geq d$ .

**A3.** Suppose  $\tilde{B} = \tilde{\nu}\tilde{\mu}^\top$ , where  $\tilde{\nu}, \tilde{\mu} \in \mathbb{R}^{K \times d}$ . Define  $\tilde{\beta} > 0$  such that

$$\tilde{\beta} < \|\tilde{\nu}_u - \tilde{\nu}_v\| \quad \text{or} \quad \tilde{\beta} < \|\tilde{\mu}_u - \tilde{\mu}_v\| \quad \text{for all } u \neq v.$$

Moreover,  $\tilde{\beta} = \hat{\beta}N$  for some constant  $\hat{\beta} > 0$ .

## 4.2 Clustering Consistency

In this section, we establish consistency results for both general networks and core-periphery networks. For general networks, we primarily focus on the directed setting.

**Proposition 1.** Let  $A \in \{0, 1\}^{N \times N}$  be an adjacency matrix of a directed graph. Let  $\tilde{A} = AA \in [0, N]^{N \times N}$  denote the directed doubled adjacency matrix. For fixed  $u, v \in \{1, \dots, N\}$ ,  $\{\tilde{A}_{uw}\tilde{A}_{vw}\}_{w=1}^N$  are conditionally independent given  $\mathcal{F} = \sigma(A_{u*}, A_{v*})$ , where  $\sigma(A_{u*}, A_{v*})$  is the information contained in the  $u$ -th and  $v$ -th rows of the matrix  $A$ . Then the following holds with probability at least  $1 - 2N^{-2}$ :

$$\|\tilde{A}\tilde{A}^\top - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F \leq \sqrt{2N^7 \log N} \quad (7)$$

and

$$\|\tilde{A}^\top\tilde{A} - \mathbb{E}(\tilde{A}^\top\tilde{A})\|_F \leq \sqrt{2N^7 \log N}. \quad (8)$$

Based on the Proposition 1, we have the misclustering rate of DASE when the graph is directed as follows.

**Theorem 4.1.** Under the directed condition in Section 4.1 and Assumptions 1, 2 and 3, suppose that the number of blocks  $K$  and the latent vector dimension  $d$  are known. Let  $\hat{\tau}^{(N)} : V \mapsto \{1, \dots, K\}$  be the estimated block assignment obtained by clustering the rows of  $\tilde{W}$  where  $\tilde{W} = [\tilde{U}|\tilde{V}]$ , satisfying Equation (5). Let  $H_K$  denote the set of permutations of  $[K]$ . Then, with probability at least  $1 - 2N^{-2}$ , we have

$$\min_{\rho \in H_K} |\{u \in V : \tau(u) \neq \rho(\hat{\tau}(u))\}| \leq \frac{2^5 3^2}{\tilde{\beta}^2 (\tilde{b}\tilde{\pi})^5} \cdot \frac{\log N}{N}.$$

The proofs of Proposition 1 and Theorem 4.1 are provided in Appendix A.2. For undirected graph, we have misclustering rate as follows:

**Corollary 4.1.1.** Under the undirected condition in Section 4.1 and Assumptions 1, 2 and 3, suppose that the number of blocks  $K$  and the latent vector dimension  $d$  are known. Let  $\hat{\tau}^{(N)} : V \mapsto \{1, \dots, K\}$  be the estimated block assignment obtained by clustering the rows of  $\tilde{W}$  where  $\tilde{W} = \tilde{U}\tilde{\Sigma}^{1/2}$ , satisfying Equation (5). Let  $H_K$  denote the set of permutations of  $[K]$ . With probability at least  $1 - 2N^{-2}$ , we have

$$\min_{\rho \in S_K} |\{u \in V : \tau(u) \neq \rho(\hat{\tau}(u))\}| \leq \frac{2^4 3^2}{\hat{\beta}^2 (\tilde{b}\tilde{\pi})^5} \cdot \frac{\log N}{N}.$$

*Proof.* Since  $A$  is undirected, the singular value decomposition of  $Q$  takes the form  $Q = U\Sigma U^\top$ . Therefore, by Lemma 3 in Appendix A.2, the misclustering rate improves by a factor of 2.  $\square$

In this paper, we focus on the core-periphery structure. To derive the misclustering rate on the core-periphery graphs, we modify Assumption 2 in the following way.

**A4.** Assume that  $\text{rank}(B) = 2$  and that  $B$  represents connection probability matrix in a core-periphery network. The within-core and within-periphery connection probabilities are  $p$  and  $s$ , respectively, while the between group connection probabilities are  $q$  and  $r$ . Assume the ordering  $s < r, q < p$ .

Let the matrix  $Q$  be

$$Q = \begin{bmatrix} pJ_{n_1 \times n_1} & qJ_{n_1 \times n_2} \\ rJ_{n_2 \times n_1} & sJ_{n_2 \times n_2} \end{bmatrix}, \quad (9)$$

where  $J_{n,m}$  is a matrix  $n \times m$  whose entries are all equal to 1.

**Proposition 2.** Let  $A \in \{0, 1\}^{N \times N}$  be an adjacency matrix of a directed graph. Let  $\tilde{A} = AA \in [0, N]^{N \times N}$  denote the doubled adjacency matrix. Let  $\tilde{Q} = \mathbb{E}(\tilde{A})$ , and suppose that the entries of  $\tilde{Q}$  lie in the range

$$[n_1 rq + n_2 s^2, n_1 p^2 + n_2 qr],$$

where  $0 \leq s < r, q < p \leq 1$ . For fixed  $u, v \in \{1, \dots, N\}$ ,  $\{\tilde{A}_{uw}\tilde{A}_{vw}\}_{w=1}^N$  are conditionally independent given  $\mathcal{F} = \sigma(A_{u*}, A_{v*})$ , where  $\sigma(A_{u*}, A_{v*})$  is the information contained in the  $u$ -th and  $v$ -th rows of the matrix  $A$ . Then, with probability at least  $1 - 2N^{2 - \frac{4}{N^2}\tilde{T}_1^2}$ ,

$$\|\tilde{A}\tilde{A}^\top - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F \leq \sqrt{2}\tilde{T}_1\sqrt{N^5 \log N}, \quad (10)$$

where

$$\tilde{T}_1 = \left[ \sum_{w \in G_1} (1 - (\pi_1 pr + \pi_2 rs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2) \right].$$

In addition, with probability at least  $1 - 2N^{2 - \frac{4}{N^2}\tilde{T}_2^2}$ ,

$$\|\tilde{A}^\top\tilde{A} - \mathbb{E}(\tilde{A}^\top\tilde{A})\|_F \leq \sqrt{2}\tilde{T}_2\sqrt{N^5 \log N}, \quad (11)$$

where

$$\tilde{T}_2 = \left[ \sum_{w \in G_1} (1 - (\pi_1 pq + \pi_2 qs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2) \right].$$

Based on the Proposition 2, we have the misclustering rate of DASE when the graph is directed and core-periphery as follows.

**Theorem 4.2.** Under the directed condition in Section 4.1 and Assumptions 1, 2 and 4, suppose that the number of blocks  $K$  and the latent vector dimension  $d$  are known. Let  $\hat{\tau}^{(N)} : V \mapsto \{1, \dots, K\}$  be the block assignment obtained by clustering the rows of  $\tilde{W}$ , where  $\tilde{W} = [\tilde{U} | \tilde{V}]$ , satisfying Equation (5). Let  $H_K$  denote the set of permutations of  $[K]$ . Then, it almost always holds that

$$\min_{\rho \in H_K} |\{u \in V : \tau(u) \neq \rho(\hat{\tau}(u))\}| \leq \frac{2^4 3^2 \tilde{T}}{\hat{\beta}^2 (\tilde{b}\tilde{\pi})^5} \frac{\log N}{N},$$

where  $\tilde{T} > 0$  and  $\tilde{T}N^2 = \tilde{T}_1^2 + \tilde{T}_2^2$ .

The proofs of Proposition 2 and Theorem 4.2 are provided in Appendix A.3.

For the purpose of comparing clustering consistency between DASE and ASE, we also derive the misclustering rate for core-periphery graphs based on the proof in (Sussman et al. 2012). The consistency bound for ASE on core-periphery graphs is as follows:

**Corollary 4.2.1.** Under the directed condition in Section 4.1 and Assumptions 1, 2 and 4, suppose that the number of blocks  $K$  and the latent vector dimension  $d$  are known. Let  $\hat{\tau}^{(N)} : V \mapsto \{1, \dots, K\}$  be the block assignment obtained by clustering the rows of  $W$ , where  $W = [U | V]$ , satisfying Equation (5). Let  $H_K$  denote the set of permutations of  $[K]$ . Then, it almost always holds that

$$\min_{\rho \in S_K} |\{u \in V : \tau(u) \neq \rho(\hat{\tau}(u))\}| \leq \frac{2^2 3^2 6 \cdot T}{\beta^2 (\tilde{b}\tilde{\pi})^5} \cdot \log N,$$

where  $T > 0$  and  $TN^2 = T_1^2 + T_2^2$ , with

$$\begin{aligned} TN^2 &:= T_1^2 + T_2^2 \\ &= \left[ \sum_{w \in G_1} (1 - r^2) + \sum_{w \in G_2} (1 - s^2) \right]^2 \\ &\quad + \left[ \sum_{w \in G_1} (1 - q^2) + \sum_{w \in G_2} (1 - s^2) \right]^2. \end{aligned}$$

The proof of Corollary 4.2.1 is provided in Appendix A.3.2.

Based on Theorem 4.2 and Corollary 4.2.1, we have a following remark to compare ASE and DASE.

**Remark 1.** *Therefore, as  $N$  increases, the misclustering rate of DASE converges to zero faster than that of ASE, indicating that DASE achieves a strictly stronger consistency guarantee in core-periphery networks.*

## 5 Simulation Study

We evaluate our proposed method, Doubled Adjacency Spectral Embedding (DASE), and compare it to spectral clustering (Ng et al. 2001) and Adjacency Spectral Embedding (ASE) (Priebe et al. 2019) using simulated data. To ensure generality, we consider both directed and undirected graphs. Similar to most existing studies, which primarily consider networks with only two clusters (Priebe et al. 2019; Sussman et al. 2012).

For each setting, we examine three simulation scenarios: (i) fixing the network size while varying the network density, (ii) fixing the network density while varying the network size, and (iii) fixing both network size and the block probability matrix  $B$  while varying the group proportions  $\pi$ . We denote the network size by the number of nodes,  $N$ . The network density is defined as the edge density:

$$\text{density} = \begin{cases} \frac{m}{N(N-1)}, & \text{if the network is directed,} \\ \frac{2m}{N(N-1)}, & \text{if the network is undirected,} \end{cases}$$

where  $m$  denotes the number of edges in the network.

To generate the adjacency matrix using the Stochastic Block Model, we define the block probability  $B$  as follows:

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = s \times \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix},$$

where  $s$  controls the expected network density.  $r_{ij}$  represents the relative proportion of edges between groups  $i$  and  $j$ . We set  $r_{11} = 1$  as the baseline ratio.

For example, we consider  $K = 2$  in simulation study. The corresponding block probability matrix  $B$  is defined as:

$$B = s \times \begin{bmatrix} 1 & 0.6 \\ 0.6 & 0.3 \end{bmatrix}$$

This structure ensures that the relative proportions of within-group and between-group edges remain consistent when varying the overall density (via  $s$ ) while keeping the network size fixed. For example, when  $K = 2$ , if there are 100 edges in the first group, then there are approximately 60 edges between two groups and 30 edges in the second group. Thus, this proportional matrix allows us to vary the density without altering the underlying connectivity structure, hence we adopt this configuration throughout our simulation study.

For computational efficiency and consistency across clustering methods, we compute only the top  $d$  largest singular values and corresponding singular vectors using `svds` for ASE and DASE, and `eigsh` for spectral clustering from the `scipy.sparse.linalg` package (Virtanen et al. 2020) in Python.

We use the Normalised Mutual Information (NMI) (Strehl et al. 2002) to assess clustering performance. NMI measures the common indices between two cluster labellings. It ranges between 0 and 1, where 0

indicates no mutual information and 1 indicates identical labels. For each simulation setting, we obtain the average NMI and its corresponding standard deviation over 50 iterations. For computational efficiency comparison, we also record the average runtime across the same 50 iterations.

In this section, simulation results for directed and undirected graphs are presented in Sections 5.1 and 5.2, respectively. For both settings, we compare the clustering performance results for spectral clustering (circle, blue), ASE (triangle, orange), and DASE (square, green) with  $k$ -means. Results using the Gaussian Mixture Model (GMM) are described in Appendix B.

## 5.1 Directed Graphs

The clustering performance on directed graphs using  $k$ -means with balanced two-cluster networks  $\pi = (0.5, 0.5)^\top$  is shown in Figure 2. Figure 2a illustrates the average NMI with the corresponding standard deviation (shaded areas) over 50 iterations for a fixed network size of  $N = 1,000$ , while varying the network density. Figures 2b, 2c, and 2d illustrate the average NMI with standard deviation (shaded areas), the standard deviation alone, and the computational cost, respectively, for a fixed expected network density of  $s = 0.05$  while varying the network size.

As shown in Figures 2a and 2b, spectral clustering fails to capture the underlying network structure. In contrast, ASE and DASE, which are designed for the core-periphery graphs, successfully recover the community structure when the network is not sparse or when the network size is large. However, when the network is sparse or the network is small, ASE not only struggles to capture the network structure but also exhibits instability, as indicated by its larger standard deviation compared with DASE in Figure 2c. Regarding computational cost in Figure 2d, DASE is slower than ASE due to the additional computation required to form  $AA$ .

We examine unbalanced two-cluster networks in Figure 3. In this setting, the network size is fixed at  $N = 1,000$ , and the block probability matrix  $B$  is fixed by

$$B = \begin{bmatrix} 0.08 & 0.048 \\ 0.048 & 0.024 \end{bmatrix},$$

which preserves the same probability ratios as in the previous simulation. Since  $B$  is fixed, the overall edge density naturally varies with the group proportions. Here, we focus on the proportion of the size of the core group ( $\pi_1$ ). As in the previous simulations, the average NMI and the corresponding standard deviation are computed over 50 iterations.

