Sankhyā A : The Indian Journal of Statistics
https://doi.org/10.1007/s13171-021-00268-x

Eigenvalues of Stochastic Blockmodel Graphs and Random Graphs with Low-Rank Edge Probability Matrices

Avanti Athreya
Johns Hopkins University Baltimore USA
Joshua Cape
University of Pittsburgh Pittsburgh USA
Minh Tang
North Carolina State University Raleigh USA

Abstract

We derive the limiting distribution for the outlier eigenvalues of the adjacency matrix for random graphs with independent edges whose edge probability matrices have low-rank structure. We show that when the number of vertices tends to infinity, the leading eigenvalues in magnitude are jointly multivariate Gaussian with bounded covariances. As a special case, this implies a limiting normal distribution for the outlier eigenvalues of stochastic blockmodel graphs and their degree-corrected or mixed-membership variants. Our result extends the classical result of Füredi and Komlós on the fluctuation of the largest eigenvalue for Erdős–Rényi graphs.

AMS (2000) *subject classification.* Primary 62H12; Secondary 05C50, 62F12.
Keywords and phrases. Random graphs, Stochastic blockmodels, Asymptotic normality, Eigenvalues distribution.

1 Introduction

The systematic study of eigenvalues of random matrices dates back to the seminal work of Wigner (1955) on the semicircle law for Wigner ensembles of symmetric or Hermitian matrices. A random $n \times n$ symmetric matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{n}$ is said to be a Wigner matrix if, for $i \leq j$, the entries a_{ij} are independent mean zero random variables with variance $\sigma_{ij}^2 = 1$ for $i < j$ and $\sigma_{ii}^2 = \sigma^2 > 0$. Much is known about the spectral properties of these matrices, such as universality of the semi-circle law for bulk eigenvalues (Erdös et al., 2010; Tao and Vu, 2010), universality of the Tracy–Widom distribution for

the largest non-spiked eigenvalue (Soshnikov, 1999), universality properties of eigenvectors (Tao and Vu, 2012; Knowles and Yin, 2013), and eigenvector delocalization (Erdös et al., 2009).

In contrast, somewhat less is known to date about the spectral properties of general random symmetric matrices $\mathbf{A} = (a_{ij})_{i,j=1}^n$ where the entries a_{ij} are independent, not necessarily mean zero random variables with possibly heterogeneous variances. Such random matrices arise naturally in many settings, with perhaps the most popular example being the adjacency matrices of (inhomogeneous) independent edge random graphs. In the case when $\mathbf{A}$ is the adjacency matrix for an Erdős–Rényi graph where the edges are i.i.d Bernoulli random variables, Arnold (1967) and Ding and Jiang (2010) show that the empirical distribution of the eigenvalues of $\mathbf{A}$ also converges to a semi-circle law. Meanwhile, the following result of Füredi and Komlós (1981) shows that the largest eigenvalue of $\mathbf{A}$ is asymptotically normally distributed when $\mathbb{E}[a_{ij}] = \mu$ and $\mathrm{Var}[a_{ij}] = \sigma^2$ for $i < j$.

THEOREM 1. Füredi and Komlós (1981) *Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ symmetric matrix where the entries a_{ij} are independent (not necessarily identically distributed) random variables uniformly bounded in magnitude by a constant C. Assume that for $i > j$, the a_{ij} have a common expectation $\mu > 0$ and variance σ^2. Furthermore, assume that $\mathbb{E}[a_{ii}] = v$ for all i. Then, the distribution of $\lambda_1(\mathbf{A})$, the largest eigenvalue of $\mathbf{A}$, can be approximated on the order $n^{-1/2}$ by a normal distribution with mean $(n-1)\mu + v + \sigma^2/\mu$ and variance $2\sigma^2$, i.e.,*

$$\lambda_1(\mathbf{A}) - (n-1)\mu - v \xrightarrow{\mathrm{d}} \mathcal{N}\left(\tfrac{\sigma^2}{\mu}, 2\sigma^2\right) \tag{1.1}$$

as $n \to \infty$. Furthermore, with probability tending to one,

$$\max_{i \geq 2} |\lambda_i(\mathbf{A})| < 2\sigma\sqrt{n} + O(n^{1/3}\log n). \tag{1.2}$$

Theorem 1 yields that for an Erdős–Rényi graph with edge probability p,

$$\lambda_1(\mathbf{A}) - np \xrightarrow{\mathrm{d}} \mathcal{N}(1-p, 2p(1-p))$$

as $n \to \infty$.

A natural generalization of (homogeneous) Erdős–Rényi random graphs is the notion of stochastic blockmodel graphs (Holland et al., 1983) where, given an integer $K \geq 1$, the entries a_{ij} for $i \leq j$ are independent Bernoulli random variables with $\mathbb{E}[a_{ij}] \in \mathcal{S}$ for some set $\mathcal{S}$ of cardinality $K(K+1)/2$. More specifically, we have the following definition.

Definition 1 (Stochastic blockmodel graphs)**.** *Let $K \geq 1$ be a positive integer and let $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_K)$ be a non-negative vector in $\mathbb{R}^K$ with $\sum_k \pi_k = 1$. Let $\mathbf{B}$ be a $K \times K$ symmetric matrix with entries $b_{k\ell} \in [0,1]$ for all $1 \leq k \leq \ell \leq K$. We say that $(\mathbf{A}, \boldsymbol{\tau}) \sim \mathrm{SBM}(\boldsymbol{\pi}, \mathbf{B})$ if the following holds. First, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$ and the τ_i are i.i.d random variables with $\Pr[\tau_i = k] = \pi_k$. Then, let $\mathbf{A} \in \{0,1\}^{n \times n}$ be an $n \times n$ symmetric binary matrix such that, conditioned on $\boldsymbol{\tau}$, for all $i \geq j$ the entries a_{ij} in $\mathbf{A}$ are independent Bernoulli random variables with $\mathbb{E}[a_{ij}] = b_{\tau_i, \tau_j}$.*

