Network Science **7** (3): 269–291, 2019
doi:10.1017/nws.2019.23

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

On spectral embedding performance and elucidating network structure in stochastic blockmodel graphs†

Joshua Cape[1]*, Minh Tang[2] and Carey E. Priebe[3]

[1]Department of Statistics, University of Michigan 1085 South University Avenue, 323 West Hall Ann Arbor, MI 48109, USA, [2]Department of Statistics, North Carolina State University 2311 Stinson Drive, 5109 SAS Hall, Campus Box 8203 Raleigh, NC 27695, USA and [3]Department of Applied Mathematics and Statistics, Johns Hopkins University 3400 North Charles Street, Whitehead Hall 100 Baltimore, MD 21218, USA (e-mails: minh_tang@ncsu.edu, cep@jhu.edu)
*Corresponding author. Email: jrcape@umich.edu

Action Editor: Stanley Wasserman

Abstract

Statistical inference on graphs often proceeds via spectral methods involving low-dimensional embeddings of matrix-valued graph representations such as the graph Laplacian or adjacency matrix. In this paper, we analyze the asymptotic information-theoretic relative performance of Laplacian spectral embedding and adjacency spectral embedding for block assignment recovery in stochastic blockmodel graphs by way of Chernoff information. We investigate the relationship between spectral embedding performance and underlying network structure (e.g., homogeneity, affinity, core-periphery, and (un)balancedness) via a comprehensive treatment of the two-block stochastic blockmodel and the class of K-blockmodels exhibiting homogeneous balanced affinity structure. Our findings support the claim that, for a particular notion of sparsity, loosely speaking, "Laplacian spectral embedding favors relatively sparse graphs, whereas adjacency spectral embedding favors not-too-sparse graphs." We also provide evidence in support of the claim that "adjacency spectral embedding favors core-periphery network structure."

Keywords: stochastic blockmodel; Laplacian matrix, adjacency matrix, spectral embedding, network structure, core-periphery, Chernoff information

1. Preface

The stochastic blockmodel (SBM) (Holland et al., 1983) is a simple yet ubiquitous network model capable of reflecting community structure that has been widely studied via spectral methods in mathematics, statistics, physics, and engineering disciplines. Each vertex in an n-vertex K-block SBM graph belongs to one of K blocks (communities), and the probability that any two vertices share an edge depends exclusively on the vertices' block assignments (memberships).

This paper provides a detailed comparison of two popular spectral embedding procedures by synthesizing recent advances in random graph limit theory. We undertake an extensive investigation of network structure for SBM graphs by considering sub-models exhibiting various functional relationships, symmetries, and geometric properties within the inherent parameter space consisting of block membership probabilities and block edge probabilities. We also provide a collection of figures depicting relative spectral embedding performance as a function of the

†This work is partially supported by the XDATA and D3M programs of the Defense Advanced Research Projects Agency (DARPA) and by the Acheson J. Duncan Fund for the Advancement of Research in Statistics at Johns Hopkins University. Part of this work was done during visits by JC and CEP to the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge under EPSCR grant EP/K032208/1. JC thanks Zachary Lubberts for productive discussions.

SBM parameter space for a range of sub-models exhibiting different forms of network structure, specifically homogeneous community structure, affinity structure, core-periphery structure, and (un)balanced block sizes. The remainder of this paper is organized as follows.

- Section 2 introduces the formal setting considered in this paper and contextualizes this work with respect to the existing statistical network analysis literature.
- Section 3 establishes notation, presents the *generalized random dot product graph* (GRDPG) *model* (of which the SBM is a special case), defines *adjacency spectral embedding* (ASE) and *Laplacian spectral embedding* (LSE), presents the corresponding spectral embedding limit theorems, and specifies the notion of sparsity considered in this paper.
- Section 4 motivates and formulates a measure of large-sample relative spectral embedding performance via Chernoff information.
- Section 5 presents a treatment of the two-block SBM and certain K-block SBMs whereby we elucidate the relationship between spectral embedding performance and network model structure.
- Section 6 offers further discussion and some concluding remarks.
- Section 7 provides additional details intended to supplement the main body of this paper.

2. Introduction

Formally, we consider the following SBM setting.

Definition 1 (K-block SBM). *Let $K \geq 2$ be a positive integer and $\boldsymbol{\pi}$ be a vector in the interior of the $(K-1)$-dimensional unit simplex in $\mathbb{R}^K$. Let $\mathbf{B} \in (0,1)^{K \times K}$ be a symmetric matrix with distinct rows. We say $(\mathbf{A}, \boldsymbol{\tau}) \sim \mathrm{SBM}(\mathbf{B}, \boldsymbol{\pi})$ with scaling factor $0 < \rho_n \leq 1$ provided the following conditions hold. Firstly, $\boldsymbol{\tau} \equiv (\tau_1, \ldots, \tau_n)^\top$, where τ_i are independent and identically distributed (i.i.d.) random variables with $\mathbb{P}[\tau_i = k] = \pi_k$. Then, $\mathbf{A} \in \{0,1\}^{n \times n}$ denotes a symmetric (adjacency) matrix such that, conditional on $\boldsymbol{\tau}$, for all $i \leq j$, the entries A_{ij} are independent Bernoulli random variables with $\mathbb{E}[A_{ij}] = \rho_n B_{\tau_i, \tau_j}$. If only $\mathbf{A}$ is observed, then we write $\mathbf{A} \sim \mathrm{SBM}(\mathbf{B}, \boldsymbol{\pi})$.*[1] ▲

The SBM is an example of an inhomogeneous Erdős–Rényi random graph model (Bollobás et al., 2007) and reduces to the classical Erdős–Rényi model (Erdős & Rényi, 1959) in the degenerate case when all the entries of $\mathbf{B}$ are identical. The SBM enjoys an extensive body of literature focused on spectral methods (von Luxburg, 2007) for statistical estimation, inference, and community detection (Fishkind et al., 2013; McSherry, 2001; Lei & Rinaldo, 2015; Rohe et al., 2011; Sussman et al., 2014; Sarkar & Bickel, 2015). Considerable effort has also been devoted to the information-theoretic and computational investigation of the SBM as a result of interest in the community detection problem; for an overview, see Abbe (2018). Popular variants of the SBM include the mixed-membership SBM (Airoldi et al., 2008) and the degree-corrected SBM (Karrer & Newman, 2011).

Within the statistics literature, substantial attention has been paid to the class of K-block SBMs with positive semidefinite block edge probability matrices $\mathbf{B}$. This is due in part to the extensive study of the RDPG model (Nickel, 2006; Young & Scheinerman, 2007; Athreya et al., 2018), a latent position random graph model (Hoff et al., 2002) which includes positive semidefinite SBMs as a special case. Notably, it was recently shown that for the RDPG model, both LSE (see Definition 3) and ASE (see Definition 3) behave approximately as random samples from Gaussian mixture models (Athreya et al., 2016; Tang & Priebe, 2018). In tandem with these limit results, the concept of Chernoff information (Chernoff, 1952) was employed in Tang and Priebe (2018) to demonstrate that neither LSE nor ASE dominates the other for subsequent inference as a spectral embedding method when the underlying inference task is to recover vertices' latent block assignments. In doing so, the results in Tang and Priebe (2018) clarify and complete the groundbreaking work in Sarkar and Bickel (2015) on comparing spectral clusterings for certain SBM graphs.

Tang and Priebe leave open the problem of comprehensively investigating Chernoff information as a measure of relative spectral embedding performance for SBM graphs. Moreover, they do not investigate how relative spectral embedding performance corresponds to underlying network model structure. This is understandable, since the positive semidefinite restriction on $\mathbf{B}$ limits the possible network structure that can be investigated under the RDPG model.

More recently, the limit theory in Tang and Priebe (2018) was extended to hold for *all* SBMs within the more flexible framework of the *generalized RDPG* (GRDPG) *model* (Rubin-Delanchy et al., 2017). This advancement now makes it possible to conduct a more comprehensive Chernoff-based analysis, and that is precisely the aim of this paper. We set forth to formulate and analyze a criterion based on Chernoff information for quantifying relative spectral embedding performance which we then further consider in conjunction with underlying network model structure. The investigation carried out in this paper is, to the best of our knowledge, among the first of its kind in the study of statistical network analysis and random graph inference.

This paper focuses on the following two models which have garnered widespread interest [e.g., see Abbe (2018) and the references therein].

(1) The two-block SBM with $\mathbf{B} = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$ and $\boldsymbol{\pi} = (\pi_1, 1-\pi_1)^\top$, where $a, b, c, \pi_1 \in (0,1)$;
(2) The $K \geq 2$ block SBM exhibiting homogeneous balanced affinity structure, that is, $B_{ij} = a$ for all $i = j$, $B_{ij} = b$ for all $i \neq j$, $0 < b < a < 1$, and $\boldsymbol{\pi} = \left(\frac{1}{K}, \ldots, \frac{1}{K}\right)^\top \in \mathbb{R}^K$.

Using the concept of Chernoff information (Section 4), we obtain an information-theoretic summary characteristic $\rho^\star \equiv \rho^\star(\mathbf{B}, \boldsymbol{\pi})$ such that the cases $\rho^\star > 1$, $\rho^\star < 1$, and $\rho^\star = 1$ correspond to the preference of spectral embedding procedure based on approximate large-sample relative performance, summarized as ASE $>$ LSE, ASE $<$ LSE, and ASE $=$ LSE, respectively. The above models' low-dimensional parameter spaces facilitate visualizing and analyzing the relationship between network structure [i.e., SBM$(\mathbf{B}, \boldsymbol{\pi})$] and embedding performance [i.e., $\rho^\star(\mathbf{B}, \boldsymbol{\pi})$].

This paper considers the task of performing inference on a single large graph. As such, we interpret the notion of *sparsity* in reference to the magnitudes of probability parameters, namely the magnitudes of the entries of $\mathbf{B}$. This notion of sparsity corresponds to the interpretation and intuition of a practitioner wanting to do statistics with an observed graph. We shall, with this understanding in mind, subsequently demonstrate that LSE is preferred as an embedding method in relatively sparse regimes, whereas ASE is preferred as an embedding method in not-too-sparse regimes.

We remark that the scaling factor ρ_n in Definition 1, which is included at the onset for the purpose of general presentation, indexes a sequence of models wherein edge probabilities change with n. Given our interest in single large graph inference where edge probabilities do not depend on n, we take ρ_n to be constant in n, which by rescaling is equivalent to setting $\rho_n \equiv 1$. Limit theorems are known for regimes where $\rho_n \to 0$ as $n \to \infty$, but these regimes are uninteresting for single graph inference from the perspective of relative spectral embedding performance (Tang & Priebe, 2018).

3. Preliminaries

3.1 Notation

In this paper, all vectors and matrices are real-valued. The symbols $:=$ and $\equiv$ are used to assign definitions and to denote formal equivalence, respectively. Given a symmetric positive definite $n \times n$ matrix $\mathbf{M}$, let $\langle \cdot, \cdot \rangle_{\mathbf{M}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ denote the real inner product induced by $\mathbf{M}$. Similarly, define the induced norm as $\| \cdot \|_{\mathbf{M}} := \sqrt{\langle \cdot, \cdot \rangle_{\mathbf{M}}}$. In particular, given the $n \times n$ identity matrix $\mathbf{I}$, denote the standard Euclidean inner product and Euclidean norm by $\langle \cdot, \cdot \rangle \equiv \langle \cdot, \cdot \rangle_{\mathbf{I}}$ and $\| \cdot \|_2 := \sqrt{\langle \cdot, \cdot \rangle}$, respectively. In this paper, $\det(\cdot)$ and $\mathrm{tr}(\cdot)$ denote the matrix determinant and

matrix trace operator, respectively. Given a diagonal matrix $\mathbf{D} := \operatorname{diag}(d_{11}, d_{22}, \ldots, d_{nn}) \in \mathbb{R}^{n\times n}$, $|\mathbf{D}|$ denotes the entrywise absolute value of $\mathbf{D}$.

The vector of all ones in $\mathbb{R}^n$ is denoted by $\mathbf{1}_n$, whereas the zero matrix in $\mathbb{R}^{m\times n}$ is denoted by $\mathbf{0}_{m,n}$. We suppress the indices for convenience when the underlying dimensions are understood, writing instead $\mathbf{1}$ and $\mathbf{0}$.

Let $\mathbb{N} := \{1, 2, 3, \ldots\}$ denote the set of natural numbers so that $[n] := \{1, 2, \ldots, n\}$ whenever $n \in \mathbb{N}$. For integers $d^+ \geq 1$, $d^- \geq 0$, and $d := d^+ + d^-$, let $\mathbf{I}_{d^-}^{d^+} := \mathbf{I}_{d^+} \bigoplus (-\mathbf{I}_{d^-}) \in \mathbb{R}^{d\times d}$ be the direct sum (diagonal) matrix consisting of the identity matrix $\mathbf{I}_{d^+} \in \mathbb{R}^{d^+\times d^+}$ and $(-\mathbf{I}_{d^-}) \in \mathbb{R}^{d^-\times d^-}$, together with the convention that $\mathbf{I}_0^{d^+} \equiv \mathbf{I}_{d^+}$. For example, $\mathbf{I}_1^1 \equiv \operatorname{diag}(1, -1) \in \mathbb{R}^{2\times 2}$ and $\mathbf{I}_1^2 \equiv \mathbf{I}_2 \oplus (-\mathbf{I}_1) \equiv \operatorname{diag}(1, 1, -1) \in \mathbb{R}^{3\times 3}$.

For integers $n \geq d \geq 1$, the set of all $n \times d$ real matrices with orthonormal columns shall be denoted by $\mathbb{O}_{n,d}$. Let $\mathbb{O}(d^+, d^-)$ denote the indefinite orthogonal group with signature (d^+, d^-), and let $\mathbb{O}_{d^+} \equiv \mathbb{O}_{d^+,d^+} \equiv \mathbb{O}(d^+, 0)$ denote the orthogonal group in $\mathbb{R}^{d^+\times d^+}$. In particular, $\mathbf{M} \in \mathbb{O}(d^+, d^-)$ has the characterization $\mathbf{M}^\top \mathbf{I}_{d^-}^{d^+} \mathbf{M} = \mathbf{I}_{d^-}^{d^+}$. In the case of the orthogonal group, this characterization reduces to the relationship $\mathbf{M}^\top \equiv \mathbf{M}^{-1}$.