As expected, spectral clustering fails to detect the network structures across different core sizes. For ASE, the accuracy is highest when the network is balanced ( $\pi_1 = \pi_2 = 0.5$ ). For both ASE and DASE, the accuracy decreases as the core group becomes more dominant. This occurs because, when the core group is large, the network increasingly resembles a single dense cluster, making it difficult to distinguish between core and periphery. In this case, DASE fails to detect the structure when 90% of the nodes belong to the core group ( $\pi = (0.9, 0.1)^\top$ ), whereas ASE fails even when the core group is relatively small.

## 5.2 Undirected Graphs

We also consider undirected, balanced two-cluster networks ( $\pi = (0.5, 0.5)^\top$ ) to evaluate the clustering performance of spectral clustering, ASE, and DASE, as shown in Figure 4. Similar to the directed case, our primary comparisons are based on clustering accuracy (average NMI) and the corresponding standard deviation (shaded areas). Figures 4a and 4b illustrate the average NMI with the corresponding standard deviation over 50 iterations for a fixed network size of  $N = 1,000$  and a fixed expected network density of  $s = 0.05$ , respectively.

As shown in Figures 4a, ASE performs better than spectral clustering, as expected for core-periphery graphs. However, as the network becomes denser, ASE still struggles to identify the underlying network structures and exhibits large variability. In contrast, DASE outperforms both methods, even when the network is sparse. When the network density is fixed as shown in Figure 4b, ASE fails to recover accurate clustering assignments even as the network size increases, and it continues to yield a larger standard deviation. In comparison, DASE converges more rapidly to the correct clustering and maintains a smaller standard deviation across network sizes.

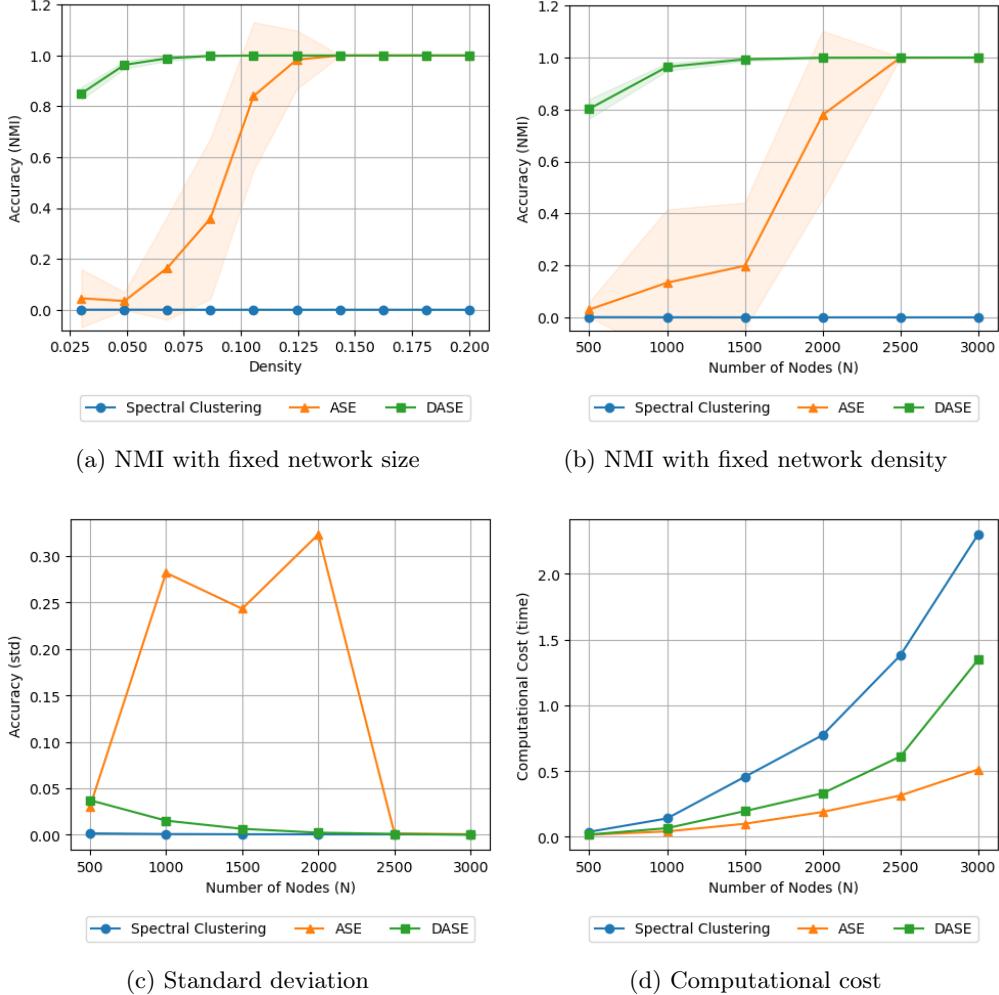


Figure 2: Figures illustrating the comparison of clustering performance on directed graphs using  $k$ -means in terms of mean accuracy (NMI), corresponding standard deviation, and mean computational cost over 50 iteration when  $K = 2$  with  $\pi = (0.5, 0.5)$ : (a) NMI with fixed network size ( $N = 1,000$ ) and varying network density; (b)-(d) NMI, standard deviation, and computational cost with fixed expected network density ( $s = 0.05$ ) and varying network sizes. In (a) and (b), shaded areas represent the standard deviations.

We examine unbalanced two-cluster networks, as shown in Figure 5, extending the earlier analysis in which only balanced settings were considered for undirected networks. In this simulation, the network size is fixed at  $N = 1,000$ , and the block probability matrix is given by

$$B = \begin{bmatrix} 0.08 & 0.048 \\ 0.048 & 0.024 \end{bmatrix},$$

while the group ratio  $\pi$  is varied.

As illustrated in Figure 5, ASE consistently struggles to identify the network structure across all ratios, even when the core and periphery groups are balanced. In contrast, DASE achieves substantially better performance over a wide range of ratios, except  $\pi_1 = 0.1$  and  $\pi_1 = 0.9$ . Even in these extreme cases ( $\pi_1 = 0.1$  and  $\pi_1 = 0.9$ ), DASE still outperform ASE, although the overall accuracy decreases for both methods. Overall, we conclude that when the core group is present but not dominant, DASE achieves better clustering performance.

To conclude the simulation results, we observe a clear difference in clustering performance between directed

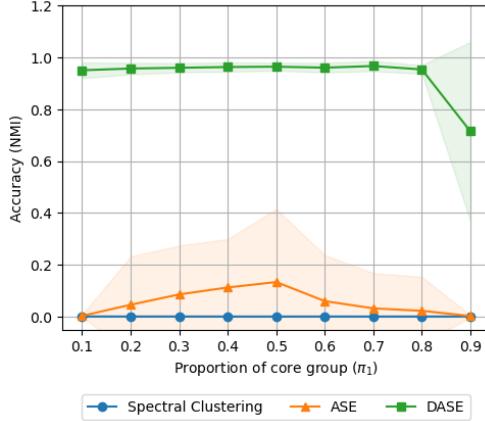


Figure 3: Figure showing the comparison of clustering performance on directed graphs using  $k$ -means in terms of mean NMI (line) and the corresponding standard deviation (shaded area) over 50 iterations when  $K = 2$ . In the simulation, the network size is fixed at ( $N = 1,000$ ), and the block probability matrix  $B$  is fixed, while varying the core group ratio ( $\pi_1$ ) from 0.1 to 0.9.

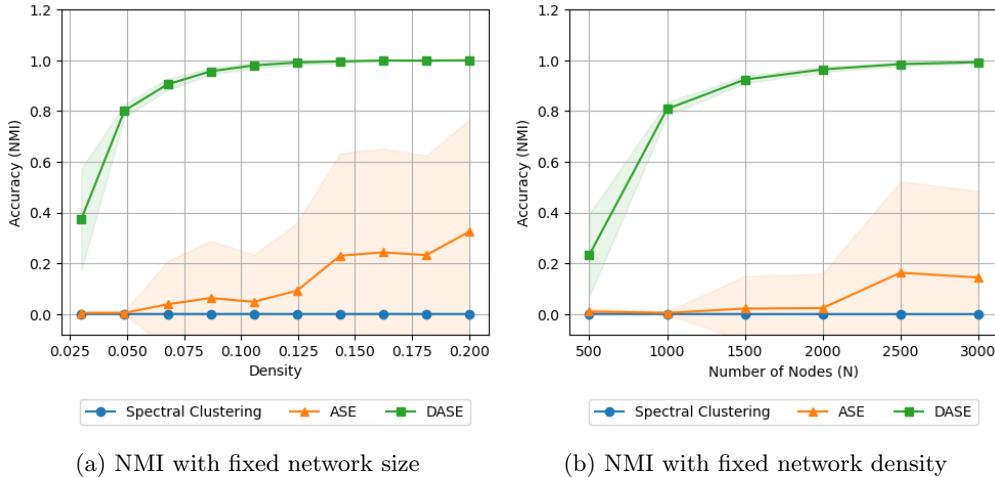


Figure 4: Figures illustrating the comparison of clustering performance on undirected graphs using  $k$ -means in terms of mean accuracy (NMI), corresponding standard deviation, and mean computational cost over 50 iteration when  $K = 2$  with  $\pi = (0.5, 0.5)$ : (a) NMI with fixed network size ( $N = 1,000$ ) and varying network density; (b) NMI with fixed expected network density ( $s = 0.05$ ) and varying network sizes. In (a) and (b), shaded areas represent the standard deviations.

and undirected graphs. ASE performs noticeably worse on undirected graphs than on directed graphs. This is expected, as directed graphs retain additional information through edge directions, while undirected graphs only preserve the presence or absence of connections.

Furthermore, DASE consistently outperforms the other two methods in terms of both NMI and standard deviation. Overall, we conclude that DASE provides superior clustering performance, particularly when the network is sparse or the network size is small, for both directed and undirected settings.

### 5.3 Chernoff Information

The Chernoff Information has been widely used as a metric for the separability between two distributions (Chernoff 1952; Chernoff 1956). Based on this measure, several works have proposed using Chernoff information to quantify community separation in spectral embedding spaces. A large Chernoff information indicates that

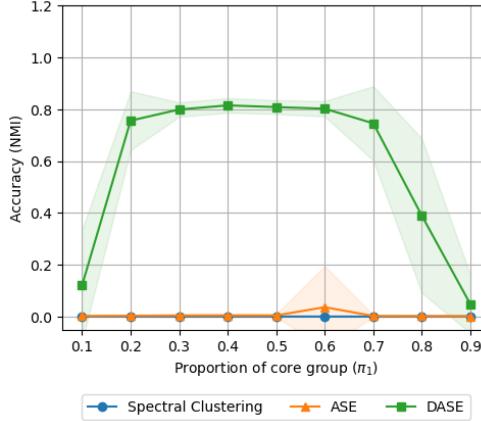


Figure 5: Figure presenting the comparison of clustering performance on undirected graphs using  $k$ -means in terms of mean NMI (line) and the corresponding standard deviation (shaded area) over 50 iterations when  $K = 2$ . In the simulation, the network size is fixed at ( $N = 1,000$ ), and the block probability matrix  $B$  is fixed, while varying the core group ratio ( $\pi_1$ ) from 0.1 to 0.9.

two embedded distributions are more statistically distinguishable.

**Lemma 1.** Let  $K$  be the number of communities in the network. Suppose the adjacency matrix  $A$  is generated from a Stochastic Block Model as defined in 2.1 with full-rank mean matrix  $M$ ,  $\text{rank}(M) = K$ . The size-adjusted Chernoff Information (Gallagher et al. 2024) is given by

$$CI = \min_{k \neq l} \sup_{t \in (0,1)} \left[ \frac{t(1-t)}{2} \{e^T M \Pi S_{kl}(t)^{-1} M e\} \right], \quad (12)$$

where  $e = (e_k - e_l)$ ,  $e_k \in \mathbb{R}^K$  is the standard basis vector with  $k$ -th element as one, zero elsewhere,  $\Pi = \text{diag}(\pi_1, \dots, \pi_K)$ , and

$$S_{kl}(t) = (1-t)\text{diag}(C_k) + t\text{diag}(C_l) \in \mathbb{R}^{K \times K}$$

is a convex combination of the diagonal matrices formed from rows  $C_k$  and  $C_l$  of the block variance matrix  $V$ .

For ASE, the mean block matrix  $M$  and the block variance matrix  $C$  are defined as

$$\begin{aligned} M_{\theta_i, \theta_j} &= B_{\theta_i, \theta_j} \in [0, 1] \quad \text{and} \\ V_{\theta_i, \theta_j} &= (1 - M_{\theta_i, \theta_j}) M_{\theta_i, \theta_j} \in [0, 1]. \end{aligned}$$

where  $\theta_i, \theta_j$  are the block labels of nodes  $i$  and  $j$ , respectively.

For DASE, the doubled adjacency  $A$  follows a Poisson-Binomial distribution, with mean and variance as follows:

$$\begin{aligned} \tilde{Q}_{ij} &= \sum_{k=1}^N Q_{ik} Q_{kj} \quad \text{and} \\ \tilde{Q}'_{ij} &= \sum_{k=1}^N (1 - Q_{ik} Q_{kj}) Q_{ik} Q_{kj}. \end{aligned}$$

Thus, the mean block matrix  $\tilde{M}$  and the block variance matrix  $\tilde{C}$  are defined as

$$\begin{aligned} \tilde{M}_{\theta_i, \theta_j} &= \frac{1}{n_{\theta_i} n_{\theta_j}} \sum_{a \in G_{\theta_i}} \sum_{b \in G_{\theta_j}} \tilde{Q}_{ab} \quad \text{and} \\ \tilde{C}_{\theta_i, \theta_j} &= \frac{1}{n_{\theta_i} n_{\theta_j}} \sum_{a \in G_{\theta_i}} \sum_{b \in G_{\theta_j}} \tilde{Q}'_{ab}, \end{aligned}$$

where  $G_{\theta_i}$  denotes the  $\theta_i$ -th group, and  $n_{\theta_i} = |G_{\theta_i}|$  denotes the size of the  $\theta_i$ -th group.