The stochastic blockmodel is among the most popular generative models for random graphs with community structure; the nodes of such graphs are partitioned into blocks (communities), and the probability of connection between any two nodes is a function of their block assignments. The adjacency matrix $\mathbf{A}$ of a stochastic blockmodel graph can be viewed as $\mathbf{A} = \mathbb{E}[\mathbf{A}] + (\mathbf{A} - \mathbb{E}[\mathbf{A}])$, where $\mathbb{E}[\mathbf{A}]$ is a low-rank deterministic matrix and $(\mathbf{A} - \mathbb{E}[\mathbf{A}])$ is a generalized Wigner matrix whose elements are independent mean zero random variables with heterogeneous variances (and noting that the variance of a_{ij} is a function of its mean, $\mathbb{E}[a_{ij}]$). We emphasize that our assumptions on $\mathbf{A} - \mathbb{E}[\mathbf{A}]$ distinguish us from existing results in the literature. For example, Péché (2006), Bordenave and Capitaine (2016), Pizzo et al. (2013), Knowles and Yin (2014), Capitaine et al. (2009), & Capitaine et al. (2012) consider finite rank additive perturbations of the random matrix $\mathbf{X}$ given by $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{P}$ under the assumption that $\mathbf{X}$ is either a Wigner matrix or is sampled from the Gaussian unitary ensemble; for these models, Capitaine et al. (2009), Capitaine et al. (2012), & Knowles and Yin (2014) derive several limiting distribution for the largest eigenvalues of $\tilde{\mathbf{X}}$, showing that these eigenvalues are asymptotically normal provided the eigenvectors of $\mathbf{P}$ are delocalized. Meanwhile, in Benaych-Georges and Nadakuditi (2011), the authors assume that $\mathbf{X}$ or $\mathbf{P}$ is orthogonally invariant; a symmetric random matrix $\mathbf{H}$ is said to be orthogonally invariant if its distribution is invariant under similarity transformations $\mathbf{H} \mapsto \mathbf{W}^{-1}\mathbf{H}\mathbf{W}$ whenever $\mathbf{W}$ is an orthogonal matrix. Finally, in O'Rourke and Renfrew (2014), the entries of $\mathbf{X}$ are assumed to be from an elliptical family of distributions, i.e., the collection $\{(X_{ij}, X_{ji})\}$ for $i < j$ consists of i.i.d copies of a random variable tuple (ξ_1, ξ_2) with $\mathbb{E}[\xi_1 \xi_2] = \rho$.

Recent years have witnessed progress towards characterizing the empirical eigenvalue distribution for stochastic blockmodels. In particular, Zhang et al. (2014) and Avrachenkov et al. (2015) derived the Stieltjes transform for the limiting empirical distribution of the bulk eigenvalues for stochastic blockmodel graphs, thereby showing that the empirical distribution of the

eigenvalues need not converge to the classical semicircle law. Zhang et al. (2014) and Avrachenkov et al. (2015) also considered the edge eigenvalues, but their characterizations rely upon inverting the Stieltjes transform and thus currently do not yield limiting distributions for these largest eigenvalues. Lei (2016) derived the limiting distribution for the largest eigenvalue of a centered and scaled version of $\mathbf{A}$. More specifically, Lei (2016) showed that there is a consistent estimate $\widehat{\mathbb{E}[\mathbf{A}]} = (\hat{a}_{ij})$ of $\mathbb{E}[\mathbf{A}]$ such that the matrix $\tilde{\mathbf{A}} = (\tilde{a}_{ij})$ with entries $\tilde{a}_{ij} = (a_{ij} - \hat{a}_{ij})/\sqrt{(n-1)\hat{a}_{ij}(1-\hat{a}_{ij})}$ has a limiting Tracy–Widom distribution, i.e., $n^{2/3}(\lambda_1(\tilde{\mathbf{A}}) - 2)$ converges to Tracy–Widom.

This paper addresses the open question of determining the limiting distribution of the outlier eigenvalues for adjacency matrices $\mathbf{A}$ whose edge probability matrices $\mathbb{E}[\mathbf{A}]$ are low-rank. In particular, we extend the result of Füredi and Komlós and show that, in the limit, these eigenvalues are jointly multivariate normal with bounded covariances. This result implies, as a special case, a limiting normal distribution for the outlier eigenvalues of stochastic blockmodels and their degree-corrected or mixed-membership variants. Other extensions of Füredi and Komlós' result to finite-rank perturbations of Wigner matrices appeared previously in capitaine et al. (2009, 2012) and Knowles and Yin (2014) but do not resolve the present problem. Finally, Chakrabarty et al. (2020) derived a related central limit theorem for the k largest eigenvalues of a finite rank inhomogeneous Erdős–R'enyi graphs; this class of random graphs is equivalent to the generalized random dot product graphs considered in this paper. While Theorem 2.3 in Chakrabarty et al. (2020) and Theorem 2 look similar at first glance, there is an important difference between the two results. The central limit theorem in Chakrabarty et al. (2020) is centered around $\mathbb{E}[\lambda_i(\mathbf{A})]$ while the central limit theorem in this paper is centered around $\lambda_i(\mathbb{E}[\mathbf{A}])$. As $\mathbf{A}$ is a single random noisy sample from the true matrix of edge probabilities, and since $\mathbb{E}[\lambda_i(\mathbf{A})] \neq \lambda_i(\mathbb{E}[\mathbf{A}])$ in general, our result is arguably more conceptually and practically useful. For example, given any value of n, the quantity $\lambda_i(\mathbb{E}[\mathbf{A}])$ can be calculated directly from the model parameters but this is not the case for $\mathbb{E}[\lambda_i(\mathbf{A})]$. Nevertheless, even though there is no explicit expression for $\mathbb{E}[\lambda_i(\mathbf{A})]$, our results establish that the difference $\mathbb{E}[\lambda_i(\mathbf{A})] - \lambda_i(\mathbb{E}[\mathbf{A}])$ converges to a constant as $n \to \infty$.

2 Main Results

We present our main results in the broader framework of generalized random dot product graphs where $\mathbb{E}[\mathbf{A}]$ is only assumed to be low rank, i.e., in contrast to blockmodels, we do not require that the entries of $\mathbb{E}[\mathbf{A}]$

take on a finite number of distinct values. We first define the notion of a (generalized) random dot product graph (Young and Scheinerman, 2007; Rubin-Delanchy et al., 2017).