3.2 The GRDPG model

A growing corpus has emerged within the statistics literature focused on the development of theory and applications for the RDPG model (Nickel, 2006; Young & Scheinerman, 2007). This latent position random graph model associates to each vertex in a graph an underlying low-dimensional vector. These vectors may be viewed as encoding structural information or attributes possessed by their corresponding vertices. In turn, the probability of two vertices sharing an edge is specified through the standard Euclidean inner (dot) product of the vertices' latent position vectors. While simple in concept and design, this model has proven successful in real-world applications in the areas of neuroscience and social networks (Lyzinski et al., 2017). On the theoretical side, the RDPG model enjoys some of the first-ever statistical theory for two-sample hypothesis testing on random graphs, both semiparametric (Tang et al., 2017*b*) and nonparametric (Tang et al., 2017*a*). For more on the RDPG model, see the survey Athreya et al. (2018) and the references therein.

More recently, the GRDPG model was introduced as an extension of the RDPG model that includes as special cases the mixed membership SBM as well as *all* (single membership) SBMs (Rubin-Delanchy et al., 2017). Effort toward developing theory for the GRDPG model has already produced new findings and raised new research questions related to the geometry of spectral methods, embeddings, and random graph inference. The present paper further contributes to these efforts.

Definition 2 (The GRDPG model). *For integers $d^+ \geq 1$ and $d^- \geq 0$ such that $d := d^+ + d^-$, let F be a distribution on a set $\mathcal{X} \subset \mathbb{R}^d$ such that $\langle \mathbf{I}_{d^-}^{d^+} x, y\rangle \equiv y^\top \mathbf{I}_{d^-}^{d^+} x \in [0, 1]$ for all $x, y \in \mathcal{X}$. We say that $(\mathbf{X}, \mathbf{A}) \sim \mathit{GRDPG}(F)$ with signature (d^+, d^-) and scaling factor $0 < \rho_n \leq 1$ if the following hold. Let $X_1, \ldots, X_n \sim F$ be independent and identically distributed random (latent position) vectors with*

$$\mathbf{X} := [X_1 | \cdots | X_n]^\top \in \mathbb{R}^{n\times d} \text{ and } \mathbf{P} := \rho_n \mathbf{X} \mathbf{I}_{d^-}^{d^+} \mathbf{X}^\top \in [0, 1]^{n\times n}. \tag{1}$$

For each $i \leq j$, the entries A_{ij} of the symmetric adjacency matrix $\mathbf{A} \in \{0, 1\}^{n\times n}$ are then generated in a conditionally independent fashion given the latent positions, namely

$$\{A_{ij} | X_i, X_j\} \sim \text{Bernoulli}(\rho_n \langle \mathbf{I}_{d^-}^{d^+} X_i, X_j\rangle). \tag{2}$$

In this setting, the conditional probability $\Pr[\mathbf{A}|\mathbf{X}]$ *can be computed explicitly as a product of Bernoulli probabilities.* ▲

To reiterate, we consider the regime $\rho_n \equiv 1$ and therefore suppress dependencies on ρ_n later in the text. In addition, moving forward we use adorned versions of the symbol ρ to denote Chernoff-related quantities unrelated to ρ_n in a manner consistent with the notation in Tang and Priebe (2018) (e.g., see Section 4).

When $d^- = 0$, the GRDPG model reduces to the RDPG model. When the distribution F is a discrete distribution on a finite collection of vectors in $\mathbb{R}^d$, then the GRDPG model coincides with the SBM, in which case the $n \times n$ edge probability matrix $\mathbf{P}$ arises as an appropriate dilation of the $K \times K$ block edge probability matrix $\mathbf{B}$. Moreover, given any valid $\mathbf{B} \in (0,1)^{K\times K}$ as in Definition 1, there exist integers d^+, d^-, and a matrix $\mathbf{X} \in \mathbb{R}^{K\times d}$ such that $\mathbf{B}$ has the (not necessarily unique) factorization $\mathbf{B} \equiv \mathbf{X}\mathbf{I}_{d^-}^{d^+}\mathbf{X}^\top$, which follows since the spectral decomposition of $\mathbf{B}$ can be written as $\mathbf{B} \equiv \mathbf{U}_\mathbf{B}\mathbf{\Lambda}\mathbf{U}_\mathbf{B}^\top = (\mathbf{U}_\mathbf{B}|\mathbf{\Lambda}|^{1/2})\mathbf{I}_{d^-}^{d^+}(\mathbf{U}_\mathbf{B}|\mathbf{\Lambda}|^{1/2})^\top := \mathbf{X}\mathbf{I}_{d^-}^{d^+}\mathbf{X}^\top$. This demonstrates the ability of the GRDPG framework in Definition 2 to model all possible SBMs formulated in Definition 1.

Remark 1 (*Nonidentifiability in the GRDPG model).* The GRDPG model possess two intrinsic sources of nonidentifiability, summarized as "uniqueness up to indefinite orthogonal transformation" and "uniqueness up to artificial dimension blow-up." More precisely, for $(\mathbf{X}, \mathbf{A}) \sim \text{GRDPG}(F)$ with signature (d^+, d^-), the following considerations must be taken into account:

(1) For any $\mathbf{Q} \in \mathbb{O}(d^+, d^-)$, $(\mathbf{X}, \mathbf{A}) \overset{\text{d}}{=} (\mathbf{Y}, \mathbf{B})$ whenever $(\mathbf{Y}, \mathbf{B}) \sim \text{GRDPG}(F \circ \mathbf{Q})$, where $F \circ \mathbf{Q}$ denotes the distribution of the latent position vector $Y_1 = \mathbf{Q}X_1$ and $\overset{\text{d}}{=}$ denotes equality in distribution. This source of nonidentifiability cannot be mitigated. See Equation (2).

(2) There exists a distribution F' on $\mathbb{R}^{d'}$ for some $d' > d$ such that $(\mathbf{X}, \mathbf{A}) \overset{\text{d}}{=} (\mathbf{Y}, \mathbf{B})$, where $(\mathbf{Y}, \mathbf{B}) \sim \text{GRDPG}(F')$. This source of nonidentifiability can be avoided by assuming, as we do in this paper, that F is nondegenerate in the sense that for $X_1 \sim F$, the second moment matrix $\mathbb{E}[X_1X_1^\top] \in \mathbb{R}^{d\times d}$ is full rank.

Definition 3 (ASE and LSE)**.** *Let $\mathbf{A} \in \{0,1\}^{n\times n}$ be a symmetric adjacency matrix with eigendecomposition $\mathbf{A} \equiv \sum_{i=1}^n \lambda_i u_i u_i^\top$ and with ordered eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ corresponding to the collection of orthonormal eigenvectors $u_1, u_2, \ldots, u_n$. Given a positive integer d such that $d \leq n$, let $\mathbf{S}_\mathbf{A} := \text{diag}(\lambda_1, \ldots, \lambda_d)$ and $\mathbf{U}_\mathbf{A} := [u_1|\ldots|u_d] \in \mathbb{O}_{n,d}$. The ASE of $\mathbf{A}$ into $\mathbb{R}^d$ is then defined to be the $n \times d$ matrix $\hat{\mathbf{X}} := \mathbf{U}_\mathbf{A}|\mathbf{S}_\mathbf{A}|^{1/2}$. [The matrix $\hat{\mathbf{X}}$ serves as a consistent estimator for $\mathbf{X}$ up to indefinite orthogonal transformation as $n \to \infty$].*

Along similar lines, define the normalized Laplacian of $\mathbf{A}$ as

$$\mathcal{L}(\mathbf{A}) := (\text{diag}(\mathbf{A}\mathbf{1}_n))^{-1/2}\mathbf{A}(\text{diag}(\mathbf{A}\mathbf{1}_n))^{-1/2} \in \mathbb{R}^{n\times n} \tag{3}$$

whose eigendecomposition is given by $\mathcal{L}(\mathbf{A}) \equiv \sum_{i=1}^n \tilde{\lambda}_i \tilde{u}_i \tilde{u}_i^\top$ with ordered eigenvalues $|\tilde{\lambda}_1| \geq |\tilde{\lambda}_2| \geq \cdots \geq |\tilde{\lambda}_n|$ corresponding to orthonormal eigenvectors $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$. Given a positive integer d such that $d \leq n$, let $\tilde{\mathbf{S}}_\mathbf{A} := \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_d)$ and let $\tilde{\mathbf{U}}_\mathbf{A} := [\tilde{u}_1|\ldots|\tilde{u}_d] \in \mathbb{O}_{n,d}$. The LSE of $\mathbf{A}$ into $\mathbb{R}^d$ is then defined to be the $n \times d$ matrix $\check{\mathbf{X}} := \tilde{\mathbf{U}}_\mathbf{A}|\tilde{\mathbf{S}}_\mathbf{A}|^{1/2}$. [The matrix $\check{\mathbf{X}}$ serves as a consistent estimator for the matrix $(\text{diag}(\mathbf{X}\mathbf{I}_{d^-}^{d^+}\mathbf{X}^\top\mathbf{1}_n))^{-1/2}\mathbf{X}$ up to indefinite orthogonal transformation as $n \to \infty$]. ▲

Remark 2 (*Consistent estimation and parametrization involving latent positions).* The matrices $\mathbf{X}$ and $(\text{diag}(\mathbf{X}\mathbf{I}_{d^-}^{d^+}\mathbf{X}^\top\mathbf{1}_n))^{-1/2}\mathbf{X}$, which are one-to-one invertible transformations of each other, may be viewed as providing different parametrizations of GRDPG graphs. As such, comparing $\hat{\mathbf{X}}$ and $\check{\mathbf{X}}$ as estimators is rather technical in nature and not immediately straightforward. In order to carry out such a comparison, we subsequently adopt an information-theoretic approach in which we consider a particular choice of f-divergence which is both analytically tractable and statistically interpretable in the current setting.

For the subsequent purposes of this paper, Theorems 1 and 2 (discussed next) state slightly weaker formulations of the corresponding limit theorems obtained in Rubin-Delanchy et al. (2017) for ASE and LSE.

Theorem 1 (ASE limit theorem for GRDPG, adapted (Rubin-Delanchy et al., 2017)). *Assume the d-dimensional GRDPG setting in Definition 2 with $\rho_n \equiv 1$. Let $\hat{\mathbf{X}}$ be the ASE into $\mathbb{R}^d$ with ith row denoted by $\hat{X}_i$. Let $\Phi(\cdot, \mathbf{\Sigma})$ denote the cumulative distribution function of the centered multivariate normal distribution in $\mathbb{R}^d$ with covariance matrix $\mathbf{\Sigma}$. Then, with respect to ASE, there exists a sequence of matrices $\mathbf{Q} \equiv \mathbf{Q}_n \in \mathbb{O}(d^+, d^-)$ such that, for any $z \in \mathbb{R}^d$,*

$$\Pr\left[\sqrt{n}\left(\mathbf{Q}\hat{X}_i - X_i\right) \leq z\right] \to \int_{\mathcal{X}} \Phi(z, \mathbf{\Sigma}(x))dF(x) \tag{4}$$

as $n \to \infty$, where for $X_1 \sim F$,

$$\mathbf{\Sigma}(x) := \mathbf{I}_{d^-}^{d^+} \mathbf{\Delta}^{-1} \mathbb{E}\left[g(x, X_1)X_1X_1^\top\right] \mathbf{\Delta}^{-1} \mathbf{I}_{d^-}^{d^+}$$

with $\mathbf{\Delta} := \mathbb{E}[X_1X_1^\top]$ and $g(x, X_1) := \langle \mathbf{I}_{d^-}^{d^+}x, X_1\rangle(1 - \langle \mathbf{I}_{d^-}^{d^+}x, X_1\rangle)$.

Theorem 2 (LSE limit theorem for GRDPG, adapted (Rubin-Delanchy et al., 2017)). *Assume the d-dimensional GRDPG setting in Definition 2 with $\rho_n \equiv 1$. Let $\check{\mathbf{X}}$ be the LSE into $\mathbb{R}^d$ with ith row denoted by $\check{X}_i$. Let $\Phi(\cdot, \mathbf{\Sigma})$ denote the cumulative distribution function of the centered multivariate normal distribution in $\mathbb{R}^d$ with covariance matrix $\mathbf{\Sigma}$. Then, with respect to LSE, there exists a sequence of matrices $\tilde{\mathbf{Q}} \equiv \tilde{\mathbf{Q}}_n \in \mathbb{O}(d^+, d^-)$ such that, for any $z \in \mathbb{R}^d$,*

$$\Pr\left[n\left(\tilde{\mathbf{Q}}\check{X}_i - \frac{X_i}{\sqrt{\sum_j \langle \mathbf{I}_{d^-}^{d^+}X_i, X_j\rangle}}\right) \leq z\right] \to \int_{\mathcal{X}} \Phi(z, \tilde{\mathbf{\Sigma}}(x))dF(x) \tag{5}$$

as $n \to \infty$, where for $X_1 \sim F$ and $\mu := \mathbb{E}[X_1]$,

$$\tilde{\mathbf{\Sigma}}(x) := \mathbf{I}_{d^-}^{d^+} \tilde{\mathbf{\Delta}}^{-1} \mathbb{E}\left[\tilde{g}(x, X_1)\left(\frac{X_1}{\langle \mathbf{I}_{d^-}^{d^+}\mu, X_1\rangle} - \frac{\tilde{\mathbf{\Delta}}\mathbf{I}_{d^-}^{d^+}x}{2\langle \mathbf{I}_{d^-}^{d^+}\mu, x\rangle}\right)\left(\frac{X_1}{\langle \mathbf{I}_{d^-}^{d^+}\mu, X_1\rangle} - \frac{\tilde{\mathbf{\Delta}}\mathbf{I}_{d^-}^{d^+}x}{2\langle \mathbf{I}_{d^-}^{d^+}\mu, x\rangle}\right)^\top\right] \tilde{\mathbf{\Delta}}^{-1} \mathbf{I}_{d^-}^{d^+}$$

with $\tilde{\mathbf{\Delta}} := \mathbb{E}\left[\langle \mathbf{I}_{d^-}^{d^+}\mu, X_1\rangle^{-1} X_1X_1^\top\right]$ and $\tilde{g}(x, X_1) := \left[\langle \mathbf{I}_{d^-}^{d^+}\mu, x\rangle^{-1}\langle \mathbf{I}_{d^-}^{d^+}x, X_1\rangle(1 - \langle \mathbf{I}_{d^-}^{d^+}x, X_1\rangle)\right]$.