To empirically compare the Chernoff information for ASE and DASE, we conduct two simple simulations based on the Stochastic Block Model. We use the probability matrix

$$B = \begin{bmatrix} 0.8 & 0.48 \\ 0.48 & 0.24 \end{bmatrix},$$

where the rank of the matrix is  $\text{rank}(B) = 2$  and the overall edge density is 0.5. We assume that the initial cluster proportions are equal, i.e.,  $\pi_1 = \pi_2 = 0.5$ .

We run two simulations. First, we vary the number of nodes  $N$  and compute the Chernoff information for ASE and DASE. The results are presented in Figure 6a.

Second, we fix the number of nodes at  $N = 1,000$  and vary the group proportion  $(\pi_1, \pi_2)$ . The probability matrix  $B$  remains unchanged to the first simulation. The overall edge density of the network is

$$\pi_1^2 p + \pi_1 \pi_2 q + \pi_1 \pi_2 r + \pi_2^2 s.$$

Thus, changing  $\pi_1$  and  $\pi_2$  naturally affects the network density. In particular, increasing the proportion of the core group results in a denser network. For this simulation, the density ranges from 0.2888 to 0.7368. The corresponding Chernoff information results are shown in Figure 6b.

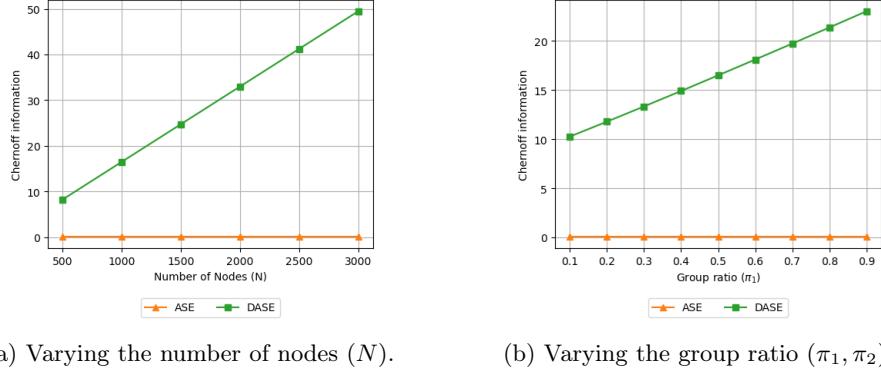


Figure 6: Chernoff Information for core-periphery networks. Comparison of performance (CI) between ASE and DASE across varying network size ( $N$ ) and varying group proportions ( $\pi$ ).

As observed, when  $K = 2$ , the Chernoff information for DASE is generally larger than that of ASE, even when the network is relatively sparse. This implies that the distribution of the two clusters obtained from DASE embedding are more easily separated than those derived from ASE.

## 6 Real Dataset

In this section, we use real-world datasets to compare the clustering methods discussed in the previous section. We evaluate clustering accuracy using Normalized Mutual Information (NMI) when the ground truth is available, and use Chernoff information to identify the preferred method. Because of the Chernoff information analysis, we mainly compare ASE and DASE in terms of clustering performance.

### 6.1 Faculty Hiring Network

We apply our proposed method to real-world datasets, one of which is the faculty hiring network (Clauset et al. 2015). This dataset includes three different departments: Business, Computer Science, and History. For our analysis, we focus on the Computer Science dataset, which contains information on gender, faculty positions, and institutions from 23 Canadian universities and 182 American universities, collected from May 2011 to March 2012. In this network, each node represents a university and each directed edge  $(u, v)$  indicates that a faculty member at university  $v$  obtained their PhD degree from university  $u$ .

The key finding of this paper is the majority of faculty members get their PhD degrees from the top 15% of institutions, which is more core institutions and have higher connectivity within the network. Therefore, they suggested that the faculty hiring network would exhibit a core-periphery pattern.

In this dataset, there are 205 institutions, 2,757 faculty hiring connections, and the density of the network is 0.0659. To better understand the network structure, we visualize the top 15% of universities highlighted in yellow, as shown in Figure 7a. For comparison of clustering results, we present a heatmap representing the core group in Figure 7b. As expected, the core universities are located at the center of the network. Consistent with the findings of (Clauset et al. 2015) these core institutions produce more faculty members than other institutions.

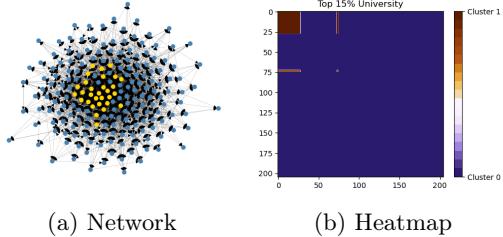


Figure 7: Faculty hiring network highlighting the top 15% of universities: (a) the network visualization with core universities shown in yellow, and (b) the corresponding heatmap illustrating the core group.

Here, we expect one core group and one periphery group, which represent the top 15% universities and the remaining ones, respectively. Hence, we assume that there are two clusters ( $K = 2$ ). The corresponding clustering results that illustrate the core regions obtained from spectral clustering, ASE and DASE with  $K = 2$  by  $k$ -means are presented in Figure 8. Here, we use heatmaps to compare the clustering results. Figures 8a-8c show the heatmaps for  $K = 2$ . As we can see, spectral clustering fails to detect the core group. In contrast, ASE and DASE have better clustering results to capture the core group, as indicated by the top-left corner in the heatmaps.

Having compared the clustering results using heatmap visualisations, we now compare the clustering performance based on Normalized Mutual Information (NMI) and Chernoff information, defined in Eq.(12), using both  $k$ -means and GMM as presented in Table 1. Since our main interest lies in comparing performance between ASE and DASE, we focus on these two methods.

For the NMI, we perform 50 iterations for  $k$ -means and GMM, and compute the mean and standard deviation. When  $K = 2$ , DASE performs better than ASE with  $k$ -means, whereas ASE performs better than DASE with GMM.

For the Chernoff information, as expected, DASE has higher values than ASE for  $K = 2$  with both  $k$ -means and GMM. Therefore, we can conclude that, for any clustering techniques ( $k$ -means and GMM), DASE yields higher Chernoff information than ASE, indicating that DASE has higher separability in the spectral embedding space than ASE.

To compare the improvement in NMI and Chernoff information, we consider their ratios between DASE and ASE, defined as

$$NMI_{ratio} = \frac{NMI_{DASE}}{NMI_{ASE}} \quad \text{and} \quad CI_{ratio} = \frac{CI_{DASE}}{CI_{ASE}}$$

A ratio greater than 1 indicates that DASE achieves higher NMI or Chernoff information than ASE, suggesting improved clustering performance and better community separation in the embedding space.

The corresponding ratio table is in Table 2. For NMI, as expected, using  $k$ -means yields greater than 1, while using GMM yields smaller than 1. On the other hand, for the Chernoff information, all ratio values,  $CI_{ratios}$ , are greater than 1. Moreover, the ratios are at least 3, indicating that DASE exhibits much higher separability in the spectral embedding space than ASE. Therefore, we can conclude that DASE is a more preferable clustering method than ASE for this network structure in all cases.

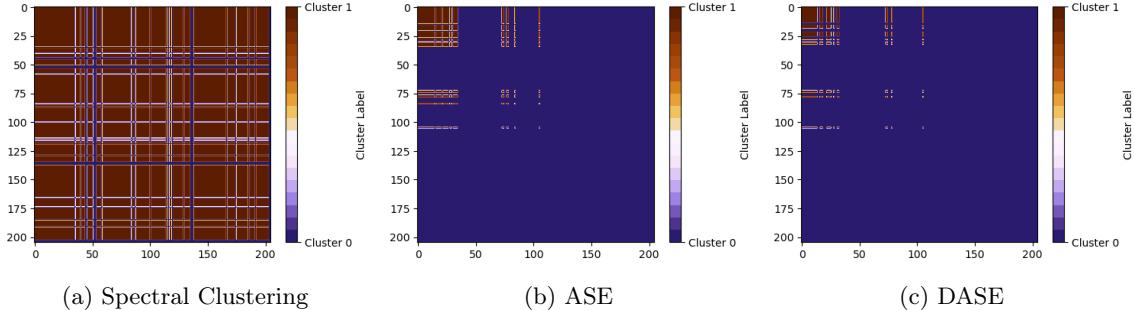


Figure 8: Heatmaps illustrating the top 15% universities in the faculty hiring network, based on clustering results obtained from spectral clustering, ASE and DASE. Figures (a)-(c) correspond to  $K = 2$  with the  $k$ -means clustering.

$K$	NMI				Chernoff Information			
	$k$ -means		GMM		$k$ -means		GMM	
	ASE	DASE	ASE	DASE	ASE	DASE	ASE	DASE
2	0.5348 (0.00)	<b>0.5371</b> (0.02)	<b>0.3588</b> (0.00)	0.3013 (0.00)	0.0187 (0.00)	<b>0.1818</b> (0.00)	0.0150 (0.00)	<b>0.1917</b> (0.00)

Table 1: Clustering performance (NMI and Chernoff information) of ASE vs DASE at  $K = 2$  using  $k$ -means and GMM. For both NMI and Chernoff information, the mean and standard deviation are computed over 50 iterations.

## 6.2 Air Traffic Network

Another real-world dataset is the crowdsourced air traffic data from the OpenSky Network during 2019 (Strohmeier et al. 2021). The dataset contains 15,439 airports, including both commercial and private operations and 801,748 flight connections. So, the edge density of this network is 0.0034. Since flight connections have inherent directions, this network is a directed graph. In this dataset, multiple connections exist between pairs of airports, making the network weighted. However, we convert the weighted edges into binary ones to indicate their connections, as our model assumes a binary adjacency matrix.



Figure 9: OpenSky air traffic network for 2019. Each node represents an airport, and each directed edges represents a flight connection. This dataset includes both commercial and private airports/flights.

For the number of clusters, we consider two different values:  $K = 2$  and  $K = 4$ . To identify one core group and one periphery group, we assume two clusters ( $K = 2$ ). Based on the profile likelihood method (Zhu et al. 2006), there are four clusters ( $K = 4$ ).

$K$	NMI		Chernoff Information	
	<i>k</i> -means	GMM	<i>k</i> -means	GMM
2	1.0042	0.8397	9.7291	12.7553

Table 2: Comparison of Chernoff information and DCPM ratios (DASE/ASE) of directed graphs using K-Means and GMM under different numbers of clusters  $K$  for the faculty hiring network.

$K$	<i>k</i> -means		GMM	
	ASE	DASE	ASE	DASE
2	0.009 (0.00)	<b>1.9623</b> (0.01)	0.0018 (0.00)	<b>0.1845</b> (0.00)
4	0.0035 (0.00)	<b>0.4973</b> (0.09)	0.0005 (0.00)	<b>0.004</b> (0.01)

Table 3: Comparison of the Chernoff information for the air traffic network using *k*-means and GMM with ASE and DASE.

In this dataset, it is not appropriate to define the ground truth using the highest-degree airports as core groups, since private airports are included. These private airports and private flights can distort the network structure and hinder the accurate identification of the true core groups. Therefore, we compare only the Chernoff information for different values of  $K$ , and not using NMI.

A comparison of ASE and DASE using *k*-means and GMM is summarised in Table 3. As we can see, for each number of clusters, DASE obtains higher Chernoff information values regardless of whether *k*-means or GMM is used.

Similar to the faculty hiring network, we compare the ratio of the Chernoff information of ASE and DASE. The corresponding ratio table is in Table 4. As we can see, all the ratio values,  $CI_{\text{ratios}}$ , are larger than 1. Moreover, the ratios are at least 7, indicating that DASE exhibits much higher separability in the spectral embedding space than ASE. Therefore, we can conclude that DASE is a more preferable clustering method than ASE for this network structure in all cases.

## 7 Conclusion

We propose *Double Adjacency Spectral Embedding (DASE)*, a clustering method designed for core-periphery graphs. This method leverages the doubled adjacency matrix, which represents the possible ways to reach a node in two steps, thereby capturing more structural information about the graph. This approach differs from the previous method for core-periphery graph, the Adjacency Spectral Embedding (ASE), which uses the single adjacency matrix, which is the indicator matrix of graph connections. In this paper, we showed that DASE is consistent for large networks. We also compared ASE and DASE using the Chernoff information to determine which method is preferred on core-periphery graphs. In addition to the theoretical comparison, we examined the simulation results. Our method outperforms other methods, spectral clustering and ASE, when the graph is sparse. Moreover, in terms of the standard deviation of clustering performance, our method is more consistent than ASE in sparse graphs. The improved performance of DASE was demonstrated through simulated examples. Additionally, we applied our method to a real dataset and compared the clustering results with those obtained using other methods.

## 8 Acknowledgements

Sinyoung Park is supported by a scholarship from the EPRSC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/S022945/1. MAN gratefully acknowledge support from EPSRC grant EP/X002195/1.

$K$	Chernoff Information	
	<i>k</i> -means	GMM
2	202.8083	104.1086
4	142.7246	7.6585

Table 4: Comparison of Chernoff information ratios (DASE/ASE) of directed graphs using K-Means and GMM under different numbers of clusters  $K$  for the faculty hiring network.