Definition 2 (Generalized random dot product graph)**.** *Let $d \geq 1$ and $p \geq 1$ and $q \geq 0$ be integers such that $p + q = d$. Let $\mathbf{I}_{p,q}$ denote the diagonal matrix whose first p diagonal elements equal 1 and the remaining q diagonal entries equal -1. Let $\mathcal{X}$ be a subset of $\mathbb{R}^d$ such that $\boldsymbol{x}^\top \mathbf{I}_{p,q} \boldsymbol{y} \in [0,1]$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$. Let F be a distribution taking values in $\mathcal{X}$. We say $(\mathbf{A}, \mathbf{X}) \sim \mathrm{GRDPG}(F)$ with signature (p,q) if the following holds. First, let $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$ and set $\mathbf{X} = [X_1 \mid \cdots \mid X_n]^\top \in \mathbb{R}^{n \times d}$. Then, let $\mathbf{A} \in \{0,1\}^{n \times n}$ be a symmetric matrix such that, conditional on $\mathbf{X}$, the entries $\{a_{ij}\}_{i \leq j}$ are independent and $a_{ij} \sim \mathrm{Bernoulli}(X_i^\top \mathbf{I}_{p,q} X_j)$. We therefore have*

$$\Pr[\mathbf{A} \mid \mathbf{X}] = \prod_{i \leq j} (X_i^\top \mathbf{I}_{p,q} X_j)^{a_{ij}} (1 - X_i^\top \mathbf{I}_{p,q} X_j)^{(1-a_{ij})}. \tag{2.1}$$

In the special case when $q = 0$, we say that $(\mathbf{A}, \mathbf{X}) \sim \mathrm{RDPG}(F)$, namely that $\mathbf{A}$ is a random dot product graph.

REMARK 1 (Low-rank structure beyond blockmodel random graphs). Any given stochastic blockmodel graph $(\mathbf{A}, \boldsymbol{\tau}) \sim \mathrm{SBM}(\boldsymbol{\pi}, \mathbf{B})$ can be represented as a generalized random dot product graph $(\mathbf{A}, \mathbf{X}) \sim \mathrm{GRDPG}(F)$ where F is a mixture of point masses. Indeed, suppose $\mathbf{B}$ is a $K \times K$ matrix and let $\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^\top$ denote the eigendecomposition of $\mathbf{B}$. Then, letting $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_K$ denote the rows of $\mathbf{U}|\boldsymbol{\Sigma}|^{1/2}$, we can define $F = \sum_{k=1}^K \pi_k \delta_{\boldsymbol{\nu}_k}$ where δ is the Dirac delta function. The signature (p,q) is given by the number of positive and negative eigenvalues of $\mathbf{B}$, respectively. This construction permits settings where $d < K$, recalling that $d = p + q$. Similar constructions show that formulations of degree-corrected stochastic blockmodel graphs (Karrer and Newman, 2011), mixed-membership stochastic blockmodel graphs (Airoldi et al., 2008), and popularity adjusted blockmodels (Noroozi et al., 2019; Sengupta and Chen, 2018) are also special cases of generalized random dot product graphs.

REMARK 2 (On identifiability in low-rank random graph models). Non-identifiability is an intrinsic property of generalized random dot product graphs. In particular, if $(\mathbf{A}, \mathbf{X}) \sim \mathrm{GRDPG}(F)$ with signature (p,q), then for any matrix $\mathbf{W}$ such that $\mathbf{W}\mathbf{I}_{p,q}\mathbf{W}^\top = \mathbf{I}_{p,q}$, we have that $(\mathbf{B}, \mathbf{Y}) \sim \mathrm{GRDPG}(F \circ \mathbf{W})$ is identically distributed to $(\mathbf{A}, \mathbf{X})$, where $F \circ \mathbf{W}$ denotes the distribution of $\mathbf{W}\xi$ for $\xi \sim F$. Any matrix $\mathbf{W}$ satisfying $\mathbf{W}\mathbf{I}_{p,q}\mathbf{W}^\top = \mathbf{I}_{p,q}$ is said to be an *indefinite orthogonal* matrix. When $q = 0$ the condition on $\mathbf{W}$ reduces to that of an orthogonal matrix.

With the above notation in place, we now state our generalization of Füredi and Komlós (1981) for the generalized random dot product graph setting.

THEOREM 2. *Let* $(\mathbf{A}, \mathbf{X}) \sim \mathrm{GRDPG}(F)$ *be a* d*-dimensional generalized random dot product graph with signature* (p, q)*. Let* $\Delta = \mathbb{E}[XX^\top]$ *where* $X \sim F$ *and suppose that* $\Delta\mathbf{I}_{p,q}$ *has* $p + q = d$ *simple eigenvalues. Let* $\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top$ *and for* $1 \le i \le d$*, let* $\hat{\lambda}_i$ *and* λ_i *be the* i*-th largest eigenvalues of* $\mathbf{A}$ *and* $\mathbf{P}$ *(in modulus), respectively. Let* $\lambda_i(\Delta\mathbf{I}_{p,q})$ *and* $\boldsymbol{\xi}_i$ *denote the* i*-th largest eigenvalue and associated (unit-norm) eigenvector pair for the matrix* $\Delta^{1/2}\mathbf{I}_{p,q}\Delta^{1/2}$*. Let* $\mu = \mathbb{E}[X]$ *and denote by* $\boldsymbol{\eta}$ *the* $d \times 1$ *vector whose elements are*

$$\eta_i = \frac{1}{\lambda_i(\Delta\mathbf{I}_{p,q})}\mathbb{E}\left[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_i(X^\top\mathbf{I}_{p,q}\mu - X^\top\mathbf{I}_{p,q}\Delta\mathbf{I}_{p,q}X)\right]. \tag{2.2}$$

Also, let $\boldsymbol{\Gamma}$ *be the* $d \times d$ *matrix whose elements are*

$$\begin{aligned}\Gamma_{ij} &= 2\big(\mathbb{E}[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_j X]^\top\mathbf{I}_{p,q}\mathbb{E}[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_j X]\big)\\ &- 2\mathrm{tr}\Big(\big[\mathbb{E}[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_j XX^\top]\mathbf{I}_{p,q}\big]^2\Big).\end{aligned} \tag{2.3}$$

We then have

$$(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2, \ldots, \hat{\lambda}_d - \lambda_d) \xrightarrow{\mathrm{d}} \mathrm{MVN}(\boldsymbol{\eta}, \boldsymbol{\Gamma})$$

as $n \to \infty$*.*

When $\mathbf{A}$ is a d-dimensional random dot product graph, Theorem 2 simplifies to the following result.