4. Spectral embedding performance

We desire to compare the large-n sample relative performance of ASE and LSE for subsequent inference, where the subsequent inference task is naturally taken to be the problem of recovering latent block assignments. Here, quantifying spectral embedding performance will correspond to approximating the large-sample optimal error rate for recovering the underlying block assignments following each of the spectral embeddings. Toward this end, we now introduce Chernoff information and Chernoff divergence as appropriate information-theoretic quantities.

Given independent and identically distributed random vectors Y_i arising from one of two absolutely continuous multivariate distributions F_1 and F_2 on $\Omega = \mathbb{R}^d$ with density functions f_1 and f_2, respectively, we are interested in testing the simple null hypothesis $\mathbb{H}_0 : F = F_1$ against the simple alternative hypothesis $\mathbb{H}_A : F = F_2$. In this framework, a statistical test T can be viewed as a sequence of mappings $T_m : \Omega^m \to \{1, 2\}$ indexed according to sample size m such that T_m returns the value two when $\mathbb{H}_0$ is rejected in favor of $\mathbb{H}_A$ and correspondingly returns the value one when $\mathbb{H}_0$ is favored. For each m, the corresponding significance level and type-II error are denoted by α_m and β_m, respectively.

Assume that the prior probability of $\mathbb{H}_0$ being true is given by $\pi \in (0, 1)$. For a given $\alpha_m^\star \in (0, 1)$, let $\beta_m^\star \equiv \beta_m^\star(\alpha_m^\star)$ denote the type-II error associated with the corresponding likelihood ratio test when the type-I error is at most $\alpha_m^\star$. Then, the *Bayes risk* in deciding between $\mathbb{H}_0$ and $\mathbb{H}_A$ given m independent random vectors $Y_1, Y_2, \ldots, Y_m$ is given by

$$\inf_{\alpha_m^\star \in (0,1)} \pi\alpha_m^\star + (1-\pi)\beta_m^\star. \tag{6}$$

The Bayes risk is intrinsically related to *Chernoff information* (Chernoff, 1952, 1956), $C(F_1, F_2)$, namely

$$\lim_{m\to\infty} \frac{1}{m}\left[\inf_{\alpha_m^\star \in (0,1)} \log\left(\pi\alpha_m^\star + (1-\pi)\beta_m^\star\right)\right] = -C(F_1, F_2) \tag{7}$$

where

$$C(F_1, F_2) := -\log\left[\inf_{t\in(0,1)} \int_{\mathbb{R}^d} f_1^t(x) f_2^{1-t}(x)dx\right] = \sup_{t\in(0,1)}\left[-\log\int_{\mathbb{R}^d} f_1^t(x) f_2^{1-t}(x)dx\right].$$

In words, the Chernoff information between F_1 and F_2 is the exponential rate at which the Bayes risk decreases as $m \to \infty$. Note that the Chernoff information is independent of the prior probability π. A version of Equation (7) also holds when considering $K \geq 3$ hypothesis with distributions $F_1, F_2, \ldots, F_K$, thereby introducing the quantity $\min_{k\neq l} C(F_k, F_l)$ [see, e.g., Tang and Priebe (2018)].

Chernoff information can be expressed in terms of the *Chernoff divergence* between distributions F_1 and F_2, defined for $t \in (0, 1)$ as

$$C_t(F_1, F_2) = -\log\int_{\mathbb{R}^d} f_1^t(x) f_2^{1-t}(x)dx \tag{8}$$

which yields the relation

$$C(F_1, F_2) = \sup_{t\in(0,1)} C_t(F_1, F_2). \tag{9}$$

Chernoff divergence is an example of an f-divergence and as such satisfies the data processing lemma (Liese & Vajda, 2006) and is invariant with respect to invertible transformations (Devroye et al., 2013). One could instead use another f-divergence for the purpose of comparing the two embedding methods, such as the Kullback–Liebler divergence. Our choice is motivated by the aforementioned relationship with Bayes risk in Equation (7).

In this paper, we explicitly consider multivariate normal distributions as a consequence of Theorems 1 and 2 when conditioning on the individual underlying latent positions for SBM graphs. In particular, given $F_1 = \mathcal{N}(\mu_1, \mathbf{\Sigma}_1)$, $F_2 = \mathcal{N}(\mu_2, \mathbf{\Sigma}_2)$, and $t \in (0, 1)$, then for $\mathbf{\Sigma}_t := t\mathbf{\Sigma}_1 + (1-t)\mathbf{\Sigma}_2$, the Chernoff information between F_1 and F_2 is given by

$$\begin{aligned} C(F_1, F_2) &= \sup_{t\in(0,1)}\left[\tfrac{t(1-t)}{2}(\mu_2-\mu_1)^\top \mathbf{\Sigma}_t^{-1}(\mu_2-\mu_1) + \tfrac{1}{2}\log\left(\frac{\det(\mathbf{\Sigma}_t)}{\det(\mathbf{\Sigma}_1)^t\det(\mathbf{\Sigma}_2)^{1-t}}\right)\right] \\ &= \sup_{t\in(0,1)}\left[\tfrac{t(1-t)}{2}\|\mu_2-\mu_1\|^2_{\mathbf{\Sigma}_t^{-1}} + \tfrac{1}{2}\log\left(\frac{\det(\mathbf{\Sigma}_t)}{\det(\mathbf{\Sigma}_1)^t\det(\mathbf{\Sigma}_2)^{1-t}}\right)\right]. \end{aligned}$$

Let $\mathbf{B} \in (0, 1)^{K\times K}$ and $\boldsymbol{\pi}$ denote the matrix of block edge probabilities and the vector of block assignment probabilities for a K-block SBM as before. This corresponds to a special case of the GRDPG model with signature (d^+, d^-), $d^+ + d^- = \text{rank}(\mathbf{B})$, and latent positions $\nu_k \in \mathbb{R}^{\text{rank}(\mathbf{B})}$. For an n-vertex SBM graph with parameters $(\mathbf{B}, \boldsymbol{\pi})$, the large-sample optimal error rate for

recovering block assignments when performing ASE can be characterized by the quantity $\rho_{\mathrm{A}} \equiv \rho_{\mathrm{A}}(\mathbf{B}, \boldsymbol{\pi}, n)$ defined by

$$\rho_{\mathrm{A}} := \min_{k \neq l} \sup_{t \in (0,1)} \left[\frac{nt(1-t)}{2} \| \nu_k - \nu_l \|^2_{\boldsymbol{\Sigma}_{kl}^{-1}(t)} + \frac{1}{2} \log \left(\frac{\det(\boldsymbol{\Sigma}_{kl}(t))}{\det(\boldsymbol{\Sigma}_k)^t \det(\boldsymbol{\Sigma}_l)^{1-t}} \right) \right] \tag{10}$$

where $\boldsymbol{\Sigma}_{kl}(t) := t\boldsymbol{\Sigma}_k + (1-t)\boldsymbol{\Sigma}_l$ for $t \in (0, 1)$.

Similarly, for LSE one has $\rho_{\mathrm{L}} \equiv \rho_{\mathrm{L}}(\mathbf{B}, \boldsymbol{\pi}, n)$ given by

$$\rho_{\mathrm{L}} := \min_{k \neq l} \sup_{t \in (0,1)} \left[\frac{nt(1-t)}{2} \| \tilde{\nu}_k - \tilde{\nu}_l \|^2_{\tilde{\boldsymbol{\Sigma}}_{kl}^{-1}(t)} + \frac{1}{2} \log \left(\frac{\det(\tilde{\boldsymbol{\Sigma}}_{kl}(t))}{\det(\tilde{\boldsymbol{\Sigma}}_k)^t \det(\tilde{\boldsymbol{\Sigma}}_l)^{1-t}} \right) \right] \tag{11}$$

where $\tilde{\boldsymbol{\Sigma}}_{kl}(t) := t\tilde{\boldsymbol{\Sigma}}_k + (1-t)\tilde{\boldsymbol{\Sigma}}_l$ and $\tilde{\nu}_k := \nu_k / (\sum_{k'} \pi_{k'} \langle \mathbf{I}_{d^-}^{d^+} \nu_{k'}, \nu_k \rangle)^{1/2}$.

To assist in the comparison and interpretation of the quantities ρ_{A} and ρ_{L} (namely for $\tilde{\nu}_k$), we are assuming throughout this paper that $n_k = n\pi_k$. The factor n in Equations (10) and (11) arises from the implicit consideration of the n-sample precision matrices $(\boldsymbol{\Sigma}_{kl}(t)/n)^{-1}$ and $(\tilde{\boldsymbol{\Sigma}}_{kl}(t)/n)^{-1}$. The logarithmic terms in Equations (10) and (11) are dominated by the preceding terms for large n, collectively motivating the following large-sample measure of relative performance $\rho^\star$. In particular, as $n \to \infty$, it holds that

$$\frac{\rho_{\mathrm{A}}}{\rho_{\mathrm{L}}} \equiv \frac{\rho_{\mathrm{A}}(n)}{\rho_{\mathrm{L}}(n)} \to \rho^\star \equiv \frac{\rho_{\mathrm{A}}^\star}{\rho_{\mathrm{L}}^\star} := \frac{\min_{k \neq l} \sup_{t \in (0,1)} \left[t(1-t) \| \nu_k - \nu_l \|^2_{\boldsymbol{\Sigma}_{kl}^{-1}(t)} \right]}{\min_{k \neq l} \sup_{t \in (0,1)} \left[t(1-t) \| \tilde{\nu}_k - \tilde{\nu}_l \|^2_{\tilde{\boldsymbol{\Sigma}}_{kl}^{-1}(t)} \right]}. \tag{12}$$

Here we have suppressed the functional dependence on the underlying model parameters $\mathbf{B}$ and $\boldsymbol{\pi}$. For large n, observe that as $\rho_{\mathrm{A}}^\star$ increases, ρ_{A} also increases, and therefore the large-sample optimal error rate corresponding to ASE decreases in light of Equation (7) and its generalization. Similarly, large values of $\rho_{\mathrm{L}}^\star$ correspond to good theoretical performance of LSE. Thus, if $\rho^\star > 1$, then ASE is to be preferred to LSE, whereas if $\rho^\star < 1$, then LSE is to be preferred to ASE. The case when $\rho^\star = 1$ indicates that neither ASE nor LSE is superior for the given parameters $\mathbf{B}$ and $\boldsymbol{\pi}$. To reiterate, we summarize these preferences as ASE $>$ LSE, ASE $<$ LSE, and ASE $=$ LSE, respectively.

In what follows, we fixate on the asymptotic quantity $\rho^\star$. For the two-block SBM and certain K-block SBMs exhibiting symmetry, Equation (12) reduces to the simpler form

$$\rho^\star = \frac{\sup_{t \in (0,1)} \left[t(1-t) \| \nu_1 - \nu_2 \|^2_{\boldsymbol{\Sigma}_{1,2}^{-1}(t)} \right]}{\sup_{t \in (0,1)} \left[t(1-t) \| \tilde{\nu}_1 - \tilde{\nu}_2 \|^2_{\tilde{\boldsymbol{\Sigma}}_{1,2}^{-1}(t)} \right]} \tag{13}$$

for canonically specified latent positions ν_1 and ν_2. In some cases, it is possible to concisely obtain analytic expressions (in t) for both the numerator and denominator. In other cases, this is not possible. A related challenge with respect to Equation (12) is analytically inverting the interpolated block conditional covariance matrices $\boldsymbol{\Sigma}_{kl}(t)$ and $\tilde{\boldsymbol{\Sigma}}_{kl}(t)$. Section 7 provides additional technical details and discussion addressing these issues.

5. Elucidating network structure

5.1 The two-block SBM

Consider the set of two-block SBMs with parameters $\boldsymbol{\pi} \equiv (\pi_1, 1-\pi_1)^\top$ and $\mathbf{B} \in \mathscr{B} := \left\{ \mathbf{B} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} : a, b, c \in (0,1) \right\}$. When $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})^\top$, then $a \geq c$ without loss of generality by symmetry.

Table 1. Summary of embedding performance in Section 5.1.1

$\rho^\star = 1 \Longleftrightarrow \boldsymbol{\psi}_{a,b} = 0$; (ASE = LSE)
$\rho^\star > 1 \Longleftrightarrow \boldsymbol{\psi}_{a,b} > 0$; (ASE > LSE)
$\rho^\star < 1 \Longleftrightarrow \boldsymbol{\psi}_{a,b} < 0$; (ASE < LSE)

In general, for any fixed choice of $\boldsymbol{\pi}$, the class of models $\mathscr{B}$ can be partitioned according to matrix rank, namely

$$\mathscr{B} \equiv \mathscr{B}_1 \bigsqcup \mathscr{B}_2 := \{\mathbf{B} : \text{rank}(\mathbf{B}) = 1; a, b, c \in (0,1)\} \bigsqcup \{\mathbf{B} : \text{rank}(\mathbf{B}) = 2; a, b, c \in (0,1)\}.$$

The collection of sub-models $\mathscr{B}_1$ further decomposes into the disjoint union of the Erdős–Rényi model with homogeneous edge probability $a = b = c \in (0,1)$ and its relative complement in $\mathscr{B}_1$ satisfying the determinant constraint $\det(\mathbf{B}) \equiv ac - b^2 = 0$. These partial sub-models can be viewed as one- and two-dimensional (parameter) regions in the open unit cube, $(0,1)^3$, respectively.