## References

- Abbe, Emmanuel (2018). “Community detection and stochastic block models: recent developments”. In: *Journal of Machine Learning Research* 18.177, pp. 1–86.
- Adamic, Lada A and Natalie Glance (2005). “The political blogosphere and the 2004 US election: divided they blog”. In: *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43.
- Athreya, Avanti et al. (2018). “Statistical inference on random dot product graphs: a survey”. In: *Journal of Machine Learning Research* 18.226, pp. 1–92.
- Borgatti, Stephen P and Martin G Everett (2000). “Models of core/periphery structures”. In: *Social Networks* 21.4, pp. 375–395.
- Chernoff, Herman (1952). “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations”. In: *The Annals of Mathematical Statistics*, pp. 493–507.
- (1956). “Large-sample theory: Parametric case”. In: *The Annals of Mathematical Statistics* 27.1, pp. 1–22.
- Clauset, Aaron, Samuel Arbesman, and Daniel B. Larremore (2015). “Systematic inequality and hierarchy in faculty hiring networks”. In: *Science Advances* 1.1, e1400005.
- Csermely, Peter et al. (2013). “Structure and dynamics of core/periphery networks”. In: *Journal of Complex Networks* 1.2, pp. 93–123.
- Cucuringu, Mihai et al. (2020). “Hermitian matrices for clustering directed graphs: insights and applications”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 983–992.
- Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the Royal Statistical Society: Series B (methodological)* 39.1, pp. 1–22.
- Derudder, Ben, Lomme Devriendt, and Frank Witlox (2007). “Flying where you don’t want to go: An empirical analysis of hubs in the global airline network”. In: *Tijdschrift voor Economische en Sociale Geografie* 98.3, pp. 307–324.
- Elliott, Andrew et al. (2020). “Core–periphery structure in directed networks”. In: *Proceedings of the Royal Society A* 476.2241, p. 20190783.
- Fjällström, Per-Olof (1998). *Algorithms for graph partitioning: A survey*. Linköping University Electronic Press.
- Fortunato, Santo (2010). “Community detection in graphs”. In: *Physics reports* 486.3-5, pp. 75–174.
- Gallagher, Ian et al. (2024). “Spectral embedding of weighted graphs”. In: *Journal of the American Statistical Association* 119.547, pp. 1923–1932.
- Gallagher, Ryan J, Jean-Gabriel Young, and Brooke Foucault Welles (2021). “A clarified typology of core–periphery structure in networks”. In: *Science Advances* 7.12, eabc9800.
- Goldenberg, Anna et al. (2010). “A survey of statistical network models”. In: *Foundations and Trends® in Machine Learning* 2.2, pp. 129–233.
- Holland, Paul W, Kathryn Blackmond Laskey, and Samuel Leinhardt (1983). “Stochastic blockmodels: First steps”. In: *Social Networks* 5.2, pp. 109–137.
- Lloyd, Stuart (1982). “Least squares quantization in PCM”. In: *IEEE Transactions on Information Theory* 28.2, pp. 129–137.
- MacQueen, J (1967). “Multivariate observations”. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1, pp. 281–297.
- McLachlan, Geoffrey J and Kaye E Basford (1988). “Mixture models: inference and applications to clustering”. In: *Statistics, textbooks and monographs*.

- Newman, Mark EJ and Michelle Girvan (2004). "Finding and evaluating community structure in networks". In: *Physical Review E* 69.2, p. 026113.
- Ng, Andrew, Michael Jordan, and Yair Weiss (2001). "On spectral clustering: Analysis and an algorithm". In: *Advances in Neural Information Processing Systems* 14.
- Pagani, Giuliano Andrea and Marco Aiello (2013). "The power grid as a complex network: a survey". In: *Physica A: Statistical Mechanics and its Applications* 392.11, pp. 2688–2700.
- Pavlopoulos, Georgios A et al. (2011). "Using graph theory to analyze biological networks". In: *BioData Mining* 4.1, p. 10.
- Priebe, Carey E et al. (2019). "On a two-truths phenomenon in spectral graph clustering". In: *Proceedings of the National Academy of Sciences* 116.13, pp. 5995–6000.
- Rohe, Karl, Sourav Chatterjee, and Bin Yu (2011). "Spectral clustering and the high-dimensional stochastic blockmodel". In: *The Annals of Mathematical Statistics* 39.4, pp. 1878–1915.
- Rombach, M Puck et al. (2014). "Core-periphery structure in networks". In: *SIAM Journal on Applied Mathematics* 74.1, pp. 167–190.
- Rombach, Puck et al. (2017). "Core-periphery structure in networks (revisited)". In: *SIAM review* 59.3, pp. 619–646.
- Rubin-Delanchy, Patrick et al. (2022). "A statistical interpretation of spectral embedding: the generalised random dot product graph". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.4, pp. 1446–1473.
- Rubinov, Mikail and Olaf Sporns (2010). "Complex network measures of brain connectivity: uses and interpretations". In: *NeuroImage* 52.3, pp. 1059–1069.
- Satuluri, Venu and Srinivasan Parthasarathy (2011). "Symmetrizations for clustering directed graphs". In: *Proceedings of the 14th international conference on extending database technology*, pp. 343–354.
- Shi, Jianbo and Jitendra Malik (2000). "Normalized cuts and image segmentation". In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8, pp. 888–905.
- Strehl, Alexander and Joydeep Ghosh (2002). "Cluster ensembles—a knowledge reuse framework for combining multiple partitions". In: *Journal of Machine Learning Research* 3.Dec, pp. 583–617.
- Strohmeier, Martin et al. (2021). "Crowdsourced air traffic data from the OpenSky Network 2019–2020". In: *Earth System Science Data* 13.2, pp. 357–366.
- Sussman, Daniel L et al. (2012). "A Consistent Adjacency Spectral Embedding for Stochastic Blockmodel Graphs". In: *Journal of the American Statistical Association* 107.499, pp. 1119–1128.
- Tang, Minh and Carey E Priebe (2018). "Limit theorems for eigenvectors of the normalized Laplacian for random graphs". In: *Annals of Statistics* 46.5, pp. 2360–2415.
- Tang, Wenli et al. (2019). "Recent advance on detecting core-periphery structure: a survey". In: *CCF Transactions on Pervasive Computing and Interaction* 1.3, pp. 175–189.
- Tang, Wenpin and Fengmin Tang (2023). "The Poisson binomial distribution—Old & new". In: *Statistical Science* 38.1, pp. 108–119.
- Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.
- Von Luxburg, Ulrike (2007). "A tutorial on spectral clustering". In: *Statistics and Computing* 17.4, pp. 395–416.
- Wang, Zhe, Yingbin Liang, and Pengsheng Ji (2020). "Spectral algorithms for community detection in directed networks". In: *Journal of Machine Learning Research* 21.153, pp. 1–45.
- Wey, Tina et al. (2008). "Social network analysis of animal behaviour: a promising tool for the study of sociality". In: *Animal behaviour* 75.2, pp. 333–344.
- Young, Stephen J and Edward R Scheinerman (2007). "Random dot product graph models for social networks". In: *International Workshop on Algorithms and Models for Web-Graph*. Springer, pp. 138–149.
- Zhang, Fuzhen and Qingling Zhang (2006). "Eigenvalue Inequalities for Matrix Product". In: *IEEE Transactions on Automatic Control* 51.9, pp. 1506–1509.
- Zhang, Xiao, Travis Martin, and Mark EJ Newman (2015). "Identification of core-periphery structure in networks". In: *Physical Review E* 91.3, p. 032803.
- Zhu, Mu and Ali Ghodsi (2006). "Automatic dimensionality selection from the scree plot via the use of profile likelihood". In: *Computational Statistics & Data Analysis* 51.2, pp. 918–930.

## A Proof of Misclustering rates

This proof follows the approach of (Sussman et al. 2012).

### A.1 Notations

Based on Section 4.1, we keep using the notations we defined. Moreover, for the proof of misclustered nodes, we define more notations related to singular value decompositions.

For ASE, we changed some notations. The singular value decompositions of the matrices  $\tilde{Q}$  and  $\tilde{A}$  are as follows:

$$Q = U\Sigma V^\top \quad \text{and} \quad A = U'\Sigma' V'^\top,$$

where  $U, U', V, V' \in \mathbb{R}^{N \times d}$  and  $\Sigma, \Sigma' \in \mathbb{R}^{d \times d}$ . Therefore, the concatenations are

$$W = [U|V] \in \mathbb{R}^{n \times 2d} \quad \text{and} \quad W' = [U'|V'] \in \mathbb{R}^{n \times 2d}.$$

For DASE, the singular value decompositions of the matrices  $\tilde{Q}$  and  $\tilde{A}$  are as follows:

$$\tilde{Q} = QQ = \tilde{U}\tilde{\Sigma}\tilde{V}^\top \quad \text{and} \quad \tilde{A} = AA = \tilde{U}'\tilde{\Sigma}'\tilde{V}'^\top,$$

where  $\tilde{U}, \tilde{U}', \tilde{V}, \tilde{V}' \in \mathbb{R}^{N \times d}$  and  $\tilde{\Sigma}, \tilde{\Sigma}' \in \mathbb{R}^{d \times d}$ . Therefore, the concatenations are

$$\tilde{W} = [\tilde{U}|\tilde{V}] \in \mathbb{R}^{n \times 2d} \quad \text{and} \quad \tilde{W}' = [\tilde{U}'|\tilde{V}'] \in \mathbb{R}^{n \times 2d}.$$

### A.2 Misclustering rate for general graphs

The proof of Theorem 4.1 relies on Proposition 1, whose proof is given below.

*Proof.* By Hoeffding's inequality, for any fixed  $i, j$ ,

$$\mathbb{P} \left( |(\tilde{A}\tilde{A}^T)_{uv} - \mathbb{E}((\tilde{A}\tilde{A}^T)_{uv})| \geq t \mid \mathcal{F} \right) \leq 2 \exp \left( -\frac{2t^2}{N^5} \right)$$

since we define  $\tilde{A}_{uv} = \sum_{w=1}^N A_{uw}A_{wv}$ , whose range is  $0 \leq \tilde{A}_{uv} \leq N$ . Let  $t^2 = 2N^5 \log N$ . Then,

$$\mathbb{P} \left( |(\tilde{A}\tilde{A}^T)_{uv} - \mathbb{E}((\tilde{A}\tilde{A}^T)_{uv})|^2 \geq 2N^5 \log N \mid \mathcal{F} \right) \leq 2N^{-4}.$$

By the union bound over all  $(u, v)$ , we have

$$\mathbb{P} \left( \|(\tilde{A}\tilde{A}^T) - \mathbb{E}(\tilde{A}\tilde{A}^T)\|_F^2 \geq 2N^7 \log N \right) \leq 2N^{-2}.$$

Therefore, with probability at least  $1 - 2N^{-2}$ , we have

$$\|\tilde{A}\tilde{A}^\top - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F \leq \sqrt{2N^7 \log N}. \tag{13}$$

□

**Lemma 2.** *It always holds that  $\sigma_{d+1}(\tilde{Q}) = 0$  and  $\sigma_1(\tilde{Q}) \leq N^2$ . It almost always holds that  $(\min_i \pi_i) \cdot \tilde{b}N^2 \leq \sigma_d(\tilde{Q})$*

*Proof.* •  $\sigma_1(\tilde{Q}) \leq N^2$

Since  $Q \in [0, 1]^{N \times N}$ ,  $\tilde{Q} = QQ \in [0, N]^{N \times N}$ . The entries of the nonnegative matrix  $\tilde{Q}\tilde{Q}^T$  is bounded by  $N^3$  as follows:

$$(\tilde{Q}\tilde{Q}^T)_{ij} = \sum_{k=1}^N \tilde{Q}_{ik}\tilde{Q}_{jk} \leq N^3.$$

Then, the sums of the rows are bounded by  $N^4$ . Therefore, we have

$$\begin{aligned} \sigma_1^2(\tilde{Q}) &= \lambda_1(\tilde{Q}\tilde{Q}^T) \leq N^4 \\ &\therefore \sigma_1(\tilde{Q}) \leq N^2. \end{aligned}$$

- $\sigma_{d+1}(\tilde{Q}) = 0$

Since the rank of the matrix  $Q$  is at most  $d$ , the rank of  $\tilde{Q}$  is at most  $d$  as well.

- $\sigma_d(\tilde{Q}) \geq N \cdot (\min_i \pi_i) \cdot \lambda_d(\tilde{B}) = \tilde{\pi} \cdot \tilde{b} N^2$

Since  $Q = ZBZ^\top$ , we have

$$\begin{aligned}\tilde{Q} &= QQ = ZPZ^\top ZPZ^\top \\ &= ZB_{\theta_i \theta_j} D_n BZ^\top \quad (\because Z^\top Z = D_n = \text{diag}(n_1, \dots, n_d) \in \mathbb{R}^{d \times d}) \\ &= Z\tilde{B}Z^\top \quad (\because \tilde{B} = BD_n B \in \mathbb{R}^{d \times d}) \\ &= Z\tilde{L}\tilde{\Lambda}\tilde{R}^\top Z^\top \quad (\because \tilde{B} = \tilde{L}\tilde{\Lambda}\tilde{R}^\top, \tilde{L}, \tilde{R} \in \mathbb{R}^{d \times d})\end{aligned}$$

Let us define  $\tilde{X} = Z\tilde{L}\tilde{\Lambda}^{1/2}$ ,  $\tilde{Y} = Z\tilde{R}\tilde{\Lambda}^{1/2}$ . Then, we have

$$\tilde{Q} = \tilde{X}\tilde{Y}^\top.$$

The nonzero eigenvalues of  $(\tilde{X}\tilde{Y}^\top)(\tilde{X}\tilde{Y}^\top)^\top$  are the same as the nonzero eigenvalues of  $\tilde{Y}^\top\tilde{Y}\tilde{X}\tilde{X}^\top$ . For the nonzero smallest eigenvalue of  $\tilde{Y}^\top\tilde{Y}$  is obtained as follows:

$$\begin{aligned}\tilde{Y}^\top\tilde{Y} &= \tilde{\Lambda}^{1/2}\tilde{R}^\top Z^\top Z\tilde{R}\tilde{\Lambda}^{1/2} \\ &= \tilde{\Lambda}^{1/2}\tilde{R}^\top D_n \tilde{R}\tilde{\Lambda}^{1/2}.\end{aligned}$$

By Rayleigh quotient,

$$x^\top \tilde{R}^\top D_n \tilde{R} x = (\tilde{R}x)^\top D_n \tilde{R} x.$$

Let  $y = \tilde{R}x$ . Then,  $\|y\|^2 = \|\tilde{R}x\|^2 = \|x\|^2$ . Then,

$$\begin{aligned}(\tilde{R}x)^\top D_n \tilde{R} x &= y^\top D_n y \\ &= \sum_{k=1}^d n_k \cdot y_k^2 \\ &\geq N \cdot \min_{i \in \{1, \dots, d\}} \pi_i \cdot \|y\|^2 \\ &= N\tilde{\pi}\|y\|^2 \quad (\because \tilde{\pi} = \min_{i \in \{1, \dots, d\}} \pi_i)\end{aligned}$$

Let  $S = \tilde{R}^\top D_n \tilde{R}$ . Then,  $\tilde{Y}^\top\tilde{Y} = \tilde{\Lambda}^{1/2}S\tilde{\Lambda}^{1/2}$ .

$$\begin{aligned}\lambda_d(\tilde{Y}^\top\tilde{Y}) &= \lambda_d(\tilde{\Lambda}^{1/2}S\tilde{\Lambda}^{1/2}) \\ &= \min_{\|z\|=1} z^\top \tilde{\Lambda}^{1/2} S \tilde{\Lambda}^{1/2} z \\ &\geq \lambda_d(S) \cdot \min_{\|z\|=1} \|\tilde{\Lambda}^{1/2} z\|^2 \\ &\geq N \cdot \tilde{\pi} \cdot \lambda_d(\tilde{\Lambda}) \\ &\geq N \cdot \tilde{\pi} \cdot \lambda_d(\tilde{B})\end{aligned}$$

Similarly, we have  $\lambda_d(\tilde{X}^\top\tilde{X}) \geq N \cdot \tilde{\pi} \cdot \lambda_d(\tilde{B})$ .