Corollary 1. *Let* $(\mathbf{A}, \mathbf{X}) \sim \mathrm{RDPG}(F)$ *be a* d*-dimensional random dot product graph and suppose that* $\Delta = \mathbb{E}[XX^\top]$ *has* d *simple eigenvalues. Let* $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$ *and let* $\lambda_i(\Delta)$ *and* $\boldsymbol{\xi}_i$ *denote the* i*-th largest eigenvalue and associated (unit-norm) eigenvector of* Δ*. Let* $\mu = \mathbb{E}[X]$ *and denote by* $\boldsymbol{\eta}$ *the* $d \times 1$ *vector with elements*

$$\eta_i = \frac{1}{\lambda_i(\Delta)^2}\mathbb{E}\left[\boldsymbol{\xi}_i^\top XX^\top\boldsymbol{\xi}_i(X^\top\mu - X^\top\Delta X)\right]. \tag{2.4}$$

Also, let $\boldsymbol{\Gamma}$ *be the* $d \times d$ *matrix whose elements are*

$$\begin{aligned}\Gamma_{ij} &= \frac{2}{\lambda_i(\Delta)\lambda_j(\Delta)}\big(\mathbb{E}[\boldsymbol{\xi}_i^\top XX^\top\boldsymbol{\xi}_j X]^\top\mathbb{E}[\boldsymbol{\xi}_i^\top XX^\top\boldsymbol{\xi}_j X]\big)\\ &- \frac{2}{\lambda_i(\Delta)\lambda_j(\Delta)}\mathrm{tr}\big(\mathbb{E}[\boldsymbol{\xi}_i^\top XX^\top\boldsymbol{\xi}_j XX^\top]\mathbb{E}[\boldsymbol{\xi}_i^\top XX^\top\boldsymbol{\xi}_j XX^\top]\big).\end{aligned} \tag{2.5}$$

We then have

$$(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2, \ldots, \hat{\lambda}_d - \lambda_d) \xrightarrow{\mathrm{d}} \mathrm{MVN}(\boldsymbol{\eta}, \boldsymbol{\Gamma})$$

as $n \to \infty$.

To illustrate Corollary 1, let $\mathbf{A}$ be an Erdős–Rényi graph with edge probability p; then F is the Dirac delta measure at $p^{1/2}$ and hence $\Delta = p$, $\boldsymbol{\xi}_1 = 1$, and $\lambda_i(\Delta) = p$. We thus recover the earlier result of Füredi and Komlós, namely that

$$\hat{\lambda}_i - np \xrightarrow{\mathrm{d}} \mathcal{N}(1 - p, 2p(1 - p)).$$

REMARK 3 (On asymptotic bias). Theorem 2 and Corollary 1 show that $(\hat{\lambda}_i - \lambda_i)_{i=1}^{d}$ converges jointly to a multivariate normal with mean $\boldsymbol{\eta}$ and covariance matrix $\boldsymbol{\Gamma}$. We therefore have that $(\mathbb{E}[\lambda_i(\mathbf{A})] - \lambda_i(\mathbb{E}[\mathbf{A}])) \to \eta_i$ as $n \to \infty$. Note however that this is only an asymptotic limiting result and that for finite n we do not have, nor do we expect to have, an explicit expression for the difference $\mathbb{E}[\lambda_i(\mathbf{A})] - \lambda_i(\mathbb{E}[\mathbf{A}])$. Furthermore, from the expression for the bias terms η_i, we have $\eta_i > 0$ if $\lambda_i > 0$ and $\eta_i < 0$ if $\lambda_i < 0$. In other words we expect $|\hat{\lambda}_i| > |\lambda_i|$. This phenomenon where $|\mathbb{E}[\hat{\lambda}_i]| > |\lambda_i|$ appears throughout random matrix theory. In particular, it is well known that the spiked eigenvalues in sample covariance matrices are often larger than the eigenvalues of underlying population-level covariance matrices (Donoho et al., 2018, Eq. (1.3)) and that the eigenvalues of finite-rank signal matrices perturbed by Wigner noise are larger than the eigenvalues of the signal matrices themselves (Benaych-Georges and Nadakuditi, 2011, Section 3.1).

Finally, while results based on centering around $\mathbb{E}[\lambda_i(\mathbf{A})]$ and the results based on centering around $\lambda_i \equiv \lambda_i(\mathbb{E}[\mathbf{A}])$ are very similar, there is a subtle difference in their utility for subsequent inference. Notably, results based on centering around $\mathbb{E}[\lambda_i(\mathbf{A})]$ may not be appropriate for one-sample hypothesis testing problems unless one incorporates the bias terms η_i, which then becomes equivalent to the use of limit results centered around $\lambda_i(\mathbb{E}[\mathbf{A}])$ instead. Note, however, that for two-sample testing problems, limit results centered around $\mathbb{E}[\lambda_i(\mathbf{A})]$ and limit results centered around $\lambda_i(\mathbb{E}[\mathbf{A}])$ should yield the same test procedure. Hypothesis testing for random graphs using the largest eigenvalues as test statistics has been considered previously, for example in Ghoshdastidar et al. (2017) & Tokuda (2018).

As an example for the aforementioned claim, albeit a contrived example, suppose we are given an Erdős–Rényi graph $\mathbf{A}$ with unknown edge probability p and that we want to test the simple null hypothesis of $\mathbb{H}_0 \colon p = p_0$ for some value of p_0. Suppose we use the test statistic $\hat{\lambda}_1 - \lambda_1$. Then, for $p = p_0$,

we have $\lambda_1 = np_0$ and furthermore the bias term η and variance term σ^2 for $\hat{\lambda}_1$ are $1-p_0$ and $2p_0(1-p_0)$, respectively. Hence determining an asymptotically level α rejection region for $\hat{\lambda}_1 - \lambda_1$ is straightforward. In contrast, the test statistic $\hat{\lambda}_1 - \mathbb{E}[\hat{\lambda}_1]$ is rather problematic as $\mathbb{E}[\hat{\lambda}_1]$ cannot be computed directly from the assumed value $p = p_0$ and the number of vertices in the observed graph $\mathbf{A}$ (at least not without resorting to either simulations or to using $\lambda_i(\mathbb{E}[\mathbf{A}]) + \eta$). The same phenomenon also holds for testing more general one-sided hypothesis, e.g., $\mathbb{H}_0 \colon p < p_0$ against $\mathbb{H}_A \colon p \geq p_0$, or two-sided hypothesis for more general models such as (1) testing $\mathbb{H}_0 \colon \mathbf{B} = \mathbf{B}_0$ against $\mathbb{H}_A \colon \mathbf{B} \neq \mathbf{B}_0$ for stochastic blockmodel graphs and (2) testing $\mathbb{H}_0 \colon \mathbf{P} = \mathbf{P}_0$ for generalized random dot product graphs.