Similarly, the collection of sub-models $\mathscr{B}_2$ further decomposes into the disjoint union of $\mathbb{PD}_2 \cap \mathscr{B}_2$ and $\mathbb{IND}_2 \cap \mathscr{B}_2$, where $\mathbb{PD}_2$ denotes the set of positive definite matrices in $\mathbb{R}^{2\times 2}$ and $\mathbb{IND}_2 := \{\mathbf{B} \in \mathscr{B}_2 : \exists \mathbf{X} \in \mathbb{R}^{2\times 2}, \text{rank}(\mathbf{X}) = 2, \mathbf{B} = \mathbf{X}\mathbf{I}_1^1\mathbf{X}^\top\}$. Here only $\mathbf{I}_0^2 \equiv \mathbf{I}_2$ and $\mathbf{I}_1^1$ are necessary for computing edge probabilities via inner products of the latent positions. Both of these partial sub-models can be viewed as three-dimensional (parameter) regions in $(0,1)^3$.

Remark 3 (*Latent position parametrization).* One might ask whether or not for our purposes there exists a "best" latent position representation for some or even every SBM. To this end, and more generally, for any $K \geq 2$ and $\mathbf{M} \in \mathbb{PD}_K \subset \mathbb{R}^{K\times K}$, there exists a unique lower-triangular matrix $\mathbf{L} \in \mathbb{R}^{K\times K}$ with positive diagonal entries such that $\mathbf{M} = \mathbf{L}\mathbf{L}^\top$ by the Cholesky matrix decomposition. This yields a canonical choice for the matrix of latent positions $\mathbf{X}$ when $\mathbf{B}$ is positive definite. In particular, for $\mathbf{B} \in \mathbb{PD}_2$, then $\mathbf{B} = \mathbf{X}\mathbf{I}_2\mathbf{X}^\top$ with $\mathbf{X} := \begin{bmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & \sqrt{ac-b^2}/\sqrt{a} \end{bmatrix}$. In contrast, for $\mathbf{B} \in \mathbb{IND}_2$, then $\mathbf{B} = \mathbf{X}\mathbf{I}_1^1\mathbf{X}^\top$ with $\mathbf{X} := \begin{bmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & \sqrt{b^2-ac}/\sqrt{a} \end{bmatrix}$, keeping in mind that in this case $b^2 - ac > 0$. The latter factorization may be viewed informally as an indefinite Cholesky decomposition under $\mathbf{I}_1^1$. For the collection of rank one sub-models $\mathscr{B}_1$, the latent positions ν_1 and ν_2 are simply taken to be real-valued scalars.

5.1.1 Homogeneous balanced network structure

We refer to the two-block SBM sub-model with $\mathbf{B} = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$ and $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})^\top$ as the *homogeneous balanced two-block SBM*. The cases when $a > b$, $a < b$, and $a = b$ correspond to the cases when $\mathbf{B}$ is positive definite, indefinite, and reduces to Erdős–Rényi, respectively. The positive definite parameter regime has the network structure interpretation of being *assortative* in the sense that the within-block edge probability a is larger than the between-block edge probability b, consistent with the affinity-based notion of community structure. In contrast, the indefinite parameter regime has the network structure interpretation of being *disassortative* in the sense that between-block edge density exceeds within-block edge density, consistent with the "opposites attract" notion of community structure.

For this SBM sub-model, $\rho^\star$ can be simplified analytically (see Section 7 for additional details) and can be expressed as a translation with respect to the value one, namely

$$\rho^\star \equiv \rho^\star_{a,b} = 1 + \frac{(a-b)^2(3a(a-1) + 3b(b-1) + 8ab)}{4(a+b)^2(a(1-a) + b(1-b))} := 1 + c_{a,b} \times \boldsymbol{\psi}_{a,b} \tag{14}$$

where $\boldsymbol{\psi}_{a,b} := 3a(a-1) + 3b(b-1) + 8ab$ and $c_{a,b} > 0$. By recognizing that $\boldsymbol{\psi}_{a,b}$ functions as a discriminating term, it is straightforward to read off the relative performance of ASE and LSE according to Table 1.

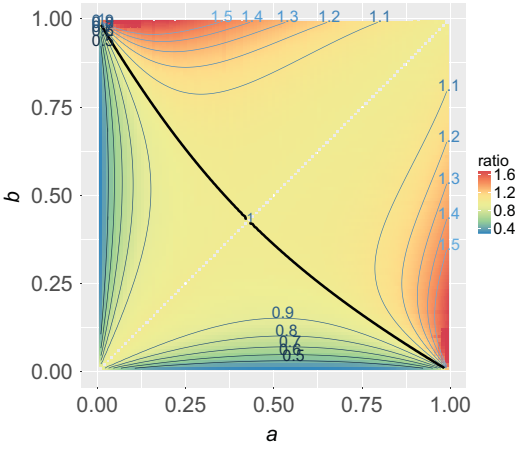

Figure 1. (Color online) The ratio $\rho^\star$ for the homogeneous balanced sub-model in Section 5.1.1. The empty diagonal depicts the Erdős–Rényi model singularity at $a = b$.

Further investigation of Equation (14) leads to the observation that ASE $<$ LSE for all values $0 < b < a \leq \frac{3}{7}$, thereby yielding a parameter region for which LSE dominates ASE. On the other hand, for any fixed $b \in (0, 1)$ there exist values $a_1 < a_2$ such that ASE $<$ LSE when $a = a_1$, whereas ASE $>$ LSE when $a = a_2$. Figure 1 demonstrates that for homogeneous balanced network structure, LSE is preferred to ASE when the entries in $\mathbf{B}$ are sufficiently small, whereas conversely ASE is preferred to LSE when the entries in $\mathbf{B}$ are not too small.

Remark 4 (*Model spectrum and ASE dominance I).* In the current setting, the largest eigenvalue of $\mathbf{B}$ is given by $\lambda_{\max}(\mathbf{B}) = a + b$, hence by Equation (14), $\lambda_{\max}(\mathbf{B}) > 1$ implies ASE $>$ LSE. This observation amounts to a network structure-based (i.e., $\mathbf{B}$-based) spectral sufficient condition for determining when ASE is preferred to LSE.

Remark 5 (*A balanced one-dimensional SBM restricted sub-model).* When $b = 1 - a$, the homogeneous balanced sub-model further reduces to a one-dimensional parameter space setting in which $\rho^\star$ simplifies to the form

$$\rho^\star = 1 + \tfrac{1}{4}(2a - 1)^2. \tag{15}$$

Here $\rho^\star \geq 1$, demonstrating that ASE uniformly dominates LSE for this restricted sub-model. Additionally, it is potentially of interest to note that in this setting the marginal covariance matrices from Theorem 1 for ASE coincide for each block. In contrast, the same behavior is not true for LSE.

5.1.2 Core-periphery network structure

We refer to the two-block SBM sub-model with $\mathbf{B} = \begin{bmatrix} a & b \\ b & b \end{bmatrix}$ and $\boldsymbol{\pi} = (\pi_1, 1 - \pi_1)^\top$ as the *core-periphery two-block SBM*. We explicitly consider the balanced (block size) regime in which $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})^\top$ and an unbalanced regime in which $\boldsymbol{\pi} = (\frac{1}{4}, \frac{3}{4})^\top$. Here, the cases $a > b$, $a < b$, and $a = b$ correspond to the cases when $\mathbf{B}$ is positive definite, indefinite, and reduces to the Erdős–Rényi model, respectively.

For this sub-model, the ratio $\rho^\star$ is not analytically tractable in general. That is to say, simple closed-form solutions do not simultaneously exist for the numerator and denominator in the definition of $\rho^\star$. As such, Figure 2 is obtained numerically by evaluating $\rho^\star$ on a grid of points in $(0, 1)^2$ followed by smoothing.

For $a > b$, graphs generated from this SBM sub-model exhibit the popular interpretation of core-periphery structure in which vertices forming a dense core are attached to surrounding periphery vertices with comparatively smaller edge connectivity. Provided the core is sufficiently

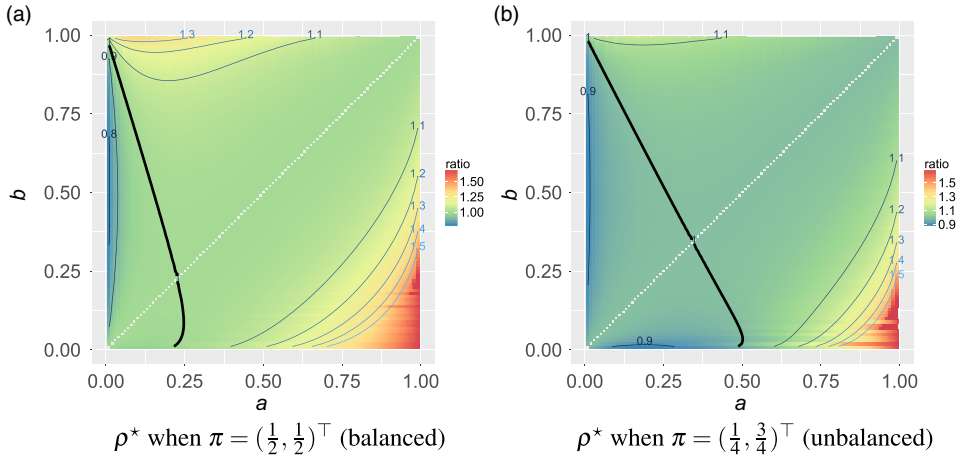

Figure 2. (Color online) The ratio $\rho^\star$ for the core-periphery sub-model in Section 5.1.2. The empty diagonal depicts the Erdős–Rényi model singularity at $a = b$.

dense, namely for $a > \frac{1}{4}$ in the balanced regime and $a > \frac{1}{2}$ in the unbalanced regime, Figure 2 demonstrates that ASE > LSE. Conversely, ASE < LSE uniformly in $0 < b < a$ for small enough values of a in both the balanced and unbalanced regime.

In contrast, when $a < b$, the sub-model produces graphs whose network structure is interpreted as having a comparatively sparse induced subgraph which is strongly connected to all vertices in the graph but for which the subgraph vertices exhibit comparatively weaker connectivity. Alternatively, the second block may itself be viewed as a dense core which is simultaneously densely connected to all vertices in the graph. Figure 2 illustrates that for the balanced regime, LSE is preferred for sparser induced subgraphs. Put differently, for large enough dense core with dense periphery, then ASE is the preferable spectral embedding procedure. Observe that LSE is preferred to ASE in only a relatively small region corresponding approximately to the triangular region where $0 < b < 1 - 4a$, which as a subset of the unit square has area $\frac{1}{8}$. Similar behavior holds for the unbalanced regime for approximately the (enlarged) triangular region of the parameter space where $0 < b < 1 - 2a$, which as a subset of the unit square has area $\frac{1}{4}$.

Figure 2 shows that as π_1 decreases from $\frac{1}{2}$ to $\frac{1}{4}$, LSE is favored in a growing region of the parameter space, albeit still in a smaller region than that for which ASE is preferred (numerically verified, not shown). Together with the observation that LSE dominates in the lower-left corner of the plots in Figure 2, where a and b have small magnitude, we are led to say in summary that LSE favors relatively sparse core-periphery network structure. To reiterate, sparsity is interpreted with respect to the parameters a and b, keeping in mind the underlying simplifying assumption that $n_k = n\pi_k$ for $k = 1, 2$.

Remark 6 (*Model spectrum and ASE dominance II).* For $0 < b < a < 1$, the largest eigenvalue of $\mathbf{B}$ is given by $\lambda_{\max}(\mathbf{B}) = \frac{1}{2}\left(a + b + \sqrt{a^2 - 2ab + 5b^2}\right)$. Numerical evaluation (not shown) demonstrates that $\lambda_{\max}(\mathbf{B}) > \frac{1}{2}$ implies ASE > LSE. Along the same lines as the discussion in Section 5.1.1, this observation provides a network structure (i.e., $\mathbf{B}$-based) spectral sufficient condition for this sub-model for determining the relative embedding performance ASE > LSE.

5.1.3 Two-block rank one sub-model

The sub-model for which $\mathbf{B} = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$ with $a, b, c \in (0, 1)$ and $\det(\mathbf{B}) = 0$ can be re-parameterized according to the assignments $a \mapsto p^2$ and $c \mapsto q^2$, yielding $\mathbf{B} = \left[\begin{smallmatrix} p^2 & pq \\ pq & q^2 \end{smallmatrix}\right]$ with $p, q \in (0, 1)$. Here

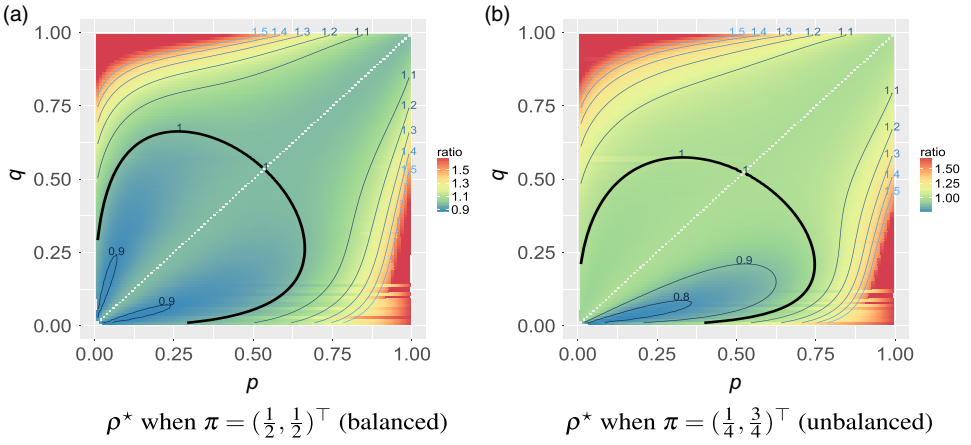

Figure 3. (Color online) The ratio $\rho^\star$ for the two-block rank one sub-model in Section 5.1.3. The empty diagonal depicts the Erdős–Rényi model singularity at $p = q$.