By the eigenvalue inequality (Zhang et al. 2006),

$$\begin{aligned}\lambda_d(\tilde{Y}^\top\tilde{Y}\tilde{X}\tilde{X}^\top) &\geq \lambda_d(\tilde{Y}^\top\tilde{Y})\lambda_d(\tilde{X}^\top\tilde{X}) \\ &\geq (N \cdot \tilde{\pi} \cdot \lambda_d(\tilde{B}))^2\end{aligned}$$

Here, we can write that  $\lambda_d(\tilde{B}) = \tilde{b}N$ , where  $\tilde{b}(> 0) \in \mathbb{R}$ . Therefore, we have

$$\sigma_d^2(\tilde{Q}) \geq \left( \tilde{\pi} \cdot \tilde{b}N^2 \right)^2,$$

where  $\tilde{\pi} = \min_{i \in \{1, \dots, d\}} \pi_i$ .

□

**Lemma 3.** *It almost always holds that there exists an orthogonal matrix  $\tilde{O} \in \mathbb{R}^{2d \times 2d}$  such that*

$$\|\tilde{W}\tilde{O} - \tilde{W}'\|_F \leq \frac{2\sqrt{2}}{\sigma_d^2(\tilde{Q})} \sqrt{N^7 \log N} \leq \frac{2\sqrt{2}}{\tilde{\pi}^2 \tilde{b}^2} \sqrt{\frac{\log N}{N}}. \quad (14)$$

*Proof.* Let  $\mathcal{S} = (\frac{1}{2}\sigma_d^2(\tilde{Q}), \infty)$ . It almost always holds that exactly  $d$  eigenvalues of  $\tilde{A}\tilde{A}^\top$  and  $\mathbb{E}(\tilde{A}\tilde{A}^\top)$  are in  $\mathcal{S}$ . Additionally, the minimum distance  $\delta$  between  $\tilde{U}$  in  $\mathcal{S}$  and any eigenvalue of  $\tilde{U}'$  not in  $\mathcal{S}$  is  $\delta > \sigma_d^2(\tilde{Q})$ . Therefore, by the Davis-Kahan Theorem, there exists an orthogonal matrix  $O_1 \in \mathbb{R}^{d \times d}$  such that

$$\|\tilde{U}\tilde{O}_1 - \tilde{U}'\|_F \leq \frac{\sqrt{2}}{\delta} \|\tilde{A}\tilde{A}^\top - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F \leq \frac{2\sqrt{N^7 \log N}}{\sigma_d^2(\tilde{Q})},$$

where  $\tilde{O}_1 \in \mathbb{R}^{d \times d}$ . Similarly, for  $\tilde{A}^\top \tilde{A}$ , we have

$$\|\tilde{V}\tilde{O}_2 - \tilde{V}'\|_F \leq \frac{2\sqrt{N^7 \log N}}{\sigma_d^2(\tilde{Q})},$$

where  $\tilde{O}_2 \in \mathbb{R}^{d \times d}$ .

By direct sum of  $\tilde{O}_1$  and  $\tilde{O}_2$ , taking  $\tilde{O} \in \mathbb{R}^{2d \times 2d}$  such that

$$\begin{aligned} \|\tilde{W}\tilde{O} - \tilde{W}'\|_F &= \sqrt{\|\tilde{U}\tilde{O}_1 - \tilde{U}'\|_F^2 + \|\tilde{V}\tilde{O}_2 - \tilde{V}'\|_F^2} \\ &\leq \frac{2\sqrt{2}\sqrt{N^7 \log N}}{\sigma_d^2(\tilde{Q})} \end{aligned}$$

Since  $\sigma_d^2(\tilde{Q}) \geq \left( \tilde{\pi} \tilde{b} N^2 \right)^2$ ,

$$\|\tilde{W}\tilde{O} - \tilde{W}'\|_F \leq \frac{2\sqrt{2}}{\tilde{\pi}^2 \tilde{b}^2} \sqrt{\frac{\log N}{N}}$$

□

**Lemma 4.** *For all  $u, v$  such that  $\tilde{X}_u \neq \tilde{X}_v$ ,  $\|\tilde{U}_u - \tilde{U}_v\| \geq \hat{\beta} \sqrt{\tilde{b} \tilde{\pi}}$ . Similarly, for all  $\tilde{Y}_u \neq \tilde{Y}_v$ ,  $\|\tilde{V}_u - \tilde{V}_v\| \geq \hat{\beta} \sqrt{\tilde{b} \tilde{\pi}}$ . Therefore,*

$$\|\tilde{W}_u - \tilde{W}_v\| \geq \hat{\beta} \frac{\sqrt{\sigma_d(\tilde{B}) \cdot \tilde{\pi} N}}{\sigma_1(\tilde{Q})} \geq \hat{\beta} \sqrt{\tilde{b} \tilde{\pi}}. \quad (15)$$

*Proof.* Define  $\tilde{E} = \tilde{\Lambda}^{1/2} \tilde{R}^\top \in \mathbb{R}^{d \times d}$ ,  $\tilde{G} = \tilde{X}\tilde{E} \in \mathbb{R}^{N \times d}$ , and  $\tilde{G}' = \tilde{G}D_n^{1/2} \in \mathbb{R}^{N \times d}$ . Since  $Q = ZBZ^\top$  and  $B \in [0, 1]^{d \times d}$  with  $\text{rank}(P) = d$ , we have

$$\tilde{Q} = \tilde{X}\tilde{Y}^\top$$

Therefore, we have

$$\begin{aligned}
\tilde{Q}\tilde{Q}^\top &= \left(\tilde{X}\tilde{Y}^\top\right)\left(\tilde{X}\tilde{Y}^\top\right)^T \\
&= \tilde{X}\tilde{Y}^\top\tilde{Y}\tilde{X}^\top \\
&= \tilde{X}\left(\tilde{\Lambda}^{1/2}\tilde{R}^\top Z^\top Z\tilde{R}\tilde{\Lambda}^{1/2}\right)\tilde{X}^\top \\
&= \tilde{X}\left(\tilde{\Lambda}^{1/2}\tilde{R}^\top\right)D_n\left(\tilde{R}\tilde{\Lambda}^{1/2}\right)\tilde{X}^\top \\
&= \tilde{X}\tilde{E}D_n\tilde{E}^\top\tilde{X}^\top \quad (\because \tilde{E} = \tilde{\Lambda}^{1/2}\tilde{R}^\top) \\
&= \tilde{G}D_n\tilde{G}^\top \quad (\because \tilde{G} = \tilde{X}\tilde{E}) \\
&= \tilde{G}'\tilde{G}'^\top \quad (\because \tilde{G}' = \tilde{G}D_n^{1/2})
\end{aligned}$$

Moreover,

$$\begin{aligned}
\tilde{Q}\tilde{Q}^\top &= \tilde{U}\tilde{\Sigma}\tilde{V}^\top\tilde{V}\tilde{\Sigma}\tilde{U}^\top \quad (\because \tilde{Q} = \tilde{U}\tilde{\Sigma}\tilde{V}^\top) \\
&= \tilde{U}\tilde{\Sigma}^2\tilde{U}^\top \\
&= \tilde{U}'\tilde{U}'^\top \quad (\because \tilde{U}' = \tilde{U}\tilde{\Sigma})
\end{aligned}$$

Thus, we have

$$\tilde{U}'\tilde{U}'^\top = \tilde{Q}\tilde{Q}^\top = \tilde{G}'\tilde{G}'^\top.$$

Let  $e \in \mathbb{R}^N$  be the vector whose  $u$ -th entry is 1, whose  $v$ -th entry is -1, and whose remaining entries are all zero. Then, we have

$$\begin{aligned}
\|\tilde{G}'_u - \tilde{G}'_v\|^2 &= e^\top \tilde{G}'\tilde{G}'^\top e \\
&= e^\top \tilde{U}'\tilde{U}'^\top e \\
&= \|\tilde{U}'_u - \tilde{U}'_v\|^2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\tilde{\beta} &\leq \|\tilde{X}_u - \tilde{X}_v\| \\
&\leq \frac{1}{\sqrt{\sigma_d(\tilde{B})}}\|\tilde{G}_u - \tilde{G}_v\| \quad (\because \tilde{G} = \tilde{X}\tilde{\Lambda}^{1/2}\tilde{R}^\top) \\
&\leq \frac{1}{\sqrt{\sigma_d(\tilde{B}) \cdot \min_{i \in \{1, \dots, d\}} n_i}}\|\tilde{G}'_u - \tilde{G}'_v\| \quad (\because \tilde{G}' = \tilde{G}D_n^{1/2}) \\
&= \frac{1}{\sqrt{\sigma_d(\tilde{B}) \cdot \tilde{\pi}N}}\|\tilde{U}'_u - \tilde{U}'_v\| \quad (\because \min_{i \in \{1, \dots, d\} n_i} = N\tilde{\pi}) \\
&\leq \frac{\sigma_1(\tilde{Q})}{\sqrt{\sigma_d(\tilde{B}) \cdot \tilde{\pi}N}}\|\tilde{U}_u - \tilde{U}_v\| \quad (\because \tilde{U}' = \tilde{U}\tilde{\Sigma})
\end{aligned}$$

where the group  $i$ -th size  $n_i \in \mathbb{R}$  with  $i \in \{1, \dots, d\}$ . Therefore, we finally have

$$\begin{aligned}
\|\tilde{U}_u - \tilde{U}_v\| &\geq \tilde{\beta} \frac{\sqrt{\sigma_d(\tilde{B}) \cdot \tilde{\pi}N}}{\sigma_1(\tilde{Q})} \\
&\geq \hat{\beta}N \frac{\sqrt{\tilde{b}\tilde{\pi}N^2}}{N^2} \\
&= \hat{\beta}\sqrt{\tilde{b}\tilde{\pi}}.
\end{aligned}$$

□

Finally, the proof of Theorem 4.1 is as follows.

*Proof.*

Let  $\hat{\psi}$  and  $\hat{\tau}$  satisfy the clustering criterion for  $\tilde{W}$  (where  $\tilde{W} = [\tilde{U}|\tilde{V}]$  takes the role of  $Z$  for Equation (4) in (Sussman et al. 2012)) Let  $C' \in \mathbb{R}^{N \times 2d}$  have row  $u$  given by  $C'_u = \hat{\psi}_{\tau(u)}$ . From Equation (5), we can say that

$$\|C' - \tilde{W}\|_F \leq \|\tilde{W}\tilde{O} - \tilde{W}'\|_F.$$

Therefore, we have

$$\begin{aligned} \|C' - \tilde{W}\tilde{O}\|_F &= \|C' - \tilde{W}' + \tilde{W}' - \tilde{W}\tilde{O}\|_F \\ &\leq \|C' - \tilde{W}'\|_F + \|\tilde{W}' - \tilde{W}\tilde{O}\|_F \\ &\leq \|\tilde{W}\tilde{O} - \tilde{W}'\|_F + \|\tilde{W}' - \tilde{W}\tilde{O}\|_F \quad (\because \text{Equation (5)}) \\ &= 2 \times \|\tilde{W}\tilde{O} - \tilde{W}'\|_F \\ &\leq 2 \times \frac{2\sqrt{2}\sqrt{N^7 \log N}}{\sigma_d^2(\tilde{Q})} \end{aligned}$$

Let  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_K$  be balls of radius

$$r = \frac{\tilde{\beta}}{3} \cdot \frac{\sqrt{\sigma_d(\tilde{B}) \cdot \tilde{\pi}N}}{\sigma_1(\tilde{Q})}$$

each centered around the  $K$  distinct rows of  $\tilde{W}$ . By Lemma 3, these balls are almost always disjoint.