Remark 4 (Disallowing versus permitting self-loops and subsequent asymptotic bias). The definition of generalized random dot product graphs as given in Definition 2 allows for self-loops, i.e., $a_{ii} \sim \text{Bernoulli}(X_i^\top \mathbf{I}_{p,q} X_i)$. If we instead disallow self-loops so that $a_{ii} = 0$ for all i, then the limit results in Theorem 2 and Corollary 1 no longer hold as stated and must be amended. More specifically, recall $\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top$ and note that if self-loops are disallowed then $\mathbb{E}[\mathbf{A}] = \mathbf{P} - \text{diag}(\mathbf{P})$ where $\text{diag}(\mathbf{P})$ is the diagonal matrix obtained by setting the off-diagonal entries of $\mathbf{P}$ to zero. Now, using the notation in Theorem 2, suppose that the eigenvalues of $\Delta\mathbf{I}_{p,q}$ are all distinct. For each i, define

$$\theta_i = \boldsymbol{\xi}_i^\top \Delta^{-1/2} \mathbb{E}\left[(X^\top \mathbf{I}_{p,q} X) X X^\top\right] \Delta^{-1/2} \boldsymbol{\xi}_i. \tag{2.6}$$

Letting $\hat{\lambda}_i$ and λ_i denote the eigenvalues of $\mathbf{A}$ and $\mathbf{P}$, respectively, we then have that

$$(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2, \ldots, \hat{\lambda}_d - \lambda_d) \xrightarrow{\text{d}} \text{MVN}(\boldsymbol{\eta} - \boldsymbol{\theta}, \boldsymbol{\Gamma}) \tag{2.7}$$

as $n \to \infty$, where $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \ldots, \theta_d)$. In words, the d largest eigenvalues $\{\hat{\lambda}_i\}_{i=1}^d$ of $\mathbf{A}$ are asymptotically jointly multivariate normal after centering, with asymptotic covariance matrix $\boldsymbol{\Gamma}$ as defined in Theorem 2. Hence, the difference between this result and Theorem 2 amounts to the role of asymptotic bias terms, i.e., the appearance of the quantities $\theta_1, \ldots, \theta_d$ when disallowing self-loops. The derivation of Eq. 2.7 is given in Remark 6 at the end of Section 4.

When the eigenvalues of $\Delta\mathbf{I}_{p,q}$ are not all simple eigenvalues, Theorem 2 can be adapted to yield the following result.

Theorem 3. *Let* $(\mathbf{X}, \mathbf{A}) \sim \text{GRDPG}(F)$ *be a* d*-dimensional generalized random dot product graph on* n *vertices with signature* (p, q) *and* $\mathbf{P} =$

$\mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top$. *For* $1 \le i \le d$, *let* $\hat{\lambda}_i$ *and* λ_i *denote the* i*-th largest eigenvalues of* $\mathbf{A}$ *and* $\mathbf{P}$ *(in modulus), respectively, and let* $\boldsymbol{v}_i$ *be the unit norm eigenvector satisfying* $(\mathbf{X}^\top\mathbf{X})^{1/2}\mathbf{I}_{p,q}(\mathbf{X}^\top\mathbf{X})^{1/2}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i$.

Define $\tilde{\boldsymbol{v}}_i = (\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i$. *Denote by* $\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}(\mathbf{X})$ *the* $d \times 1$ *vector with elements*

$$\tilde{\eta}_i = \frac{1}{\lambda_i}\sum_{s=1}^{n}\tilde{\boldsymbol{v}}_i^\top X_s X_s^\top \tilde{\boldsymbol{v}}_i \sum_{t=1}^{n} X_s^\top \mathbf{I}_{p,q} X_t (1 - X_s^\top \mathbf{I}_{p,q} X_t), \tag{2.8}$$

and denote by $\tilde{\boldsymbol{\Gamma}}$ *the* $d \times d$ *matrix whose elements are*

$$\tilde{\Gamma}_{ij} = 2\Big(\sum_{s=1}^{n}\sum_{t=1}^{n}\tilde{\boldsymbol{v}}_i^\top X_s X_s^\top \tilde{\boldsymbol{v}}_j \tilde{\boldsymbol{v}}_i^\top X_t X_t^\top \tilde{\boldsymbol{v}}_j X_s^\top \mathbf{I}_{p,q} X_t (1 - X_s^\top \mathbf{I}_{p,q} X_t)\Big) \tag{2.9}$$

We then have

$$\tilde{\boldsymbol{\Gamma}}^{-1/2}(\hat{\lambda}_1 - \lambda_1 - \tilde{\eta}_1, \hat{\lambda}_2 - \lambda_2 - \tilde{\eta}_2, \ldots, \hat{\lambda}_d - \lambda_d - \tilde{\eta}_d) \xrightarrow{\mathrm{d}} \mathrm{MVN}(\mathbf{0}, \mathbf{I}_d)$$

as $n \to \infty$, *where* $\mathbf{I}_d$ *denotes the* $d \times d$ *identity matrix.*

The main difference between Theorem 3 and Theorem 2 is that we do not claim the quantities $(\tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_d)$ and the matrix $\tilde{\boldsymbol{\Gamma}}$ in Theorem 3 (which, for $(\mathbf{A}, \mathbf{X}) \sim \mathrm{GRDPG}(F)$ are functions of the underlying latent positions $\mathbf{X}$) converge as $n \to \infty$. This difference stems mainly from the fact that when the eigenvalues of $\Delta\mathbf{I}_{p,q}$ are not simple eigenvalues, then $\frac{\mathbf{X}^\top\mathbf{X}}{n} \to \Delta$ almost surely as $n \to \infty$, but the eigenvectors $\{\boldsymbol{v}_i\}_{i=1}^d$ of $(\frac{\mathbf{X}^\top\mathbf{X}}{n})^{1/2}\mathbf{I}_{p,q}(\frac{\mathbf{X}^\top\mathbf{X}}{n})^{1/2}$ do not necessarily converge to those of $\Delta^{1/2}\mathbf{I}_{p,q}\Delta^{1/2}$. Finally, we emphasize that the potential multiplicity of eigenvalues refers to the $d \times d$ matrix $\Delta\mathbf{I}_{p,q}$ and not the matrix $\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top$. Indeed, the proof of Theorem 3 shows that the eigenvalues of $\mathbf{P}$ are, with high probability, still simple eigenvalues; however, the distances between these eigenvalues are potentially much smaller than those in the setting where all the eigenvalues of $\Delta\mathbf{I}_{p,q}$ are distinct. Finally, echoing our earlier discussion in Remark 4, an almost identical result to Theorem 3 also holds when we disallow self-loops in our observed graphs. Concretely, for each i, let

$$\tilde{\theta}_i = \sum_{s=1}^{n}\tilde{\boldsymbol{v}}_i^\top X_s X_s^\top \tilde{\boldsymbol{v}}_i X_s^\top \mathbf{I}_{p,q} X_s.$$