$\text{rank}(\mathbf{B}) = 1$ and $\mathbf{B}$ is positive semidefinite, corresponding to the one-dimensional RDPG model with latent positions given by the scalars p and q with associated probabilities π_1 and π_2, respectively. Explicit computation yields the expression

$$\rho^\star = \frac{\left(\sqrt{p}+\sqrt{q}\right)^2\left(\pi_1 p^2+\pi_2 q^2\right)^2\left(\sqrt{\pi_1 p(1-p^2)+\pi_2 q(1-pq)}+\sqrt{\pi_1 p(1-pq)+\pi_2 q(1-q^2)}\right)^2}{4(\pi_1 p+\pi_2 q)^2\left(\sqrt{\pi_1 p^4(1-p^2)+\pi_2 pq^3(1-pq)}+\sqrt{\pi_1 p^3 q(1-pq)+\pi_2 q^4(1-q^2)}\right)^2} \tag{16}$$

whereby $\rho^\star$ is given as an explicit, closed-form function of the parameter values p, q, and π_1, with $\pi_2 = 1 - \pi_1$. The simplicity of this sub-model together with its analytic tractability with respect to both $\mathbf{B}$ and $\boldsymbol{\pi}$ makes it particularly amenable to study for the purpose of elucidating network structure. In the following, consideration of this sub-model further illustrates the relationship between sparsity (based on edge probabilities) and relative embedding performance.

Figure 3 demonstrates how LSE favors sparse graphs in the sense of the edge probabilities, p and q, as well as how relative embedding performance changes in light of (un)balanced block sizes, reflected by π_1. Here the underlying $\mathbf{B}$ matrix is always positive semidefinite, and both cases $p > q$ and $p < q$ correspond to a modified notion of core-periphery structure. In particular, when $p > q$, then $p^2 > pq > q^2$, yielding a hierarchy of core-periphery structure when passing from vertices that are both in block one to vertices that are in different blocks and finally to vertices that are both in block two. Note the similar behavior in the bottom-right triangular regions in Figure 3(a) and (b) and in the same bottom-right triangular region in Figure 2.

Remark 7 (*The two-block polynomial p SBM restricted sub-model).* Consider the restricted sub-model in which $\mathbf{B} = \begin{bmatrix} p^2 & p^{\gamma+1} \\ p^{\gamma+1} & p^{2\gamma} \end{bmatrix}$, for $\gamma > 1$ and $\pi_1 \in (0,1)$. Provided $\gamma \gg 1$ and π_1 is fixed, then $\rho^\star$ in Equation (16) satisfies the approximate behavior

$$\rho^\star \approx \frac{\left(1+\sqrt{1-p^2}\right)^2}{4(1-p^2)}. \tag{17}$$

The above approximation exceeds the value one since $1 > \sqrt{1-p^2}$ for all $p \in (0,1)$ and is simultaneously agnostic with respect to π_1. Moreover, for large values of γ, the block edge probability matrix is approximately of the form $\mathbf{B} \approx \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix}$ with $p_1 \gg p_2 \approx p_3$, where p_2 and p_3 are very small. This restricted sub-model can therefore be viewed as exhibiting an extremal version

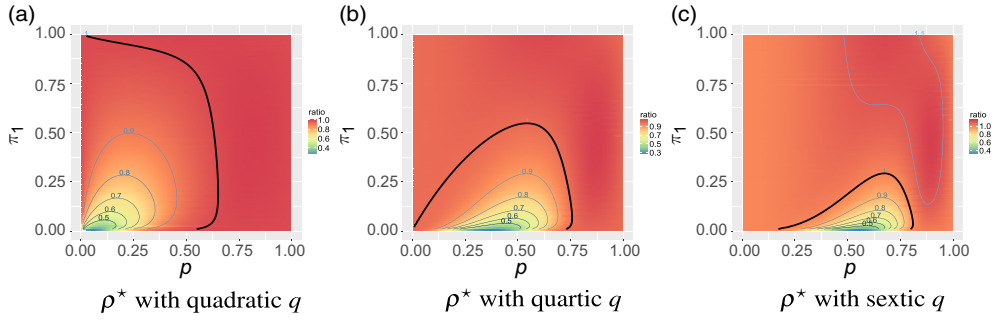

Figure 4. (Color online) The ratio $\rho^\star$ for $p, \pi_1 \in (0, 1)$ when $q = p^\gamma$, $\gamma \in \{2, 4, 6\}$ in Section 5.1.3.

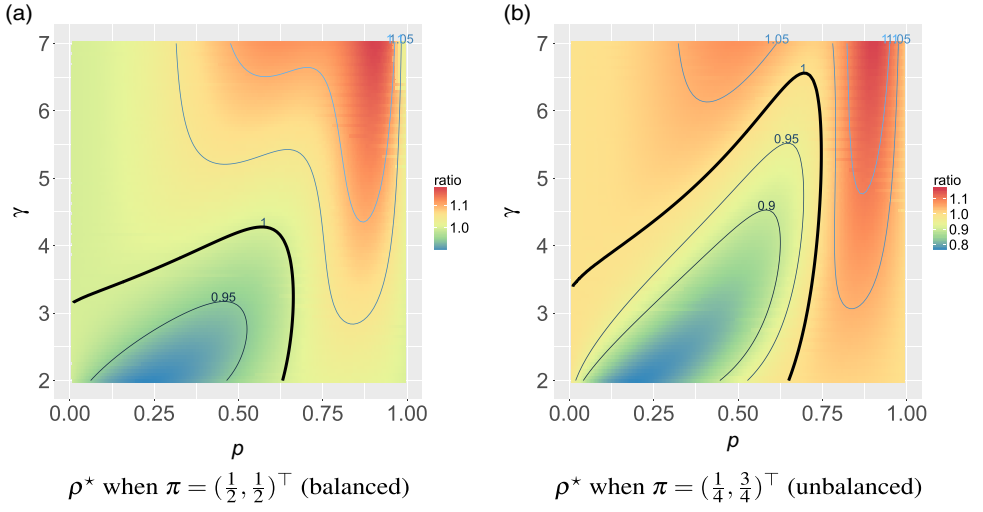

Figure 5. (Color online) The ratio $\rho^\star$ for $p \in (0, 1)$, $\gamma \in [2, 7]$ when $q = p^\gamma$ in Section 5.1.3.

of core-periphery structure corresponding to the extremal regions in Figure 2, where ASE is preferred.

In Figure 4, the progression from left to right corresponds to tending toward the approximation presented in Equation (17). For larger values of γ when $q = p^\gamma$ (not shown), the region where ASE > LSE continues to expand. We do not discuss or pursue the taking of limits within the parameter space(s) in light of degenerate boundary value behavior and in order to avoid possible misinterpretation.

Figure 5 offers a different perspective in which γ is allowed to vary continuously for both the balanced and the unbalanced regime. As in Figure 3, Figure 5 demonstrates that LSE is preferred for network structure wherein the block with comparatively higher edge probability exhibits smaller block membership size.

5.1.4 Full-rank two-block SBMs

This section presents a macroscopic view of full-rank two-block SBMs with $\mathbf{B} = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$, $(a, b, c) \in (0, 1)^3$, for the regimes $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})^\top$ and $\boldsymbol{\pi} = (\frac{1}{4}, \frac{3}{4})^\top$. The parameter space is partitioned via the latent space geometry of $\mathbf{B}$, namely according to whether $\mathbf{B}$ is either positive definite or indefinite.

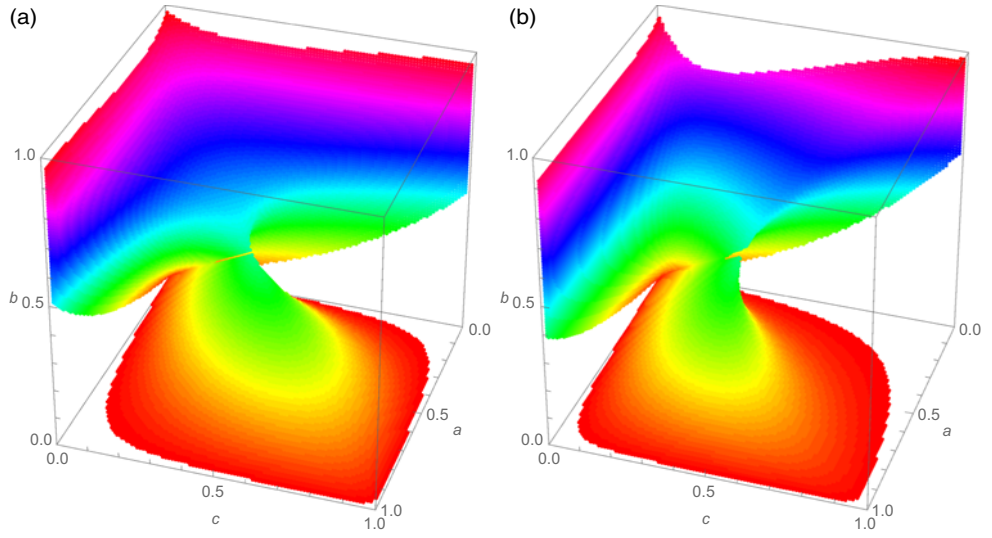

Figure 6. (Color online) The parameter region where ASE $<$ LSE for full rank $\mathbf{B}$ in Section 5.1.4. The plots depict numerical evaluations of $\rho^\star$ for $a, b, c \in [0.01, 0.99]$ with step size 0.01. The color scale corresponds to b values for the plotted triples (a, b, c).

Figure 6(a) and (b) each present a three-dimensional view of the parameter space region where ASE $<$ LSE. The separate positive definite and indefinite parameter regions exhibiting ASE $<$ LSE can be seen extending from faces of the unit cube. Specifically, the conic-like region rising up from the $b = 0$ face corresponds to $\mathbf{B}$ for which $\mathbf{B} \in \mathbb{PD}_2$, whereas the hyperbolic-like regions extending from the $a = 0$ and $c = 0$ faces correspond to $\mathbf{B}$ for which $\mathbf{B} \in \mathbb{IND}_2$.

For the balanced case reflected in Figure 6(a), let $a \geq c$ without loss of generality by symmetry, and hence $\rho^\star$ is symmetric about the plane defined by $a = c$. For the unbalanced case shown in Figure 6(b), symmetry no longer holds, and geometric warping behavior can be seen with respect to the $a = c$ plane. Figure 7(a) and (b) provides a bird's-eye view of the three-dimensional positive definite parameter region from the top-down vantage point of $b = 1$. The latter provides another view of the warping phenomenon observed for $\boldsymbol{\pi} = (\frac{1}{4}, \frac{3}{4})^\top$ that holds in general for all unbalanced regimes.

In both block size regimes depicted in Figure 6, the colored parameter region occupies less than one fourth of the unit cube volumetrically, thereby quantitatively providing a coarse overall sense in which ASE is to be preferred to LSE for numerous two-block SBMs.

5.2 The K-blockmodel with homogeneous balanced affinity network structure

This section generalizes the analysis in Section 5.1.1 to the setting of K-block homogeneous balanced affinity SBMs, where $B_{ij} = a$ for all $i = j$, $B_{ij} = b$ for all $i \neq j$, $0 < b < a < 1$, and $\pi_i = \frac{1}{K}$ for $1 \leq i \leq K$.

Theorem 3. *For K-block homogeneous balanced affinity SBMs as in Section 5.2, the ratio $\rho^\star$ in Equation (12) can be expressed analytically as*

$$\rho^\star = 1 + \frac{(a-b)^2(3a(a-1)+3b(b-1)(K-1)+4abK)}{4(a+(K-1)b)^2(a(1-a)+b(1-b))} := 1 + c_{a,b,K} \times \boldsymbol{\psi}_{a,b,K} \tag{18}$$

where $\boldsymbol{\psi}_{a,b,K} := 3a(a-1) + 3b(b-1)(K-1) + 4abK$ and $c_{a,b,K} > 0$.

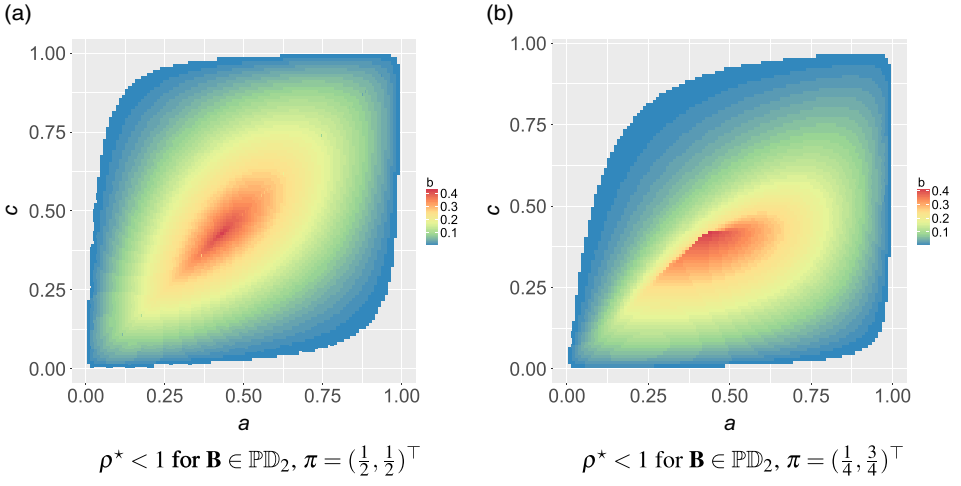

Figure 7. A top-down projected view of the positive definite region where ASE < LSE in Section 5.1.4, with a, b, and c corresponding to length, depth, and width, respectively. The plots depict numerical evaluations of $\rho^\star$ for $a, b, c \in [0.01, 0.99]$ with step size 0.01.

As in Table 1, the function $\boldsymbol{\psi}_{a,b,K}$ is the discriminating term that explicitly characterizes the relative performance of ASE and LSE.