We want to find the number of misclustering nodes, i.e.,

$$\|C'_u - \tilde{W}_u\tilde{O}\| > r.$$

To find the number of  $u$ -th rows to satisfy this inequality. Since the norm we use, we have

$$\|C' - \tilde{W}\tilde{O}\|_F = \sqrt{\sum_u \|C'_u - \tilde{W}_u\tilde{O}\|^2}.$$

Therefore, we have

$$\begin{aligned} \sqrt{\sum_u \|C'_u - \tilde{W}_u\tilde{O}\|^2} &= \|C' - \tilde{W}\tilde{O}\|_F \leq \frac{2^{5/2}\sqrt{N^7 \log N}}{\sigma_d^2(\tilde{Q})}. \\ \Rightarrow \sum_u \|C'_u - \tilde{W}_u\tilde{O}\|^2 &\leq \frac{2^5 N^7 \log N}{\sigma_d^4(\tilde{Q})}. \end{aligned}$$

Let  $M = \{u : \|C'_u - \tilde{W}_u\tilde{O}\| > r\}$  be a set of the misclustered nodes, which is the same idea of the expression

$$\min_{\rho \in S_K} |\{u \in V : \tau(u) \neq \rho(\hat{\tau}(u))\}|.$$

Then, we have

$$\sum_{u \in M} \|C'_u - \tilde{W}_u\tilde{O}\|^2 \geq |M|r^2.$$

Therefore,

$$|M|r^2 \leq \sum_{u \in M} \|C'_u - \tilde{W}_u\tilde{O}\|^2 \leq \sum_u \|C'_u - \tilde{W}_u\tilde{O}\|^2 \leq \frac{2^5 \cdot N^7 \log N}{\sigma_d^4(\tilde{Q})}.$$

Hence, by substituting

$$r = \frac{\tilde{\beta}}{3} \cdot \frac{\sqrt{\sigma_d(\tilde{B})\tilde{\pi}N}}{\sigma_1(\tilde{Q})},$$

the number of the misclustered nodes is bounded as follows:

$$\begin{aligned} |M| &\leq \frac{2^5 3^2 \sigma_1^2(\tilde{Q}) N^7 \log N}{\tilde{\beta}^2 \sigma_d(\tilde{B})(\tilde{\pi}N) \sigma_d^4(\tilde{Q})} \\ &\leq \frac{2^5 3^2 N^4 N^7 \log N}{(\hat{\beta}N)^2 (\tilde{b}N)(\tilde{\pi}N)(\tilde{\pi}\tilde{b}N^2)^4} \quad (\because \sigma_1(\tilde{Q}) \leq N^2, \tilde{\beta} = \hat{\beta}N, \sigma_d(\tilde{B}) = \tilde{b}N, \sigma_d(\tilde{Q}) \geq (\tilde{\pi}\tilde{b}N^2)) \\ &= \frac{2^5 3^2}{\hat{\beta}^2 (\tilde{b}\tilde{\pi})^5} \cdot \frac{\log N}{N}, \end{aligned}$$

where constants  $\hat{b}, \tilde{b}, \tilde{\pi} > 0$  do not depend on the network size  $N$ .  $\square$

### A.3 Misclustered nodes for core-periphery case

In this section, we focus on the consistency of the core-periphery structure. Let the matrix  $Q$  be

$$Q = \begin{bmatrix} pJ_{n_1 \times n_1} & qJ_{n_1 \times n_2} \\ rJ_{n_2 \times n_1} & sJ_{n_2 \times n_2} \end{bmatrix}, \quad (16)$$

where  $J_{n,m}$  is a matrix  $n \times m$  whose entries are all equal to 1. So, we define

$$\tilde{Q} = QQ = \begin{bmatrix} (n_1p^2 + n_2qr) J_{n_1 \times n_1} & (n_1pq + n_2qs) J_{n_1 \times n_2} \\ (n_1pr + n_2rs) J_{n_2 \times n_1} & (n_1qr + n_2s^2) J_{n_2 \times n_2} \end{bmatrix} =: \begin{bmatrix} \tilde{Q}_{11} J_{n_1 \times n_1} & \tilde{Q}_{12} J_{n_1 \times n_2} \\ \tilde{Q}_{21} J_{n_2 \times n_1} & \tilde{Q}_{22} J_{n_2 \times n_2} \end{bmatrix}$$

and

$$\tilde{Q}\tilde{Q}^\top = \begin{bmatrix} (n_1\tilde{Q}_{11}^2 + n_2\tilde{Q}_{12}^2) J_{n_1 \times n_1} & (n_1\tilde{Q}_{11}\tilde{Q}_{21} + n_2\tilde{Q}_{12}\tilde{Q}_{22}) J_{n_1 \times n_2} \\ (n_1\tilde{Q}_{21}\tilde{Q}_{11} + n_2\tilde{Q}_{22}\tilde{Q}_{12}) J_{n_2 \times n_1} & (n_1\tilde{Q}_{21}^2 + n_2\tilde{Q}_{22}^2) J_{n_2 \times n_2} \end{bmatrix}.$$

Here, for the core-periphery structure, we assume that

$$0 \leq s < q, r < p \leq 1.$$

#### A.3.1 Proof of Theorem 4.2

Here, we mainly focus on the proof of Theorem 4.2, which is the DASE case. Here, in order to consider the Hoeffding's inequality, we need to express the bound for  $\|\tilde{A}\tilde{A}^\top - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|$ . First, we have

$$\mathbb{E}((\tilde{A}\tilde{A}^\top)_{uv}) = \mathbb{E}\left(\sum_{w=1}^N \tilde{A}_{uw}\tilde{A}_{vw}\right).$$

Since we know the bound for  $\tilde{Q}$  and  $\tilde{Q}\tilde{Q}^\top$ , we want to express  $\mathbb{E}(\tilde{A}_{uw}\tilde{A}_{vw})$  in a form of  $\tilde{Q}$  as follows:

$$\mathbb{E}(\tilde{A}_{uw}\tilde{A}_{vw}) = \tilde{Q}_{uw}\tilde{Q}_{vw} + \text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw})$$

**Lemma 5.** For all  $u, v, w \in \{1, \dots, N\}$ ,

$$\text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}) \geq 0,$$

where  $\tilde{A} = AA$  and entries of  $A$  are independent Bernoulli random variables.

*Proof.*

$$\text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}) = \mathbb{E}(\tilde{A}_{uw}\tilde{A}_{vw}) - \mathbb{E}(\tilde{A}_{uw})\mathbb{E}(\tilde{A}_{vw}).$$

Here, we can write

$$\mathbb{E}(\tilde{A}_{uw}\tilde{A}_{vw}) = \mathbb{E}\left(\sum_{l=1}^N \sum_{h=1}^N A_{ul}A_{lw}A_{vh}A_{hw}\right)$$

If  $l \neq h$ ,

$$\mathbb{E}(\tilde{A}_{uw}\tilde{A}_{vw}) = \mathbb{E}\left(\sum_{l=1}^N \sum_{h=1}^N A_{ul}A_{lw}A_{vh}A_{hw}\right) = \left(\sum_{l=1}^N Q_{ul}Q_{lw}\right) \left(\sum_{h=1}^N Q_{vh}Q_{hw}\right).$$

Therefore,

$$\text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}) = 0.$$

On the other hand, if  $l = h$ ,

$$\mathbb{E}(\tilde{A}_{uw}\tilde{A}_{vw}) = \mathbb{E}\left(\sum_{l=1}^N A_{ul}A_{lw}^2A_{vl}\right) = \sum_{l=1}^N Q_{ul}Q_{lw}Q_{vl}.$$

Therefore,

$$\begin{aligned} \text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}) &= \sum_{l=1}^N Q_{ul}Q_{lw}Q_{vl} - \sum_{l=1}^N Q_{ul}Q_{lw}^2Q_{vl} \\ &= \sum_{l=1}^N Q_{ul}Q_{vl}Q_{lw}(1 - Q_{lw}) \\ &\geq 0. \end{aligned}$$

□

**Lemma 6.** For  $u, v \in \{1, \dots, N\}$  and  $0 \leq s < r, q < p \leq 1$ ,

$$|(\tilde{A}\tilde{A}^\top)_{uv} - \mathbb{E}(\tilde{A}\tilde{A}^\top)_{uv}| \leq N^2 \left[ \sum_{w \in G_1} (1 - (\pi_1 pr + \pi_2 rs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2) \right],$$

where  $G_1$  and  $G_2$  are the core and periphery groups, respectively.

*Proof.* Since  $(\tilde{A}\tilde{A}^\top)_{uv} = \sum_{w=1}^N \tilde{A}_{uw}\tilde{A}_{vw}$ ,

$$(\tilde{A}\tilde{A}^\top)_{uv} \in [0, N^3].$$

Since  $\text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}) \geq 0$ ,

$$\begin{aligned} \tilde{Q}_{uw}\tilde{Q}_{vw} &\leq \mathbb{E}(\tilde{A}_{uw}\tilde{A}_{vw}) = \tilde{Q}_{uw}\tilde{Q}_{vw} + \text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}) \\ \Rightarrow (\tilde{Q}\tilde{Q}^\top)_{uv} &\leq \mathbb{E}((\tilde{A}\tilde{A}^\top)_{uv}) = (\tilde{Q}\tilde{Q}^\top)_{uv} + \sum_{w=1}^N \text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}). \end{aligned}$$

By

$$(\tilde{Q}\tilde{Q}^\top)_{uv} \in [n_1\tilde{Q}_{21}^2 + n_2\tilde{Q}_{22}^2, n_1\tilde{Q}_{11}^2 + n_2\tilde{Q}_{12}^2],$$

we have

$$-(n_1\tilde{Q}_{11}^2 + n_2\tilde{Q}_{12}^2 + N^2p^3(1-p)) \leq (\tilde{A}\tilde{A}^\top)_{uv} - \mathbb{E}(\tilde{A}\tilde{A}^\top)_{uv} \leq N^3 - (n_1\tilde{Q}_{21}^2 + n_2\tilde{Q}_{22}^2),$$

since  $\text{Cov}(\tilde{A}_{uw}, \tilde{A}_{vw}) \in [Ns^3(1-s), Np^3(1-p)]$ . Therefore,

$$\begin{aligned}
|(\tilde{A}\tilde{A}^\top)_{uv} - \mathbb{E}(\tilde{A}\tilde{A}^\top)_{uv}| &\leq N^3 - (n_1\tilde{Q}_{21}^2 + n_2\tilde{Q}_{22}^2) \\
&= N^3 - \left( \sum_{w \in G_1} (n_1pr + n_2rs)^2 + \sum_{w \in G_2} (n_1qr + n_2s^2)^2 \right) \\
&= N^3 - N^2 \left( \sum_{w \in G_1} (\pi_1pr + \pi_2rs)^2 + \sum_{w \in G_2} (\pi_1qr + \pi_2s^2)^2 \right) \quad (\because n_1 = \pi_1N, n_2 = \pi_2N) \\
&= N^2 \left[ \sum_{w \in G_1} (1 - (\pi_1pr + \pi_2rs)^2) + \sum_{w \in G_2} (1 - (\pi_1qr + \pi_2s^2)^2) \right] \quad (N = n_1 + n_2),
\end{aligned}$$

where  $G_1$  and  $G_2$  are clusters 1 and 2, which refer to core and periphery groups, respectively.  $\square$

As in the general graph setting, we begin by proving Proposition 2 for core-periphery graphs.

*Proof.* By Lemma 6, we have

$$|(\tilde{A}\tilde{A}^\top)_{uv} - \mathbb{E}(\tilde{A}\tilde{A}^\top)_{uv}| \leq N^2 \left[ \sum_{w \in G_1} (1 - (\pi_1pr + \pi_2rs)^2) + \sum_{w \in G_2} (1 - (\pi_1qr + \pi_2s^2)^2) \right],$$

where  $G_1$  and  $G_2$  are the core and periphery groups, respectively.

For applying to Hoeffding's inequality,

$$\mathbb{P}(|(\tilde{A}\tilde{A}^\top)_{uv} - \mathbb{E}(\tilde{A}\tilde{A}^\top)_{uv}| \geq t) \leq 2 \exp\left(-\frac{2t^2}{N^5}\right).$$

Let  $t = \sqrt{2}\tilde{T}_1\sqrt{N^3 \log N}$ , where

$$\tilde{T}_1 = \left[ \sum_{w \in G_1} (1 - (\pi_1pr + \pi_2rs)^2) + \sum_{w \in G_2} (1 - (\pi_1qr + \pi_2s^2)^2) \right].$$

By Hoeffding's inequality,

$$\mathbb{P}(|(\tilde{A}\tilde{A}^\top)_{uv} - \mathbb{E}(\tilde{A}\tilde{A}^\top)_{uv}|^2 \geq 2\tilde{T}_1^2 N^3 \log N) \leq 2N^{-\frac{4}{N^2}\tilde{T}_1^2}.$$

By the union bound over all  $(u, v)$ , we have

$$\mathbb{P}\left(\|(\tilde{A}\tilde{A}^\top) - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F^2 \geq 2\tilde{T}_1^2 N^5 \log N\right) \leq 2N^{2-\frac{4}{N^2}\tilde{T}_1^2}.$$

Therefore, with probability at least  $1 - 2N^{2-\frac{4}{N^2}\tilde{T}_1^2}$ , we have

$$\begin{aligned}
\|(\tilde{A}\tilde{A}^\top) - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F &\leq \sqrt{2}\tilde{T}_1\sqrt{N^5 \log N} \\
&= \sqrt{2} \left[ \sum_{w \in G_1} (1 - (\pi_1pr + \pi_2rs)^2) + \sum_{w \in G_2} (1 - (\pi_1qr + \pi_2s^2)^2) \right] \sqrt{N^5 \log N}.
\end{aligned}$$

Similarly, for  $\|\tilde{A}^\top\tilde{A} - \mathbb{E}(\tilde{A}^\top\tilde{A})\|_F$ , we have

$$\mathbb{P}\left(\|(\tilde{A}\tilde{A}^\top) - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F^2 \geq 2\tilde{T}_2^2 N^5 \log N\right) \leq 2N^{2-\frac{4}{N^2}\tilde{T}_2^2},$$

where

$$\tilde{T}_2 = \left[ \sum_{w \in G_1} (1 - (\pi_1 pq + \pi_2 qs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2) \right].$$

Therefore, with probability at least  $1 - 2N^{2-\frac{4}{N^2}[\sum_{w \in G_1}(1-(\pi_1 pq + \pi_2 qs)^2) + \sum_{w \in G_2}(1-(\pi_1 qr + \pi_2 s^2)^2)]^2}$ ,

$$\|\tilde{A}^\top \tilde{A} - \mathbb{E}(\tilde{A}^\top \tilde{A})\|_F \leq \sqrt{2} \left[ \sum_{w \in G_1} (1 - (\pi_1 pq + \pi_2 qs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2) \right] \sqrt{N^5 \log N}. \quad (17)$$

□

When we assume the core-periphery structure, we can get the singular values of  $\tilde{Q}$  and  $\tilde{B}$  directly.