Then, for $\hat{\lambda}_i = \lambda_i(\mathbf{A})$ where $\mathbf{A}$ has no self-loops, as $n \to \infty$,

$$\tilde{\boldsymbol{\Gamma}}^{-1/2}(\hat{\lambda}_1 - \lambda_1 - \tilde{\eta}_i + \tilde{\theta}_1, \ldots, \hat{\lambda}_d - \lambda_d - \tilde{\eta}_d + \tilde{\theta}_d) \xrightarrow{\mathrm{d}} \mathrm{MVN}(\mathbf{0}, \mathbf{I}_d).$$

Remark 5 (On consistency of plug-in estimators). Given the limit results in Theorem 2 and Theorem 3, it is natural to inquire about the estimation of the quantities $\tilde{\eta}_i$ or η_i together with estimation of the covariance matrices $\tilde{\mathbf{\Gamma}}$ and $\mathbf{\Gamma}$. A sketch of the consistency of proposed estimators for these quantities is given below. First, consider $\tilde{\eta}_i$. From Eq. 4.19, which will be derived subsequently, we have

$$\tilde{\eta}_i = \lambda_i^{-1} \sum_{s=1}^{n} u_{is}^2 \sum_{t=1}^{n} X_s^\top \mathbf{I}_{p,q} X_t (1 - X_s^\top \mathbf{I}_{p,q} X_t).$$

Here entries u_{is} are the elements of the eigenvector $\boldsymbol{u}_i$ corresponding to the i-th largest eigenvalue of $\mathbf{P} = \mathbb{E}[\mathbf{A}]$. We can estimate $\boldsymbol{u}_i$ using the eigenvector $\hat{\boldsymbol{u}}_i$ associated with the i-th largest eigenvalue of $\mathbf{A}$, while we can estimate λ_i using $\hat{\lambda}_i$. Finally, to estimate X_s, we use the *adjacency spectral embedding* of $\mathbf{A}$ as discussed in Rubin-Delanchy et al. (2017). More specifically, let $\hat{\mathbf{X}} = (|\hat{\lambda}_1|^{1/2}\hat{\boldsymbol{u}}_1, \ldots, |\hat{\lambda}_d|^{1/2}\hat{\boldsymbol{u}}_d)$ be the $n \times d$ matrix whose columns are the eigenvectors $\hat{\boldsymbol{u}}_i$ of $\mathbf{A}$ subsequently *scaled* by the square root of the absolute value of the corresponding eigenvalues $\hat{\lambda}_i$. Let $\hat{X}_s$ denote the s-th row of $\hat{\mathbf{X}}$. Then, as a consequence of Rubin-Delanchy et al. (2017), we have

$$\max_{s,t} |\hat{X}_s^\top \mathbf{I}_{p,q} \hat{X}_t - X_s^\top \mathbf{I}_{p,q} X_t| \xrightarrow{\text{a.s.}} 0 \tag{2.10}$$

as $n \to \infty$. Equation 2.10 implies that the estimator

$$\hat{\eta}_i = \hat{\lambda}_i^{-1} \sum_{s=1}^{n} \hat{u}_{is}^2 \sum_{t=1}^{n} \hat{X}_s^\top \mathbf{I}_{p,q} \hat{X}_t (1 - \hat{X}_s^\top \mathbf{I}_{p,q} \hat{X}_t)$$

is consistent for $\tilde{\eta}_i$, i.e., that $\hat{\eta}_i - \tilde{\eta}_i \longrightarrow 0$ almost surely as $n \to \infty$.

The estimation procedure for $\tilde{\Gamma}_{ij}$ is similar. Specifically, from the discussion near the end of Section 4, it holds that

$$\begin{aligned} \tilde{\Gamma}_{ij} &= 2 \sum_k \sum_\ell u_{ik}^2 u_{j\ell}^2 p_{k\ell}(1 - p_{k\ell}) - \sum_k p_{kk}(1 - p_{kk}) u_{ik}^2 u_{j\ell}^2 \\ &= 2 \sum_k \sum_\ell u_{ik}^2 u_{j\ell}^2 p_{k\ell}(1 - p_{k\ell}) + o_{\mathbb{P}}(1). \end{aligned}$$

Recalling that $p_{k\ell} = X_k^\top \mathbf{I}_{p,q} X_\ell$, a consistent estimator for $\tilde{\Gamma}_{ij}$ is consequently given by

$$\hat{\Gamma}_{ij} = 2 \sum_k \sum_\ell \hat{u}_{ik}^2 \hat{u}_{j\ell}^2 \hat{X}_k^\top \mathbf{I}_{p,q} \hat{X}_\ell (1 - \hat{X}_k^\top \mathbf{I}_{p,q} \hat{X}_\ell).$$

3 Examples

This section provides several simulation examples to illustrate the theoretical results in Section 2. For the first example, consider a stochastic blockmodel with $K = 2$ blocks and parameters

$$\mathbf{B} = \begin{bmatrix} 0.3 & 0.5 \\ 0.5 & 0.3 \end{bmatrix}, \quad \boldsymbol{\pi} = (0.3, 0.7).$$

We sample $m = 1000$ graphs each having $n = 4000$ vertices from this stochastic blockmodel; we emphasize that for each graph, the number of vertices assigned to each block is a random variable and will vary between the sampled graphs. For each graph, we compute the two largest eigenvalues in magnitude; plots of the empirical distributions for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2)$ for these $m = 1000$ graphs are given in Fig. 1. Note that the eigenvalues (λ_1, λ_2) depend on the edge probability matrix which in turns depends on the number of vertices assigned to each block and thus will also vary between the sampled graphs; furthermore, with high probability λ_1 is positive and λ_2 is negative. For example, if the sampled graph has 1200 and 2800 vertices assigned to block 1 and block 2, respectively, then $(\lambda_1, \lambda_2) = (1547.42, -347.42)$ and if the sampled graph has 1250 and 2750 vertices assigned to block 1 and block 2, respectively then $(\lambda_1, \lambda_2) = (1553.94, -353.94)$.