Here $\boldsymbol{\psi}_{a,b,K}$ satisfies $(4ab - 3(a - b^2))K < \boldsymbol{\psi}_{a,b,K} < (4ab)K$, and there are explicit constants $c^{(1)}_{a,b}$ and $c^{(2)}_{a,b}$ depending only on a and b such that $\frac{1}{K}c^{(1)}_{a,b} < c_{a,b,K} \times \boldsymbol{\psi}_{a,b,K} < \frac{1}{K}c^{(2)}_{a,b}$. Viewing a and b as fixed, we can summarize the above behavior using the asymptotic notation $c_{a,b,K} \times \boldsymbol{\psi}_{a,b,K} = \Theta_{a,b}(\frac{1}{K})$. Equation (18) can therefore be summarized in terms of K as

$$\rho^\star = 1 + \Theta_{a,b}\left(\frac{1}{K}\right) \tag{19}$$

demonstrating that $\rho^\star \to 1$ as $K \to \infty$. In words, for the class of SBMs under consideration, ASE and LSE in a sense have asymptotically (in K) equivalent embedding performance (via $\rho^\star$). This amounts to a statement concerning a sequence of models with a necessarily growing number of vertices in order to ensure the underlying assumption of equal block sizes.

Rewriting the level-set $\boldsymbol{\psi}_{a,b,K} = 0$, which holds if and only if $\rho^\star = 1$, yields the equation

$$\left(\frac{1-a}{b}\right)\frac{1}{K} + \left(\frac{1-b}{a}\right)\frac{K-1}{K} = \frac{4}{3} \tag{20}$$

together with the observation that ASE > LSE (resp., ASE < LSE) when the left-hand side of Equation (20) is less than (resp., greater than) the value $\frac{4}{3}$. The above equation, perhaps interestingly, depicts a convex combination (in terms of K) involving the variables $\frac{1-a}{b}$ and $\frac{1-b}{a}$, where the value $\frac{4}{3}$ is interpretable as a Chernoff-based information theoretic threshold.

The observation that $\boldsymbol{\psi}_{a,b,K} > (4ab - 3(a - b^2))K$ in the context of Equation (18) implies a sufficient condition for determining a parameter region in which ASE > LSE *uniformly in K*. Specifically, the condition $(4ab - 3(a - b^2)) > 0$, equivalently written as $\frac{a-b^2}{ab} < \frac{4}{3}$, ensures that $\boldsymbol{\psi}_{a,b,K} > 0$ and hence that $\rho^\star > 1$.

Remark 8 (*Detectability and phase transitions in random graph models).* With respect to the random graph literature, the setting considered in this paper corresponds to a strong consistency regime (i.e., exact recovery) in which the block membership of each individual vertex is recovered almost surely for graphs on n vertices with $n \to \infty$. For different regimes where

edge probabilities are allowed to decrease as a function of n, numerous deep and fascinating detectability and phase transition phenomena are known, some of which also employ Chernoff divergence and related considerations (Abbe, 2018). In the context of homogeneous balanced affinity SBMs, the signal-to-noise ratio (SNR) $:= \frac{(a-b)^2}{K(a+(K-1)b)}$ has been shown to function as an important information-theoretic quantity. Here too the SNR appears, albeit with respect to $c_{a,b,K}$, in the sense that

$$c_{a,b,K} := \frac{(a-b)^2}{4(a+(K-1)b)^2(a(1-a)+b(1-b))} \equiv \left(\frac{(a-b)^2}{K(a+(K-1)b)}\right) \tilde{c}_{a,b,K} = \text{SNR} \times \tilde{c}_{a,b,K}$$

for some constant $\tilde{c}_{a,b,K} > 0$. Perhaps more interestingly,

$$c_{a,b,K} \equiv \frac{1}{4}\left(\frac{\lambda_{\min}(\mathbf{B}(K))}{\lambda_{\max}(\mathbf{B}(K))}\right)^2 \left(\frac{1}{\sigma^2(B_{11}(K))+\sigma^2(B_{22}(K))}\right)$$

where $\sigma^2(B_{ij}(K))$ is the edge variance corresponding to a pair of vertices in blocks i and j, together with eigenvalues $\lambda_{\min}(\mathbf{B}(K)) = a - b$ and $\lambda_{\max}(\mathbf{B}(K)) = a + (K-1)b$. The constant factor $\frac{1}{4}$ could just as easily be absorbed by redefining $\boldsymbol{\psi}_{a,b,K}$. It may well prove fruitful to further investigate these observations in light of existing results in the literature.

6. Discussion and conclusions

Loosely speaking, LSE may be viewed as a degree-normalized version of ASE in light of Equation (3). As such, the analysis presented in this paper seeks to understand normalization in the context of spectral methods (Sarkar & Bickel, 2015; von Luxburg, 2007). Moreover, our work together with Rubin-Delanchy et al. (2017) addresses network models exhibiting indefinite geometry, an area that has received comparatively limited attention in the statistical network analysis literature. The ability of indefinite modeling considerations to reflect widely observed disassortative community structure is encouraging and suggests future research activity in this and related directions.

Core-periphery network structure, broadly construed, is considered to be ubiquitous in real-world networks (Csermely et al., 2013; Holme, 2005; Leskovec et al., 2009), though it must be emphasized that usage of the term *core-periphery* is at times ambiguous and inconsistent in the literature. Our usage is similar to Jeub et al. (2015) and Zhang et al. (2015) wherein core-periphery structure is taken to mean that the edge probabilities in the SBM $\mathbf{B}$ matrix satisfy the relationship $a > b > c$. More specifically, in Section 5.1.2, we first use *core-periphery* to describe the setting where $b = c$. This formulation, which conveniently enables two-dimensional visualizations for fixed block membership vectors, also captures different versions of idealized core-periphery structure as proposed in Borgatti and Everett (2000), particularly when $1 \approx a > b \approx 0$ and when $0 \approx a < b \approx 1$. Subsequently in Section 5.1.3, we investigate restricted rank SBM sub-models and pay particular attention to parameter regions of the abovementioned core-periphery form $a > b > c$. At a more macroscopic level, Figure 6 illustrates that ASE is most often preferred to LSE in this parameter region, provided the probabilities (parameters) a, b, and c are not too small. With this understanding and the ability of the SBM to serve as a building block for hierarchically modeling complex network structure, our findings pertaining to spectral embedding for core-periphery structure may be of particular interest.

The approach adopted in this paper is asymptotic and theoretical in nature. Our findings mesh with the contemporary, related work in Priebe et al. (2019), which carries over the ideas discussed here to real-data connectome analysis and empirically driven simulation examples. In terms of future work, it would be interesting to conduct Chernoff-based analysis for other matrix-valued graph representations. Another valuable future contribution would be to obtain precise finite sample results for describing spectral embedding performance.

To reiterate, this paper examines the information-theoretic relationship between the performance of two competing, widely popular graph embeddings and subsequent vertex clustering with an eye toward underlying network model structure. The findings presented in Section 5 support the claim that, for *sparsity* interpreted as $\mathbf{B}$ having entries that are small, loosely speaking, "LSE favors relatively sparse graphs, whereas ASE favors not-too-sparse graphs." Moreover, our results provide evidence in support of the claim that "ASE favors certain core-periphery network structure." Of course, caution must be exercised when making such general assertions, since the findings in this paper demonstrate intricate and nuanced functional relationships linking spectral embedding performance to network model structure. Nevertheless, we believe such summarative statements are both faithful and useful for conveying a high-level, macroscopic overview of the investigation presented in this work.

7. Supplementary material

7.1 Latent position geometry

All SBMs in Definition 1 can be formulated as instantiations of GRDPG models possessing inherent latent position (vector) structure. Earlier observations for the two-block SBM in Section 5 are summarized in the following table, for which the implicit underlying vector $\boldsymbol{\pi}$ may be viewed as an additional parameter space dimension that weights the latent positions ν_1 and ν_2 by π_1 and π_2, respectively.

Model geometry:	Canonical latent positions:
Positive definite $\mathbf{B}(a,b,c)$	$\boldsymbol{\nu}_1 = (\sqrt{a}, 0)^\top$, $\boldsymbol{\nu}_2 = (b/\sqrt{a}, \sqrt{ac-b^2}/\sqrt{a})^\top$ in $\mathbb{R}^2$
Indefinite $\mathbf{B}(a,b,c)$	$\boldsymbol{\nu}_1 = (\sqrt{a}, 0)^\top$, $\boldsymbol{\nu}_2 = (b/\sqrt{a}, \sqrt{b^2-ac}/\sqrt{a})^\top$ in $\mathbb{R}^2$
Rank one $\mathbf{B}(p^2, pq, q^2)$	$\boldsymbol{\nu}_1 = p$, $\boldsymbol{\nu}_2 = q$ in $\mathbb{R}$

For the homogeneous balanced affinity two-block network structure investigated in Section 5.1.1, the latent position geometry can be equivalently reparameterized as two vectors on the circle of radius $r := \sqrt{a}$ separated by the angle $\theta := \arccos(b/a)$. This behavior generalizes to the homogeneous balanced affinity K-blockmodel. More specifically, when $\mathbf{B} \equiv \mathbf{B}(K) \in (0,1)^{K\times K}$ has value a on the main diagonal and value b on the off-diagonal with $0 < b < a < 1$, we can write $\mathbf{B} = \mathbf{X}\mathbf{X}^\top$ via the Cholesky decomposition, with $\mathbf{X} = [x_1|x_2|\ldots|x_K]^\top$. For $K = 2, 3, 4$, the matrix $\mathbf{X}(K)$ is given by

$$\mathbf{X}(2) := \begin{bmatrix} \sqrt{a} & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{\frac{(a-b)(a+b)}{a}} \end{bmatrix} \tag{21}$$

$$\mathbf{X}(3) := \begin{bmatrix} \sqrt{a} & 0 & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{\frac{(a-b)(a+b)}{a}} & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{\frac{(a-b)(a+b)}{a}}\,\frac{b}{a+b} & \sqrt{\frac{(a-b)(a+2b)}{a+b}} \end{bmatrix} \tag{22}$$

$$\mathbf{X}(4) := \begin{bmatrix} \sqrt{a} & 0 & 0 & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{\frac{(a-b)(a+b)}{a}} & 0 & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{\frac{(a-b)(a+b)}{a}}\,\frac{b}{a+b} & \sqrt{\frac{(a-b)(a+2b)}{a+b}} & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{\frac{(a-b)(a+b)}{a}}\,\frac{b}{a+b} & \sqrt{\frac{(a-b)(a+2b)}{a+b}}\,\frac{b}{a+2b} & \sqrt{\frac{(a-b)(a+3b)}{a+2b}} \end{bmatrix} \tag{23}$$

By induction, for $K \geq 3$, the entries of the vector x_K are given by

$$x_K = \left(x_{K-1}^{(1)}, x_{K-1}^{(2)}, \ldots, x_{K-1}^{(K-2)}, \left(\tfrac{b}{a+(K-2)b}\right) x_{K-1}^{(K-1)}, \sqrt{\tfrac{(a-b)(a+(K-1)b)}{a+(K-2)b}} \right)^{\top} \in \mathbb{R}^K. \tag{24}$$

Only $\mathbf{I}_0^K$ and $\mathbf{I}_{K-1}^1$ are necessary with respect to combining possible inner products on account of the sign-flip involving $a - b$. Beginning with the second row in each of the $\mathbf{X}$ matrices, the first column of each matrix can be written in the more illuminating form $\sqrt{a}\frac{b}{a}$.

For this specific K-blockmodel, symmetry with respect to equally spaced vectors on the $\sqrt{a}$-radius sphere in $\mathbb{R}^K$ together with block membership balancedness translates into shared covariance structure such that Equation (12) reduces to Equation (13). The first two rows of $\mathbf{X}$ are ideal candidates to serve as canonical latent positions for subsequent computation, since these vectors are maximally sparse in the sense of having the fewest nonzero entries and merely become zero-inflated as a function of K. These geometric considerations are crucial in the subsequent proof of Theorem 3.

7.2 Analytic derivations for the two-block SBM

The value of $\rho^\star$ in Equation (14) for the homogeneous balanced two-block SBM can be computed by brute force; however, such an approach offers only limited insight and understanding of how the covariance structure in Theorems 1 and 2 interact to yield differences in relative spectral embedding performance as measured via Chernoff information. This section offers a different approach to understanding $\rho^\star$ as a covariance-based spectral quantity.

The following lemma is a general matrix analysis observation that establishes a correspondence between the inverse of a convex combination of 2×2 matrices and the inverses of the original 2×2 matrices. The proof of Lemma 4 follows directly from elementary computations and is therefore omitted. Extending Lemma 4 to $n \times n$ invertible matrices is intractable in general.