**Lemma 7.** *Let us define  $Q \in [0, 1]^{N \times N}$  in Equation (9), and  $\text{rank}(Q) = d = 2$ . For singular values of  $\tilde{Q} \in [0, 1]^{N \times N}$ ,*

$$\sigma_{1,2}^2(\tilde{Q}) = \frac{\text{diag}(\tilde{Q}\tilde{Q}^\top) \pm \sqrt{\text{diag}^2(\tilde{Q}\tilde{Q}^\top) - 4\det(\tilde{Q}\tilde{Q}^\top)}}{2}.$$

For singular values of  $\tilde{B} \in \mathbb{R}^{K \times K}$ ,

$$\sigma_{1,2}^2(\tilde{B}) = \frac{\text{diag}(\tilde{B}\tilde{B}^\top) \pm \sqrt{\text{diag}^2(\tilde{B}\tilde{B}^\top) - 4\det(\tilde{B}\tilde{B}^\top)}}{2}.$$

**Lemma 8.** *It almost always holds that there exists an orthogonal matrix  $\tilde{O} \in \mathbb{R}^{2d \times 2d}$  such that*

$$\|\tilde{W}\tilde{O} - \tilde{W}'\|_F \leq \frac{2\sqrt{\tilde{T}_1^2 + \tilde{T}_2^2}}{\sigma_2^2(\tilde{Q})} \sqrt{N^5 \log N} \leq \frac{2\sqrt{\tilde{T}_1^2 + \tilde{T}_2^2}}{\tilde{\pi}^2 \tilde{b}^2} \sqrt{\frac{\log N}{N^3}}. \quad (18)$$

*Proof.* Similar to Lemma 3,  $\|\tilde{U}\tilde{O}_1 - \tilde{U}'\|_F$  is bounded as

$$\|\tilde{U}\tilde{O}_1 - \tilde{U}'\|_F \leq \frac{\sqrt{2}}{\sigma_2^2(\tilde{Q})} \|\tilde{A}\tilde{A}^\top - \mathbb{E}(\tilde{A}\tilde{A}^\top)\|_F \leq \frac{2\tilde{T}_1}{\sigma_2^2(\tilde{Q})} \sqrt{N^5 \log N},$$

where  $\tilde{T}_1 = [\sum_{w \in G_1} (1 - (\pi_1 pr + \pi_2 rs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2)]$  and  $\tilde{O}_1 \in \mathbb{R}^{d \times d}$ . For  $\|\tilde{V}\tilde{O}_2 - \tilde{V}'\|_F$ ,

$$\|\tilde{V}\tilde{O}_2 - \tilde{V}'\|_F \leq \frac{\sqrt{2}}{\sigma_2^2(\tilde{Q})} \|\tilde{A}^\top \tilde{A} - \mathbb{E}(\tilde{A}^\top \tilde{A})\|_F \leq \frac{2\tilde{T}_2}{\sigma_2^2(\tilde{Q})} \sqrt{N^5 \log N},$$

where  $\tilde{T}_2 = [\sum_{w \in G_1} (1 - (\pi_1 pq + \pi_2 qs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2)]$  and  $\tilde{O}_2 \in \mathbb{R}^{d \times d}$ . Therefore, by the direct sum, we have

$$\|\tilde{W}\tilde{O} - \tilde{W}'\|_F \leq \frac{2\sqrt{\tilde{T}_1^2 + \tilde{T}_2^2}}{\sigma_2^2(\tilde{Q})} \sqrt{N^5 \log N},$$

where  $\tilde{O} \in \mathbb{R}^{2d \times 2d}$ .

□

Therefore, the proof of Theorem 4.2 is as follows.

*Proof.* Similar to the proof of Theorem 4.1,

$$\sum_u \|C'_u - \tilde{W}_u \tilde{O}\|^2 \leq \frac{2^4(\tilde{T}_1^2 + \tilde{T}_2^2)}{\sigma_2^4(\tilde{Q})} N^5 \log N.$$

Let  $M = \{u : \|C'_u - \tilde{W}_u O\| > r\}$  be a set of the misclustered nodes, we have

$$\sum_{u \in M} \|C'_u - \tilde{W}_u \tilde{O}\|^2 \geq |M|r^2.$$

Therefore,

$$|M|r^2 \leq \sum_u \|C'_u - \tilde{W}_u \tilde{O}\|^2 \leq \frac{2^4(\tilde{T}_1^2 + \tilde{T}_2^2)}{\sigma_2^4(\tilde{Q})} N^5 \log N.$$

By substituting

$$r = \frac{\tilde{\beta}}{3} \frac{\sqrt{\sigma_2(\tilde{B}) \cdot \tilde{\pi}N}}{\sigma_1(\tilde{Q})},$$

the number of misclustered nodes is bounded as follows:

$$\begin{aligned} |M| &\leq \frac{2^4 3^2 (\tilde{T}_1^2 + \tilde{T}_2^2) \sigma_1^2(\tilde{Q}) \cdot N^5 \log N}{\tilde{\beta}^2 \sigma_2(\tilde{B})(\tilde{\pi}N) \sigma_2^4(\tilde{Q})} \\ &\leq \frac{2^4 3^2 (\tilde{T}_1^2 + \tilde{T}_2^2) (N^2)^2 \cdot N^5 \log N}{(\hat{\beta}N)^2 (\tilde{b}N)(\tilde{\pi}N)(\tilde{b}\tilde{\pi}N^2)^4} \quad (\because \sigma_1(\tilde{Q}) \leq N^2, \tilde{\beta} = \hat{\beta}N, \sigma_d(\tilde{B}) = \tilde{b}N, \sigma_d(\tilde{Q}) \geq (\tilde{\pi}\tilde{b}N^2)) \\ &= \frac{2^4 3^2 (\tilde{T}_1^2 + \tilde{T}_2^2) \cdot N^9 \log N}{\hat{\beta}^2 (\tilde{b}\tilde{\pi})^5 N^{12}}. \end{aligned}$$

Let us define

$$\begin{aligned} \tilde{T}N^2 &:= \tilde{T}_1^2 + \tilde{T}_2^2 \\ &= \left[ \sum_{w \in G_1} (1 - (\pi_1 pr + \pi_2 rs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2) \right]^2 \\ &\quad + \left[ \sum_{w \in G_1} (1 - (\pi_1 pq + \pi_2 qs)^2) + \sum_{w \in G_2} (1 - (\pi_1 qr + \pi_2 s^2)^2) \right]^2. \end{aligned}$$

Therefore, we have

$$|M| \leq \frac{2^4 3^2 \tilde{T}}{\hat{\beta}^2 (\tilde{b}\tilde{\pi})^5} \frac{\log N}{N},$$

where constants  $\hat{b}, \tilde{b}, \tilde{\pi}, \tilde{T} > 0$  are not dependent on the network size  $N$ .  $\square$

### A.3.2 Proof of Corollary 4.2.1

In this section, we focus on the proof of Corollary 4.2.1, which corresponds to the ASE case, for the purpose of comparing ASE and DASE. The proof follows the same structure as that of DASE and is based on the general ASE analysis in (Sussman et al. 2012).

**Proposition 3.** Let  $A \in \{0, 1\}^{N \times N}$  be an adjacency matrix of a directed graph. Let  $Q = \mathbb{E}(A)$  and the entries of  $Q$  is in a range  $[s, p]$ , where  $0 \leq s < p \leq 1$ . Suppose that  $A$  is conditionally independent given  $Q$ . Then, with probability at least  $1 - 2N^{2 - \frac{4}{N^2} [\sum_{w \in G_1} (1 - r^2) + \sum_{w \in G_2} (1 - s^2)]^2}$ ,

$$\|AA^\top - QQ^\top\|_F \leq \sqrt{3} \left[ \sum_{w \in G_1} (1 - r^2) + \sum_{w \in G_2} (1 - s^2) \right] \sqrt{N \log N}, \quad (19)$$

and with probability at least  $1 - 2N^{2 - \frac{4}{N^2} [\sum_{w \in G_1} (1 - q^2) + \sum_{w \in G_2} (1 - s^2)]^2}$ ,

$$\|A^\top A - Q^\top Q\|_F \leq \sqrt{3} \left[ \sum_{w \in G_1} (1 - q^2) + \sum_{w \in G_2} (1 - s^2) \right] \sqrt{N \log N}. \quad (20)$$

*Proof.* Let us consider  $\mathbb{E}((AA^\top)_{uv})$  for the cases  $u = v$  and  $u \neq v$ . When  $u = v$ ,

$$\mathbb{E}\left(\sum_{w=1}^N A_{uw}^2\right) = \sum_{w=1}^N Q_{uw}.$$

When  $u \neq v$ ,

$$\mathbb{E}\left(\sum_{w=1}^N A_{uw}A_{vw}\right) = \sum_{w=1}^N Q_{uw}Q_{vw}.$$

Since we assume that the matrix  $Q$  is defined as Equation 16, for  $u \neq v$ ,

$$QQ^\top = \begin{bmatrix} (n_1 p^2 + n_2 q^2) J_{n_1 \times n_1} & (n_1 p r + n_2 q s) J_{n_1 \times n_2} \\ (n_1 p r + n_2 q s) J_{n_2 \times n_1} & (n_1 r^2 + n_2 s^2) J_{n_2 \times n_2} \end{bmatrix}.$$

Since  $s < r, q < p$ ,

$$\begin{aligned} (AA^\top)_{ij} - (QQ^\top)_{ij} &\in [-(n_1 p^2 + n_2 q^2), N - (n_1 r^2 + n_2 s^2)] \\ \Rightarrow |(AA^\top)_{ij} - (QQ^\top)_{ij}| &\leq \sum_{w \in G_1} (1 - r^2) + \sum_{w \in G_2} (1 - s^2), \end{aligned}$$

where  $G_1$  and  $G_2$  are clusters 1 and 2, which refer to core group and periphery group, respectively.

Let us consider  $u \neq v$  first. By Hoeffding's inequality, let  $t = \frac{\sqrt{2}}{N} T_1 \sqrt{N \log N}$ , where

$$T_1 = \sum_{w \in G_1} (1 - r^2) + \sum_{w \in G_2} (1 - s^2),$$

we have

$$\mathbb{P}\left(|(AA^\top)_{uv} - (QQ^\top)_{uv}|^2 \geq \frac{2}{N} T_1^2 \log N\right) \leq 2N^{-\frac{4}{N^2} T_1^2}.$$

When  $u = v$ ,

$$|(AA^\top)_{uu} - (QQ^\top)_{uu}|^2 \leq T_1^2.$$

By the union bound over all  $(u, v)$ , we have

$$\mathbb{P}\left(\|(AA^\top) - (QQ^\top)\|_F^2 \geq 3T_1^2 N \log N\right) \leq 2N^{2-\frac{4}{N^2} T_1^2},$$

since  $2T_1^2 N \log N + T_1^2 N \leq 3T_1^2 N \log N$ . Therefore, with probability at least  $1 - 2N^{2-\frac{4}{N^2} T_1^2}$ ,

$$\|AA^\top - QQ^\top\|_F \leq \sqrt{3} T_1 \sqrt{N \log N}.$$

Similarly, for  $\|A^\top A - Q^\top Q\|_F$ , we have

$$\mathbb{P}\left(\|A^\top A - Q^\top Q\|_F^2 \geq 3T_2^2 N \log N\right) \leq 2N^{2-\frac{4}{N^2} T_2^2},$$

where

$$T_2 = \sum_{w \in G_1} (1 - q^2) + \sum_{w \in G_2} (1 - s^2).$$

Therefore, with probability at least  $1 - 2N^{2-\frac{4}{N^2} T_2^2}$ ,

$$\|A^\top A - Q^\top Q\|_F \leq \sqrt{3} T_2 \sqrt{N \log N}.$$

□

**Lemma 9.** Let us defind  $Q \in [0, 1]^{N \times N}$  in Equation (16), and  $\text{rank}(Q) = 2$ . For singular values of  $Q$ ,

$$\sigma_{1,2}^2(Q) = \frac{\text{diag}(QQ^\top) \pm \sqrt{\text{diag}^2(QQ^\top) - 4\det(QQ^\top)}}{2}.$$

For singular values of  $B \in \mathbb{R}^{K \times K}$ ,

$$\sigma_{1,2}^2(B) = \frac{\text{diag}(BB^\top) \pm \sqrt{\text{diag}^2(BB^\top) - 4\det(BB^\top)}}{2}.$$

**Lemma 10.** It almost always holds that there exists an orthogonal matrix  $O \in \mathbb{R}^{2d \times 2d}$  such that

$$\|\tilde{W}O - \tilde{W}'\|_F \leq \frac{2\sqrt{\tilde{T}_1^2 + \tilde{T}_2^2}}{\sigma_2^2(\tilde{Q})} \sqrt{N^5 \log N} \leq \frac{2\sqrt{\tilde{T}_1^2 + \tilde{T}_2^2}}{\tilde{\pi}^2 b^2} \sqrt{\frac{\log N}{N^3}}. \quad (21)$$

*Proof.* Similar to Lemma 8,  $\|UR_1 - U'\|$  is bounded as

$$\|UO_1 - U'\|_F \leq \frac{\sqrt{2}}{\sigma_2^2(Q)} \|AA^\top - QQ^\top\|_F \leq \frac{\sqrt{6}T_1}{\sigma_2^2(Q)} \sqrt{N \log N},$$

where  $O_1 \in \mathbb{R}^{d \times d}$ . For the bound  $\|VO_2 - V'\|_F$ ,

$$\|VO_2 - V'\|_F \leq \frac{\sqrt{2}}{\sigma_2^2(Q)} \|A^\top A - Q^\top Q\|_F \leq \frac{\sqrt{6}T_2}{\sigma_2^2(Q)} \sqrt{N \log N},$$

where  $O_2 \in \mathbb{R}^{d \times d}$ . Therefore, by the direct sum, there exists  $P \in \mathbb{R}^{2d \times 2d}$ , we have

$$\|WO - W'\|_F \leq \frac{\sqrt{6}\sqrt{T_1^2 + T_2^2}}{\sigma_2^2(Q)} \sqrt{N \log N}.$$

□

**Lemma 11.** By Lemma 3 in (Sussman et al. 2012), we have

$$\|W_u - W_v\| \geq \beta\sqrt{b\tilde{\pi}}N^{-1/2}$$

for all  $u, v$  such that  $\tau(u) \neq \tau(v)$ .