Figure 1 indicates that the empirical distributions are well-approximated by the limiting normal distributions given in Theorem 2; in particular a collection of goodness-of-fit tests fail to reject the null hypothesis that the empirical distribution for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2)$ correspond to random samples from a bivariate normal distribution.

We also repeat the same simulation but for graphs with $n = 2000$ vertices and $n = 6000$ vertices. We then look at the sample covariance matrix for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2)$ for these values of $n \in \{2000, 4000, 6000\}$ and compare them to the theoretical covariances given in Eq. 2.3. The results are presented in Table 1. We see that since the eigenvalues of $\Delta\mathbf{I}_{p,q}$ are distinct in this case, the sample covariance matrices appear to converge to the theoretical covariance matrix.

For the next example, consider a stochastic blockmodel with parameters

$$\mathbf{B} = \begin{bmatrix} 0.5 & 0.3 & 0.3 \\ 0.3 & 0.5 & 0.3 \\ 0.3 & 0.3 & 0.5 \end{bmatrix}, \quad \boldsymbol{\pi} = (\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}).$$

This choice of parameters leads to Δ having repeated eigenvalues, explicitly $\frac{11}{30}$, $\frac{2}{30}$, and $\frac{2}{30}$. We once again sample $m = 1000$ graphs, each graph having

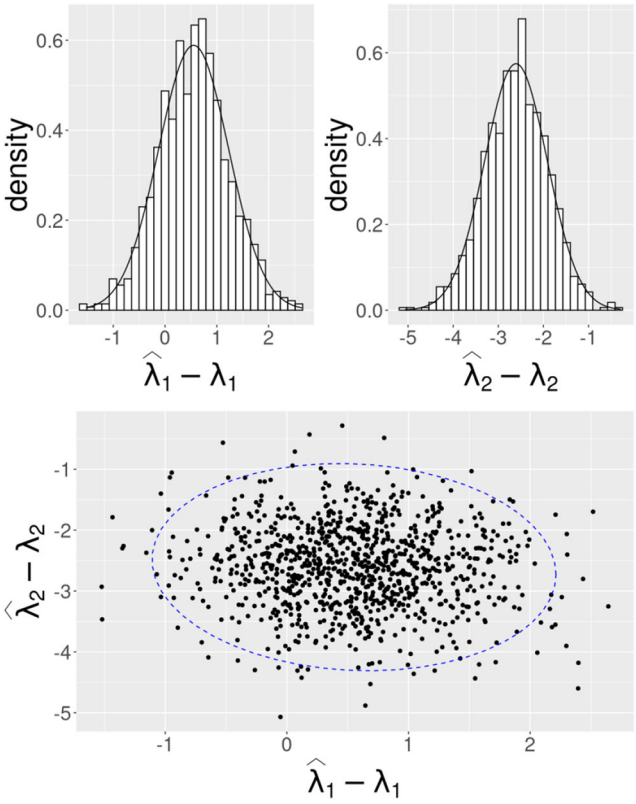

Figure 1: Plot of the empirical marginal distributions for $\hat{\lambda}_1 - \lambda_1$ and $\hat{\lambda}_2 - \lambda_2$, and the empirical joint distribution for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2)$. The empirical distributions are based on 1000 samples of two-block SBM graphs on $n = 4000$ vertices. The SBM parameters are $\mathbf{B} = \begin{pmatrix} 0.3 & 0.5 \\ 0.5 & 0.3 \end{pmatrix}$ and $\boldsymbol{\pi} = (0.3, 0.7)$. The dashed lines for the univariate plots are the probability density functions for the univariate normal distributions with parameters given in Theorem 2, while the dashed line for the bivariate plot demarcates the 95% theoretical confidence region for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2)$

$n = 4000$ vertices, from this stochastic blockmodel and plot the empirical distributions for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2, \hat{\lambda}_3 - \lambda_3)$ based on these m graph. The results are presented in Fig. 2. Once again, the empirical distributions are well-approximated by the appropriate univariate and multivariate normal distributions. Indeed, a suite of goodness-of-fit tests fails to reject the null hypothesis that the empirical distribution for $(\hat{\lambda}_i - \lambda_i)_{i=1}^3$ corresponds to random samples from a multivariate normal distribution.

Table 1: Empirical estimates of the covariance matrix for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2)$ as n varies

$n = 2000$	$n = 4000$	$n = 6000$	Theoretical
$\begin{bmatrix} 0.452 & -0.027 \\ -0.027 & 0.405 \end{bmatrix}$	$\begin{bmatrix} 0.455 & -0.038 \\ -0.038 & 0.455 \end{bmatrix}$	$\begin{bmatrix} 0.464 & -0.034 \\ -0.034 & 0.446 \end{bmatrix}$	$\begin{bmatrix} 0.457 & -0.037 \\ -0.037 & 0.457 \end{bmatrix}$

For each choice of n, the estimates are obtained from 1000 Monte Carlo replicates using the same setting as for Fig. 1. Here, the matrix $\mathbf{\Delta I}_{p,q}$ has distinct eigenvalues and hence the empirical covariances are approximately the same as the theoretical covariance matrix from Eq. 2.3