Lemma 4. *Let $\mathbf{M}_0, \mathbf{M}_1 \in \mathbb{R}^{2\times 2}$ be two invertible matrices. For each $t \in [0, 1]$ define the matrix $\mathbf{M}_t := (1-t)\mathbf{M}_0 + t\mathbf{M}_1$. Provided $\mathbf{M}_t$ is invertible, then the inverse matrix $\mathbf{M}_t^{-1}$ can be expressed as*

$$\mathbf{M}_t^{-1} \equiv \frac{(1-t)\mathbf{M}_0^{-1} + \det(\mathbf{M}_1\mathbf{M}_0^{-1})t\mathbf{M}_1^{-1}}{\det(\mathbf{M}_1\mathbf{M}_0^{-1})t^2 + \operatorname{tr}(\mathbf{M}_1\mathbf{M}_0^{-1})t(1-t) + (1-t)^2}. \tag{25}$$

If, in the context of Lemma 4, $\det(\mathbf{M}_1\mathbf{M}_0^{-1}) = 1$, then Equation (25) simplifies to

$$\mathbf{M}_t^{-1} \equiv \frac{(1-t)\mathbf{M}_0^{-1} + t\mathbf{M}_1^{-1}}{t^2 + \operatorname{tr}(\mathbf{M}_1\mathbf{M}_0^{-1})t(1-t) + (1-t)^2}.$$

which is nearly a convex combination of the inverse matrices $\mathbf{M}_0^{-1}$ and $\mathbf{M}_1^{-1}$ modulo division by a degree-two polynomial in the parameter t. If, in addition, $\operatorname{tr}(\mathbf{M}_1\mathbf{M}_0^{-1}) \neq -2$ (which always holds when $\mathbf{M}_0$ and $\mathbf{M}_1$ are both positive definite), then the inverse matrix at the value $t = \frac{1}{2}$ further simplifies to

$$\mathbf{M}_{1/2}^{-1} \equiv \left(\frac{2}{2+\operatorname{tr}(\mathbf{M}_1\mathbf{M}_0^{-1})}\right)\left(\mathbf{M}_0^{-1} + \mathbf{M}_1^{-1}\right) \tag{26}$$

For the homogeneous balanced two-block SBM considered in Section 5.1.1, one can explicitly check that the above $\det(\cdot)$ and $\operatorname{tr}(\cdot)$ conditions are satisfied. Moreover, the value $t^\star = \frac{1}{2}$ achieves the supremum in both the numerator and denominator of $\rho^\star$ in Equation (12). With these observations in hand, it follows by subsequent computations that for both the positive definite and indefinite regimes,

$$\begin{aligned}
\rho^\star &= \frac{\|\nu_1 - \nu_2\|^2_{\boldsymbol{\Sigma}^{-1}_{1,2}(1/2)}}{\|\tilde{\nu}_1 - \tilde{\nu}_2\|^2_{\tilde{\boldsymbol{\Sigma}}^{-1}_{1,2}(1/2)}} \\
&= \left(\frac{\left(\frac{2}{2+\mathrm{tr}(\boldsymbol{\Sigma}(\nu_1)\boldsymbol{\Sigma}^{-1}(\nu_2))} \right)}{\left(\frac{2}{2+\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}(\nu_1)\tilde{\boldsymbol{\Sigma}}^{-1}(\nu_2))} \right)} \right) \times \left(\frac{(\nu_1 - \nu_2)^\top (\boldsymbol{\Sigma}^{-1}(\nu_1) + \boldsymbol{\Sigma}^{-1}(\nu_2))(\nu_1 - \nu_2)}{(\tilde{\nu}_1 - \tilde{\nu}_2)^\top (\tilde{\boldsymbol{\Sigma}}^{-1}(\nu_1) + \tilde{\boldsymbol{\Sigma}}^{-1}(\nu_2))(\tilde{\nu}_1 - \tilde{\nu}_2)} \right) \\
&= \left(\frac{2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}(\nu_1)\tilde{\boldsymbol{\Sigma}}^{-1}(\nu_2))}{2 + \mathrm{tr}(\boldsymbol{\Sigma}(\nu_1)\boldsymbol{\Sigma}^{-1}(\nu_2))} \right) \times 1 \\
&= 1 + \frac{\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}(\nu_1)\tilde{\boldsymbol{\Sigma}}^{-1}(\nu_2)) - \mathrm{tr}(\boldsymbol{\Sigma}(\nu_1)\boldsymbol{\Sigma}^{-1}(\nu_2))}{2 + \mathrm{tr}(\boldsymbol{\Sigma}(\nu_1)\boldsymbol{\Sigma}^{-1}(\nu_2))} \\
&= 1 + \frac{(a-b)^2(3a(a-1) + 3b(b-1) + 8ab)}{4(a+b)^2(a(1-a) + b(1-b))}.
\end{aligned}$$

7.3 Proof of Theorem 3

This section is dedicated to proving Theorem 3 for $K \geq 2$ block SBMs exhibiting homogeneous balanced affinity structure. The proof is divided into two parts which separately evaluate the suprema in the numerator and denominator of $\rho^\star$ in Equation (12). By invoking underlying symmetries in latent space and the covariance structure of the ASE and LSE limit results, respectively, we shall leverage the (considerably simpler) ASE computations (numerator) when working with LSE (denominator). Further simplifying the numerator and denominator yields the more easily interpretable (shifted) expression of $\rho^\star$ provided in Equation (18).

Proof: First recall the discussion of latent space geometry in Section 7.1, specifically that for the homogeneous balanced affinity K-block SBM, the canonical latent positions can be arranged row-wise as a lower-triangular matrix $\mathbf{X}$ where each latent position vector has norm $\sqrt{a}$ and each pair of distinct latent position vectors has common inner-product b. This rotational symmetry implies rotational symmetry for the block-conditional covariance matrices in Theorems 1 and 2, and as such, the formulation of $\rho^\star$ in Equation (18) can be reduced to simply working with the latent position pair $\{\nu_1, \nu_2\}$ without loss of generality. This pair is attractive, since the nonzero entries of these vectors remain unchanged for all $K \geq 2$. Moreover, one need only work with the standard inner product since $d^- = 0$.

7.3.1 Proof of Theorem 3: ASE (numerator)

Let $g(x, X_1) := \langle x, X_1 \rangle (1 - \langle x, X_1 \rangle)$ and for $0 < t < 1$ define $g_t(x_1, x_2, X_1) := tg(x_1, X_1) + (1-t)g(x_2, X_1)$. By Theorem 1, $\boldsymbol{\Sigma}(x) = \boldsymbol{\Delta}^{-1}\mathbb{E}[g(x, X_1)X_1X_1^\top]\boldsymbol{\Delta}^{-1}$, and therefore $\boldsymbol{\Sigma}_{1,2}(t) := t\boldsymbol{\Sigma}(\nu_1) + (1-t)\boldsymbol{\Sigma}(\nu_2) = \boldsymbol{\Delta}^{-1}\mathbb{E}[g_t(\nu_1, \nu_2, X_1)X_1X_1^\top]\boldsymbol{\Delta}^{-1}$. Evaluating the inner expectation yields

$$\begin{aligned}
\mathbb{E}[g_t(\nu_1, \nu_2, X_1)X_1X_1^\top] &= \sum_{i=1}^{K} \tfrac{1}{K}(t\langle \nu_1, \nu_i \rangle (1 - \langle \nu_1, \nu_i \rangle) + (1-t)\langle \nu_2, \nu_i \rangle (1 - \langle \nu_2, \nu_i \rangle))\nu_i\nu_i^\top \\
&= b(1-b)\boldsymbol{\Delta} + \left(\tfrac{a(1-a) - b(1-b)}{K} \right) \left[t\nu_1\nu_1^\top + (1-t)\nu_2\nu_2^\top \right] \\
&= b(1-b)\boldsymbol{\Delta} + \mathbf{N}(c_0\mathbf{D}_t)\mathbf{N}^\top
\end{aligned}$$

where $\mathbf{N} := [\nu_1 | \nu_2] \in \mathbb{R}^{K \times 2}$, $c_0 := \left(\frac{a(1-a) - b(1-b)}{K} \right)$, and $\mathbf{D}_t := \mathrm{diag}(t, 1-t)$. Clearly $c_0\mathbf{D}_t$ is invertible, as is $\boldsymbol{\Delta}$ since the underlying distribution F is nondegenerate. Moreover, $\mathbf{X}$ is also invertible

since the K-blockmodel under consideration is also rank K. The relation $\mathbf{X}^\top\mathbf{X} = K\boldsymbol{\Delta}$ implies $\boldsymbol{\Delta}^{-1} = K\mathbf{X}^{-1}(\mathbf{X}^\top)^{-1}$ and therefore $\mathbf{X}\boldsymbol{\Delta}^{-1}\mathbf{X}^\top = K\mathbf{I}$, so $\nu_i^\top\boldsymbol{\Delta}^{-1}\nu_j = K\mathbb{I}_{ij}$, where $\mathbb{I}_{ij}$ denotes the indicator function for indices i and j. Thus, $(c_0\mathbf{D}_t)^{-1} + \frac{1}{b(1-b)}\mathbf{N}^\top\boldsymbol{\Delta}^{-1}\mathbf{N} = (c_0\mathbf{D}_t)^{-1} + \frac{K}{b(1-b)}\mathbf{I}$, which is also invertible. By an application of the Sherman–Morrison–Woodbury matrix inversion formula (Horn and Johnson (2012), Section 0.7.4), then

$$\begin{aligned}\mathbb{E}[\,g_t(\nu_1,\nu_2,X_1)X_1X_1^\top]^{-1} &= \left(b(1-b)\boldsymbol{\Delta} + \mathbf{N}(c_0\mathbf{D}_t)\mathbf{N}^\top\right)^{-1}\\ &= \left(\tfrac{1}{b(1-b)}\right)\boldsymbol{\Delta}^{-1} - \left(\tfrac{1}{b(1-b)}\right)^2\boldsymbol{\Delta}^{-1}\mathbf{N}\left(\tfrac{1}{c_0}\mathbf{D}_t^{-1} + \tfrac{K}{b(1-b)}\mathbf{I}\right)^{-1}\mathbf{N}^\top\boldsymbol{\Delta}^{-1}\end{aligned}$$

For $\nu := \nu_1 - \nu_2 = \left(\frac{a-b}{\sqrt{a}}, -\sqrt{\frac{(a-b)(a+b)}{a}}, 0, \ldots, 0\right)^\top \in \mathbb{R}^K$, then $\nu^\top\boldsymbol{\Delta}\nu = \frac{2}{K}(a-b)^2$ and $\mathbf{N}^\top\nu = (a-b)(1,-1)^\top \in \mathbb{R}^2$. These observations together with subsequent computations yield the following chain of equalities:

$$\begin{aligned}\|\nu\|^2_{\boldsymbol{\Sigma}_{1,2}^{-1}(t)} &= \nu^\top\left(\boldsymbol{\Delta}^{-1}\mathbb{E}[\,g_t(\nu_1,\nu_2,X_1)X_1X_1^\top]\boldsymbol{\Delta}^{-1}\right)^{-1}\nu\\
&= \nu^\top(\boldsymbol{\Delta}\mathbb{E}[\,g_t(\nu_1,\nu_2,X_1)X_1X_1^\top]^{-1}\boldsymbol{\Delta})\nu\\
&= \nu^\top\boldsymbol{\Delta}\left(\tfrac{1}{b(1-b)}\boldsymbol{\Delta}^{-1} - \left(\tfrac{1}{b(1-b)}\right)^2\boldsymbol{\Delta}^{-1}\mathbf{N}\left(\tfrac{1}{c_0}\mathbf{D}_t^{-1} + \tfrac{K}{b(1-b)}\mathbf{I}\right)^{-1}\mathbf{N}^\top\boldsymbol{\Delta}^{-1}\right)\boldsymbol{\Delta}\nu\\
&= \nu^\top\left(\tfrac{1}{b(1-b)}\boldsymbol{\Delta} - \left(\tfrac{1}{b(1-b)}\right)^2\mathbf{N}\left(\tfrac{1}{c_0}\mathbf{D}_t^{-1} + \tfrac{K}{b(1-b)}\mathbf{I}\right)^{-1}\mathbf{N}^\top\right)\nu\\
&= \left(\tfrac{1}{b(1-b)}\right)\nu^\top\boldsymbol{\Delta}\nu - \left(\tfrac{1}{b(1-b)}\right)^2\nu^\top\mathbf{N}\left(\tfrac{1}{c_0}\mathbf{D}_t^{-1} + \tfrac{K}{b(1-b)}\mathbf{I}\right)^{-1}\mathbf{N}^\top\nu\\
&= \left(\tfrac{2(a-b)^2}{b(1-b)K}\right) - \left(\tfrac{a-b}{b(1-b)}\right)^2(1,-1)\left(\tfrac{1}{c_0}\mathbf{D}_t^{-1} + \tfrac{K}{b(1-b)}\mathbf{I}\right)^{-1}(1,-1)^\top\\
&= \left(\tfrac{2(a-b)^2}{b(1-b)K}\right) - \left(\tfrac{a-b}{b(1-b)}\right)^2\operatorname{tr}\left(\left(\tfrac{1}{c_0}\mathbf{D}_t^{-1} + \tfrac{K}{b(1-b)}\mathbf{I}\right)^{-1}\right)\\
&= \left(\tfrac{2(a-b)^2}{b(1-b)K}\right) - \left(\tfrac{a-b}{b(1-b)}\right)^2\left(\tfrac{(a(1-a)-b(1-b))b(1-b)t}{((a(1-a)-b(1-b))t+b(1-b))K} + \tfrac{(a(1-a)-b(1-b))b(1-b)(1-t)}{((a(1-a)-b(1-b))(1-t)+b(1-b))K}\right)\\
&= \tfrac{(a-b)^2(a(a-1)+b(b-1))}{(a(1-a)+(a(a-1)-b(b-1))t)(b(b-1)+(a(a-1)-b(b-1))t)K}\end{aligned}$$

In particular,

$$\sup_{t\in(0,1)}\left[t(1-t)\|\nu\|^2_{\boldsymbol{\Sigma}_{1,2}^{-1}(t)}\right] = \tfrac{1}{K}\tfrac{(a-b)^2}{a(1-a)+b(1-b)} \tag{27}$$

where by underlying symmetry the supremum is achieved at $t^\star = \frac{1}{2}$ over the entire parameter region $0 < b < a < 1$.