*Proof.* Proof in (Sussman et al. 2012). Since we define  $\tilde{\pi} = \min_{i \in [K]} \pi_i$  and  $b = \lambda_d(B)$ , the result of (Sussman et al. 2012) is equivalent to

$$\|W_u - W_v\| \geq \beta\sqrt{b\tilde{\pi}}N^{-1/2}$$

for all  $u, v$  such that  $\tau(u) \neq \tau(v)$ .

□

We now present the proof of Corollary 4.2.1 as follows.

*Proof.* Similar to the proof of Theorem 4.2,

$$\sum_u \|C'_u - W_u O\|^2 \leq \frac{2^2 6(T_1^2 + T_2^2)}{\sigma_2^4(Q)} N \log N.$$

Let  $M = \{u : \|C'_u - W_u O\| > r\}$  be a set of the misclustered nodes, we have

$$\sum_{u \in M} \|C'_u - W_u O\|^2 \geq |M|r^2.$$

Therefore,

$$|M|r^2 \leq \sum_u \|C'_u - W_u O\|^2 \leq \frac{2^2 6(T_1^2 + T_2^2)}{\sigma_2^4(Q)} N \log N.$$

By substituting

$$r = \frac{\beta}{3} \sqrt{b\tilde{\pi}} N^{-1/2},$$

the number of misclustered nodes is bounded as follows:

$$\begin{aligned} |M| &\leq \frac{2^2 3^2 6(T_1^2 + T_2^2)}{\beta^2(b\tilde{\pi})N^{-1}\sigma_d^4(Q)} N \log N \\ &\leq \frac{2^2 3^2 6(T_1^2 + T_2^2)}{\beta^2(b\tilde{\pi})(b\tilde{\pi}N)^4} N^2 \log N \quad (\because \sigma_d(Q) \geq b\tilde{\pi}N) \\ &= \frac{2^2 3^2 6(T_1^2 + T_2^2)}{\beta^2(b\tilde{\pi})^5 N^2} \log N \end{aligned}$$

Let us define

$$\begin{aligned} TN^2 &:= T_1^2 + T_2^2 \\ &= \left[ \sum_{w \in G_1} (1 - r^2) + \sum_{w \in G_2} (1 - s^2) \right]^2 \\ &\quad + \left[ \sum_{w \in G_1} (1 - q^2) + \sum_{w \in G_2} (1 - s^2) \right]^2. \end{aligned}$$

Therefore, we have

$$|M| \leq \frac{2^2 3^2 6T}{\beta^2(b\tilde{\pi})^5} \log N,$$

where constants  $b, \beta, \tilde{\pi}, T > 0$  do not dependent on the network size  $N$ .  $\square$

## B Simulation Study

In this section, we do simulation for directed and undirected graphs using the Gaussian Mixture Model (GMM), as described in Sections B.1 and B.2, respectively. For both cases, we compare the clustering performance of spectral clustering (circle, blue), ASE (triangle, orange) and DASE (square, green).

Similar to the  $k$ -means clustering simulations, we consider three types of simulations: (1) fixing the network size and varying the network density with balanced networks, (2) fixing the network density and varying the network size with balanced networks, and (3) fixing the network size and density but varying the size of core groups ( $\pi_1$ ).

### B.1 Directed Graphs

Figure 10 shows the clustering performance measured by the average Normalised Mutual Information (NMI), along with the corresponding standard deviations (shaded areas). When the network size is fixed and the network density varies, both ASE and DASE recover the true labels even when the network density is sparse, as shown in Figure 10a. However, the standard deviation of ASE is larger than that of DASE, indicating that DASE is more stable in sparse settings.

For fixed network density and varying network sizes, the simulation results are shown in Figures 10b-10d. As seen in Figure 10b, both ASE and DASE outperform spectral clustering, as expected. Once again, the standard deviation of ASE is again larger than that of DASE, as shown in Figures 10b and 10c, confirming that DASE is more stable than ASE.

We also evaluate the computational cost of the three methods, with results shown in Figure 10d. Spectral

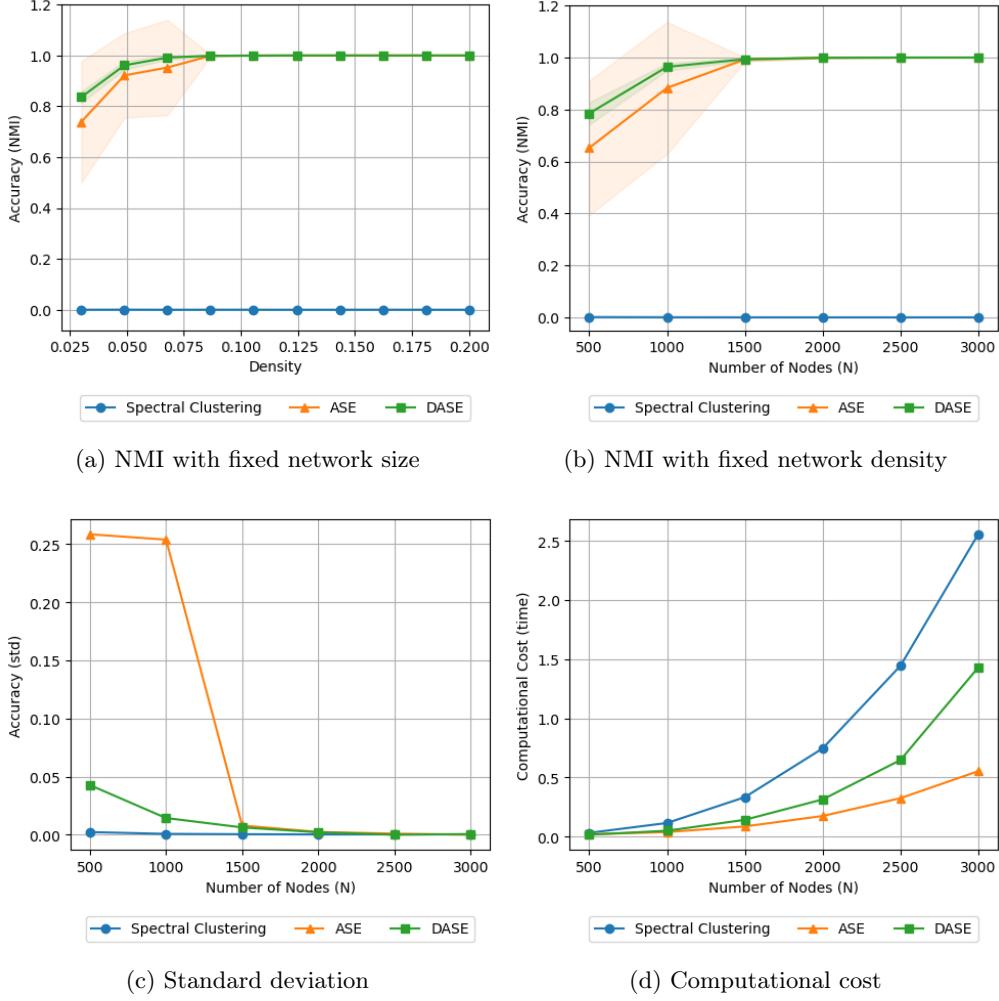


Figure 10: Figures illustrating the comparison of clustering performance on directed graphs using GMM in terms of mean accuracy (NMI), corresponding standard deviation, and mean computational cost over 50 iteration when  $K = 2$  with  $\pi = (0.5, 0.5)$ : (a) NMI with fixed network size ( $N = 1,000$ ) and varying network density; (b)-(d) NMI, standard deviation, and computational cost with fixed expected network density ( $s = 0.05$ ) and varying network sizes. In (a) and (b), shaded areas represent the standard deviations.

clustering is the most computationally expensive method, whereas ASE is the least expensive. DASE is more costly than ASE as it requires computing the doubled adjacency matrix, increasing the overall computational burden.

So far, we have examined the clustering performance on balanced networks. Here, we also consider unbalanced two-cluster networks, as shown in Figure 11. Across different proportions of the core group, ASE almost recovers the true labels, but its performance fluctuates substantially, and its standard deviation is larger than that of DASE. In contrast, DASE not only achieves better clustering accuracy than ASE but also remains more stable across different ratios. Across different ratios, DASE exhibits its largest standard deviations and reduced clustering performance when  $\pi_1 = 0.9$ , where the network structure resembles a single dense group. Overall, these results confirm that DASE remains stable even in unbalanced settings.

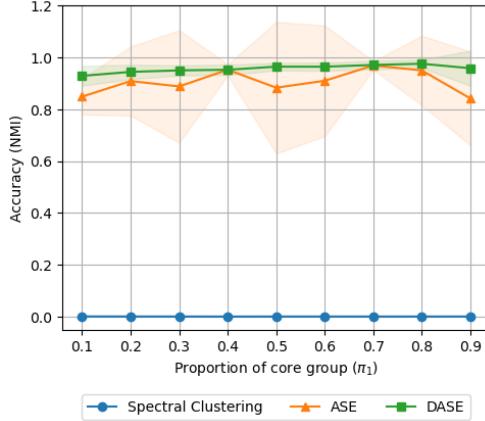


Figure 11: Figure showing the comparison of clustering performance on directed graphs using GMM in terms of mean NMI (line) and the corresponding standard deviation (shaded area) over 50 iterations when  $K = 2$ . In the simulation, the network size is fixed at ( $N = 1,000$ ), and the block probability matrix  $B$  is fixed, while varying the core group ratio ( $\pi_1$ ) from 0.1 to 0.9.

## B.2 Undirected Graphs

We next consider undirected, balanced two-cluster networks to evaluate the clustering performance of spectral clustering, ASE and DASE, as shown in Figure 12. Here, we focus on the clustering accuracy under fixed network size and fixed network density, presented in Figures 12a and 12b, respectively.

For fixed network size and varying network density in Figure 12a, ASE performs better than spectral clustering but worse than DASE, even when the network is not particularly sparse. In addition, the standard deviation of ASE is larger than that of DASE.

For fixed network density and varying network size, shown in Figure 12b, the results are consistent with those under fixed network size: DASE successfully recovers the true label when  $N > 2,500$ , whereas the accuracy of ASE falls below 0.4 when  $N = 3,000$ .

We also examine unbalanced, undirected two-cluster networks, as shown in Figure 13. As expected, DASE consistently outperforms ASE and spectral clustering across different core-group ratios. In contrast, clustering performance of ASE drops when the two groups are closer in size and increases again as the core group becomes dominant.

DASE, on the other hand, performs well except  $\pi_1 = 0.1$  and  $\pi_1 = 0.9$ , where either the core group or the periphery group is dominant. Although its accuracy decreases in these cases, it still performs better than ASE across all settings.

Here, we note one additional observation: directed networks exhibits better clustering performance than undirected networks. This is because converting a directed network into an undirected one leads to the loss of edge information. Therefore, we expect both ASE and DASE to perform better on directed networks, where no information is lost.

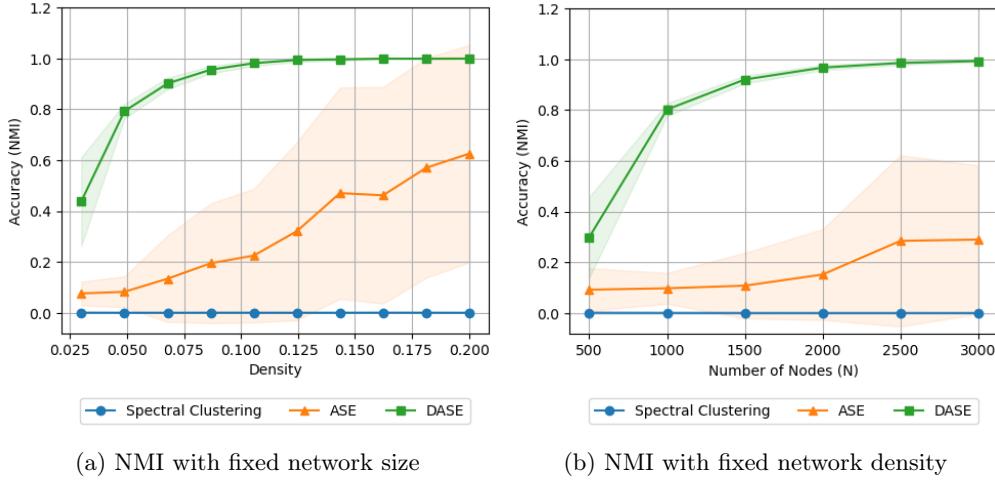


Figure 12: Figures illustrating the comparison of clustering performance on undirected graphs using GMM in terms of mean accuracy (NMI), corresponding standard deviation, and mean computational cost over 50 iteration when  $K = 2$  with  $\pi = (0.5, 0.5)$ : (a) NMI with fixed network size ( $N = 1,000$ ) and varying network density; (b) NMI with fixed expected network density ( $s = 0.05$ ) and varying network sizes. In (a) and (b), shaded areas represent the standard deviations.

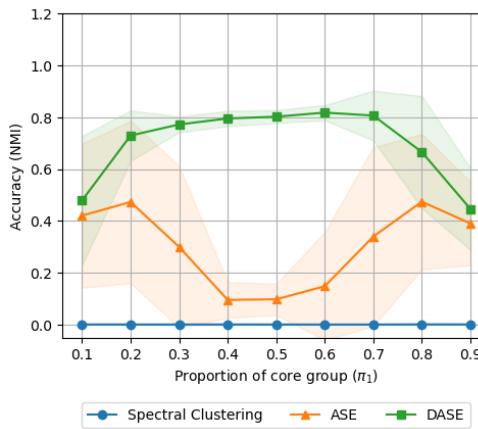


Figure 13: Figure presenting the comparison of clustering performance on undirected graphs using GMM in terms of mean NMI (line) and the corresponding standard deviation (shaded area) over 50 iterations when  $K = 2$ . In the simulation, the network size is fixed at ( $N = 1,000$ ), and the block probability matrix  $B$  is fixed, while varying the core group ratio ( $\pi_1$ ) from 0.1 to 0.9.