7.3.2 Proof of Theorem 3: LSE (denominator)

Recall that for this model $\mathbf{I}_{d^-}^{d^+} \equiv \mathbf{I}_d$ since $d^- = 0$. From Theorem 2 for LSE, the block conditional covariance matrix for each latent position x can be written in the modified form

$$\tilde{\boldsymbol{\Sigma}}(x) = \mathbb{E}\left[\left(\tfrac{g(x,X_1)}{\langle x,\mu\rangle}\right)\left(\tfrac{\tilde{\boldsymbol{\Delta}}^{-1}X_1}{\langle X_1,\mu\rangle} - \tfrac{x}{2\langle x,\mu\rangle}\right)\left(\tfrac{\tilde{\boldsymbol{\Delta}}^{-1}X_1}{\langle X_1,\mu\rangle} - \tfrac{x}{2\langle x,\mu\rangle}\right)^\top\right]$$

We begin with several preliminary observations in order to define the quantities c_1, c_2, and c_3. Namely, for each latent position (row) x of $\mathbf{X}$,

$$\langle x, \mu \rangle = \left(\frac{a+(K-1)b}{K} \right) =: c_1 \tag{28}$$

$$\mathbb{E}[\, g(x, X_1)] = \left(\frac{a(1-a)+(K-1)b(1-b)}{K} \right) =: c_2 \tag{29}$$

$$\mathbb{E}[\, g(x, X_1) X_1] := \left(\frac{a(1-a)-b(1-b)}{K} \right) x + b(1-b)\mu =: c_3 x + b(1-b)\mu \tag{30}$$

Subsequent computations yield

$$\begin{aligned}
\mathbf{\Delta} x &= \left(\tfrac{a-b}{K} \right) x + b\mu \\
\left[\mathbf{\Delta} - \left(\tfrac{a-b}{K} \right) \mathbf{I} \right] x x^\top &= b \mu x^\top \\
\langle \mathbf{\Delta} x, x \rangle &= \left(\tfrac{a^2+(K-1)b^2}{K} \right) \\
\tilde{\mathbf{\Delta}} \equiv \mathbb{E} \left[\tfrac{1}{\langle X_1, \mu \rangle} X_1 X_1^\top \right] &= \tfrac{1}{c_1} \mathbf{\Delta}
\end{aligned}$$

The above observations allow us to write $\tilde{\mathbf{\Sigma}}(x)$ as

$$\begin{aligned}
&\mathbb{E} \left[\left(\tfrac{g(x,X_1)}{\langle x, \mu \rangle} \right) \left(\tfrac{\tilde{\mathbf{\Delta}}^{-1} X_1}{\langle X_1, \mu \rangle} - \tfrac{x}{2\langle x, \mu \rangle} \right) \left(\tfrac{\tilde{\mathbf{\Delta}}^{-1} X_1}{\langle X_1, \mu \rangle} - \tfrac{x}{2\langle x, \mu \rangle} \right)^\top \right] \\
&= \tilde{\mathbf{\Delta}}^{-1} \mathbb{E} \left[\tfrac{g(x,X_1)}{\langle x, \mu \rangle} \left(\tfrac{X_1}{\langle X_1, \mu \rangle} - \tfrac{\tilde{\mathbf{\Delta}} x}{2\langle x, \mu \rangle} \right) \left(\tfrac{X_1}{\langle X_1, \mu \rangle} - \tfrac{\tilde{\mathbf{\Delta}} x}{2\langle x, \mu \rangle} \right)^\top \right] \tilde{\mathbf{\Delta}}^{-1} \\
&= \tfrac{1}{c_1} \mathbf{\Delta}^{-1} \mathbb{E} \left[g(x, X_1) \left(X_1 - \tfrac{1}{2c_1} \mathbf{\Delta} x \right) \left(X_1 - \tfrac{1}{2c_1} \mathbf{\Delta} x \right)^\top \right] \mathbf{\Delta}^{-1}
\end{aligned}$$

Expanding the term inside the expectation and applying linearity of expectation allows us to analyze each piece in turn. The first term in the expansion can be analyzed via the previous computations for ASE. For the second term,

$$\begin{aligned}
\mathbb{E} \left[\tfrac{1}{2c_1} g(x, X_1) X_1 x^\top \mathbf{\Delta} \right] &= \tfrac{1}{2c_1} \mathbb{E}[\, g(x, X_1) X_1] x^\top \mathbf{\Delta} \\
&= \tfrac{1}{2c_1} \left(c_3 x x^\top + b(1-b) \mu x^\top \right) \mathbf{\Delta} \\
&= \tfrac{1}{2c_1} \left(c_3 x x^\top + (1-b) \left[\mathbf{\Delta} - (\tfrac{a-b}{K}) \mathbf{I} \right] x x^\top \right) \mathbf{\Delta} \\
&= \left(\tfrac{1-b}{2c_1} \right) \mathbf{\Delta} x x^\top \mathbf{\Delta} + \left(\tfrac{K c_3 - (a-b)(1-b)}{2c_1 K} \right) x x^\top \mathbf{\Delta} \\
&= \left(\tfrac{1-b}{2c_1} \right) \mathbf{\Delta} x x^\top \mathbf{\Delta} + \left(\tfrac{a(b-a)}{2c_1 K} \right) x x^\top \mathbf{\Delta}
\end{aligned}$$

Note that the transpose of this matrix corresponds to the third term in the implicit expansion of interest (not shown). Finally, the fourth term simply reduces to the form

$$\mathbb{E} \left[g(x, X_1) \left(\tfrac{1}{2c_1} \mathbf{\Delta} x \right) \left(\tfrac{1}{2c_1} \mathbf{\Delta} x \right)^\top \right] = c_2 \left(\tfrac{1}{2c_1} \mathbf{\Delta} x \right) \left(\tfrac{1}{2c_1} \mathbf{\Delta} x \right)^\top = \left(\tfrac{c_2}{4c_1^2} \right) \mathbf{\Delta} x x^\top \mathbf{\Delta}$$

Thus,

$$\begin{aligned}
&\mathbb{E}\left[g(x, X_1) \left(X_1 - \tfrac{1}{2c_1}\mathbf{\Delta} x \right) \left(X_1 - \tfrac{1}{2c_1}\mathbf{\Delta} x \right)^\top \right] \\
&= \mathbb{E}[\, g(x, X_1) X_1 X_1^\top] - \mathbb{E}[\tfrac{1}{2c_1} g(x, X_1) X_1 x^\top \mathbf{\Delta}] - \mathbb{E}[\tfrac{1}{2c_1} g(x, X_1) X_1 x^\top \mathbf{\Delta}]^\top \\
&\quad + \mathbb{E}[\, g(x, X_1)(\tfrac{1}{2c_1}\mathbf{\Delta} x)(\tfrac{1}{2c_1}\mathbf{\Delta} x)^\top] \\
&= \mathbb{E}[\, g(x, X_1) X_1 X_1^\top] - \left(\tfrac{a(b-a)}{2c_1 K} \right) x x^\top \mathbf{\Delta} - \left(\tfrac{a(b-a)}{2c_1 K} \right) \mathbf{\Delta} x x^\top + \left(\tfrac{c_2}{4c_1^2} - \tfrac{1-b}{c_1} \right) \mathbf{\Delta} x x^\top \mathbf{\Delta}
\end{aligned}$$

Let $\mathbf{M}_1 \equiv \mathbf{M}_1(t) := \mathbf{N}\mathbf{D}_t\mathbf{N}^\top$ and $\mathbf{M}_2 := \mathbf{\Delta}$ with respect to the notation introduced earlier in the derivation for ASE. By completing the appropriate matrix product, there are explicit constants d_i depending on a, b, and K, such that

$$\begin{aligned}
\tilde{\mathbf{\Sigma}}_{1,2}(t) &= t\tilde{\mathbf{\Sigma}}(\nu_1) + (1-t)\tilde{\mathbf{\Sigma}}(\nu_2) \\
&= \mathbf{\Delta}^{-1} \left(d_1 \mathbf{\Delta} + d_2 \mathbf{N}\mathbf{D}_t\mathbf{N}^\top + d_3 \mathbf{N}\mathbf{D}_t\mathbf{N}^\top \mathbf{\Delta} + d_3 \mathbf{\Delta}\mathbf{N}\mathbf{D}_t\mathbf{N}^\top + d_4 \mathbf{\Delta}\mathbf{N}\mathbf{D}_t\mathbf{N}^\top \mathbf{\Delta} \right) \mathbf{\Delta}^{-1} \\
&= \mathbf{\Delta}^{-1} \left(\left(d_1 \mathbf{M}_2 + d_5 \mathbf{M}_1 \right) + (\mathbf{I} + d_6 \mathbf{M}_2)(d_7 \mathbf{M}_1)(\mathbf{I} + d_6 \mathbf{M}_2) \right) \mathbf{\Delta}^{-1} \\
&= :\mathbf{\Delta}^{-1} \left(\mathbf{M}_3 + \mathbf{M}_4 \right) \mathbf{\Delta}^{-1}
\end{aligned}$$

where $\mathbf{M}_3 \equiv \mathbf{M}_3(t) := d_1\mathbf{M}_2 + d_5\mathbf{M}_1(t)$ and $\mathbf{M}_4 \equiv \mathbf{M}_4(t) = (\mathbf{I} + d_6\mathbf{M}_2)(d_7\mathbf{M}_1(t))(\mathbf{I} + d_6\mathbf{M}_2)$.

Note that $\tilde{\nu}_k := \left(\frac{1}{\langle \nu_k, \mu \rangle} \right)^{1/2} \times \nu_k = \left(\frac{K}{a+(K-1)b} \right)^{1/2} \times \nu_k$ for $k = 1, 2$, so

$$\|\tilde{\nu}\|^2_{\tilde{\mathbf{\Sigma}}^{-1}_{1,2}(t)} = \tilde{\nu}^\top \tilde{\mathbf{\Sigma}}^{-1}_{1,2}(t) \tilde{\nu} = \left(\tfrac{K}{a+(K-1)b} \right) \nu^\top \mathbf{\Delta} \left(\mathbf{M}_3 + \mathbf{M}_4 \right)^{-1} \mathbf{\Delta} \nu$$

The above matrix inversion can again be carried out via the Sherman–Morrison–Woodbury formula. We omit the algebraic details. Subsequent computations and simplification yield

$$\sup_{t \in (0,1)} \left[t(1-t) \|\tilde{\nu}\|^2_{\tilde{\mathbf{\Sigma}}^{-1}_{1,2}(t)} \right] = \frac{4(a-b)^2(a+(K-1)b)^2}{4(a(1-a)+b(1-b))(a+(K-1)b)^2 K + (a-b)^2 K(3a(a-1)+3b(b-1)(K-1)+4abK)} \tag{31}$$

By underlying symmetry, the supremum is achieved at $t^\star = \frac{1}{2}$ over the entire parameter region $0 < b < a < 1$. Taken together, Equations (27) and (31) simplify to yield $\rho^\star$ as in Equation (18), thereby completing the proof. □

Conflict of interest. The authors declare no conflicts of interest.

Notes

1 The distinct row assumption removes potential redundancy with respect to block connectivity and labeling. Namely, if rows k and k' of $\mathbf{B}'$ are identical, then their corresponding blocks are indistinguishable and can without loss of generality be merged to form a reduced block edge probability matrix $\mathbf{B}$ with corresponding combined block membership probability $\pi_k + \pi_{k'}$. We also remark that Definition 1 implicitly permits vertex self-loops, a choice that we make for mathematical expediency. Whether or not self-loops are disallowed does not alter the asymptotic results and conclusions presented in this paper.

References

Abbe, E. (2018). Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, *18*(177), 1–86.

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

Athreya, A., Fishkind, D. E., Tang, M., Priebe, C. E., Park, Y., Vogelstein, J. T., … Sussman, D. L. (2018). Statistical inference on random dot product graphs: A survey. *Journal of Machine Learning Research*, *18*(226), 1–92.

Athreya, A., Priebe, C. E., Tang, M., Lyzinski, V., Marchette, D. J., & Sussman, D. L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, *78*(1), 1–18.

Bollobás, B., Janson, S., & Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, *31*(1), 3–122.

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, *21*(4), 375–395.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, *23*(4), 493–507.

Chernoff, H. (1956). Large-sample theory: Parametric case. *The Annals of Mathematical Statistics*, *27*(1), 1–22.

Csermely, P., London, A., Wu, L.-Y., & Uzzi, B. (2013). Structure and dynamics of core-periphery networks. *Journal of Complex Networks*, *1*(2), 93–123.

Devroye, L., Györfi, L., & Lugosi, G. (2013). *A Probabilistic Theory of Pattern Recognition*. Vol. 31. Springer.

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae (Debrecen)*, *6*, 290–297.

Fishkind, D. E., Sussman, D. L., Tang, M., Vogelstein, J. T., & Priebe, C. E. (2013). Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. *SIAM Journal on Matrix Analysis and Applications*, *34*(1), 23–39.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*(2), 109–137.

Holme, P. (2005). Core-periphery organization of complex networks. *Physical Review E*, *72*, 046111.

Horn, R. A., & Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press.

Jeub, L. G. S., Balachandran, P., Porter, M. A., Mucha, P. J., & Mahoney, M. W. (2015). Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Physical Review E*, *91*(1), 012821.

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

Lei, J., & Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, *43*(1), 215–237.

Leskovec, J., Lang, K. J., Dasgupta, A., & Mahoney, M. W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, *6*(1), 29–123.

Liese, Friedrich, & Vajda, Igor. (2006). On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, *52*(10), 4394–4412.

Lyzinski, V., Tang, M., Athreya, A., Park, Y., & Priebe, C. E. (2017). Community detection and classification in hierarchical stochastic blockmodels. *IEEE Transactions on Network Science and Engineering*, *4*(1), 13–26.

McSherry, F. (2001). Spectral partitioning of random graphs. *In Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pp. 529–537.

Nickel, C. L. M. (2006). Random dot product graphs: A model for social networks. *Ph.D. thesis, Johns Hopkins University*.

Priebe, C. E., Park, Y., Vogelstein, J. T., Conroy, J. M., Lyzinski, V., Tang, M., … Bridgeford, E. (2019). On a two-truths phenomenon in spectral graph clustering. *Proceedings of the National Academy of Sciences*, *116*(13), 5995–6000.

Rohe, K., Chatterjee, S., & Yu, B. (2011). Spectral clustering and the high-dimensional stochastic block model. *Annals of Statistics*, *39*(4), 1878–1915.

Rubin-Delanchy, P., Priebe, C. E., Tang, M., & Cape, J. (2017). A statistical interpretation of spectral embedding: The generalised random dot product graph. *arxiv preprint arxiv:1709.05506*.

Sarkar, P., & Bickel, P. J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Annals of Statistics*, *43*(3), 962–990.

Sussman, D. L., Tang, M., Fishkind, D. E., & Priebe, C. E. (2014). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, *107*(499), 1119–1128.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., & Priebe, C. E. (2017a). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, *23*(3), 1599–1630.

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., Park, Y., & Priebe, C. E. (2017b). A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, *26*(2), 344–354.

Tang, M., & Priebe, C. E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Annals of Statistics*, *46*(5), 2360–2415.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416.

Young, S., & Scheinerman, E. (2007). Random dot product graph models for social networks. *Algorithms and Models for the Web-Graph*, *4863*, 138–149.

Zhang, X., Martin, T., & Newman, M. E. J. (2015). Identification of core-periphery structure in networks. *Physical Review E*, *91*(3), 032803.

Cite this article: Cape J., Tang M., and Priebe C. E. (2019). On spectral embedding performance and elucidating network structure in stochastic blockmodel graphs. *Network Science* **7**, 269–291. https://doi.org/10.1017/nws.2019.23