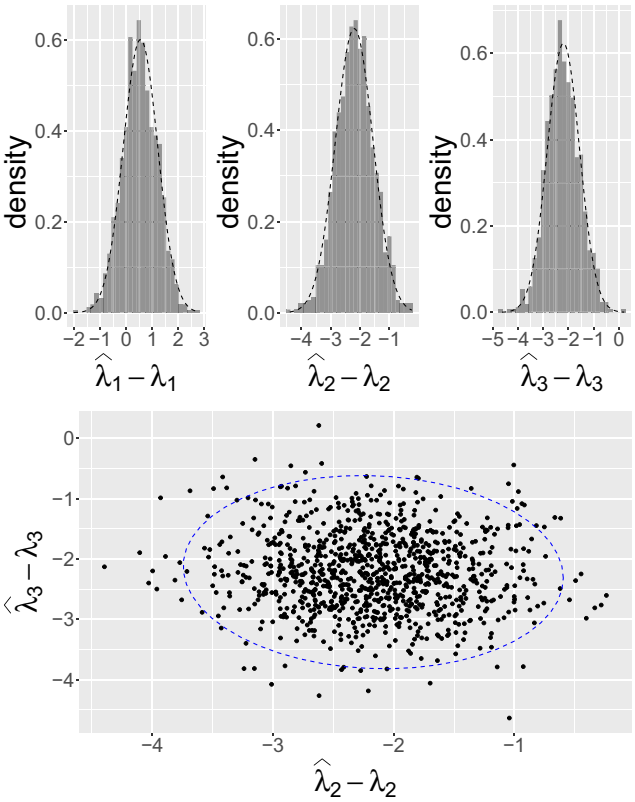

Figure 2: Plot of the empirical marginal distributions for $\hat{\lambda}_1 - \lambda_1$, $\hat{\lambda}_2 - \lambda_2$, and $\hat{\lambda}_3 - \lambda_3$, and the empirical joint distribution for $(\hat{\lambda}_2 - \lambda_2, \hat{\lambda}_3 - \lambda_3)$. These empirical distributions are based on 1000 samples of a three-block SBM graphs on $n = 4000$ vertices. The SBM parameters are $\mathbf{B} = 0.2\mathbf{I} + 0.3\mathbf{1}\mathbf{1}^\top$ and $\boldsymbol{\pi} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The dashed lines for the univariate plots are the marginal probability density functions for the normal distributions with mean $\tilde{\eta}_i$ and variance σ_i^2 as specified in Theorem 3. The dashed line for the bivariate plot is the 95% confidence ellipse. Note that Δ has repeated eigenvalues given by $\frac{11}{30}$, $\frac{2}{30}$ and $\frac{2}{30}$

Next, we repeat the same simulation but now for graphs with $n = 2000$ and $n = 6000$ vertices. We then look at the sample covariance matrix for $(\hat{\lambda}_i - \lambda_i)_{i=1}^3$ for each $n \in \{2000, 4000, 6000\}$ and compare the sample covariance matrices to the theoretical covariances given in Eq. 2.3; note that the expression in Eq. 2.3 assumes that the eigenvalues of Δ are all distinct and is

thus not necessarily appropriate for the current setting. The results are presented in Table 2. Table 2 indicates that, in contrast to Table 1, the sample covariance matrices fluctuates and, furthermore, do not appear to converge to the theoretical covariance matrix given in Eq. 2.3. While this could due to finite sample noise, we believe that it may also be a consequence of the fact that the eigenvalues of Δ are not distinct (see also the discussion after the statement of Theorem 3). The question of whether or not $\tilde{\mathbf{\Gamma}}$ converge to some deterministic matrix as $n \to \infty$ is currently an open question.

4 Proof of Theorem 2 and Theorem 3

Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d$ be the eigenvectors corresponding to the non-zero eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ of $\mathbf{P}$. Similarly, let $\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \ldots \hat{\boldsymbol{u}}_d$ be the eigenvectors corresponding to the eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_d$ of $\mathbf{A}$. A sketch of the proof of Theorem 2 and Theorem 3 is as follows. First we derive an approximation of $\hat{\lambda}_i - \lambda_i$ as a linear combination of two quadratic forms, $\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i$ and $\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i$, namely

$$\hat{\lambda}_i - \lambda_i = \frac{\lambda_i}{\hat{\lambda}_i}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i + \frac{\lambda_i}{\hat{\lambda}_i^2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i + O_{\mathbb{P}}(n^{-1}). \tag{4.1}$$

The quantity $\lambda_i^{-1}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i$ is a function of the $n(n+1)/2$ independent random variables $\{a_{rs} - p_{rs}\}_{r\le s}$ and hence is concentrated around its expectation, i.e.,

$$\frac{\lambda_i}{\hat{\lambda}_i^2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i = \mathbb{E}\left[\lambda_i^{-1}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i\right] + O_{\mathbb{P}}(n^{-1/2}), \tag{4.2}$$

where the expectation is with respect to $\mathbf{A}$ and conditional on $\mathbf{P}$.

Letting $\tilde{\eta}_i = \mathbb{E}[\lambda_i^{-1}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i]$, after some straightforward algebraic manipulations we obtain the expression for $\tilde{\eta}_i$ given in Eq. 2.8. When the eigenvalues of $\Delta\mathbf{I}_{p,q}$ are distinct, we derive the almost sure limit $\tilde{\eta}_i \longrightarrow \eta_i$, where η_i is defined in Eq. 2.2.

Next, with u_{is} denoting the s-th element of the vector $\boldsymbol{u}_i$, the expression

$$\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i = \sum_{r<s} 2u_{is}u_{ir}(a_{rs}-p_{rs}) + \sum_r u_{ir}^2(a_{rr}-p_{rr})$$

is, conditional on $\mathbf{X}$, a sum of independent mean zero random variables. The Lindeberg–Feller central limit theorem therefore yields that, for $n \to \infty$, then

$$\tilde{\Gamma}_{ii}^{-1/2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{4.3}$$

Table 2: Empirical estimates of the covariance matrix for $(\hat{\lambda}_1 - \lambda_1, \hat{\lambda}_2 - \lambda_2, \hat{\lambda}_3 - \lambda_3)$ as n changes

$n = 2000$			$n = 4000$			$n = 6000$			Theoretical		
0.432	0.024	0.048	0.436	0.040	0.045	0.430	0.028	0.043	0.446	0.027	0.027
0.024	0.453	0.023	0.040	0.455	−0.002	0.028	0.489	0.017	0.027	0.460	0.013
0.048	0.023	0.421	0.045	−0.002	0.500	0.043	0.017	0.432	0.027	0.013	0.460

The estimates, for each value of n, are obtained from 1000 Monte Carlo replicates using the same setting as for Fig. 2. Here, the matrix Δ has repeated eigenvalues and hence the empirical covariances fluctuate and do not appear to converge to the theoretical covariance matrix given in Eq. 2.3 which assumes distinct eigenvalues

Here $\tilde{\Gamma}_{ii}$ is as defined in Eq. 2.9. Consequently, for each $1 \leq i \leq d$, then $\tilde{\Gamma}_{ii}^{-1/2}(\lambda_i - \hat{\lambda}_i - \tilde{\eta}_i) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ as $n \to \infty$. When the eigenvalues of $\Delta \mathbf{I}_{p,q}$ are distinct, then $\tilde{\Gamma}_{ii} \longrightarrow \Gamma_{ii}$ almost surely, where Γ_{ii} is given in Eq. 2.3. The joint distribution of $(\hat{\lambda}_i - \lambda_i)_{i=1}^d$ in both Theorem 2 and Theorem 3 then follows from the Cramer–Wold device.

We now provide detailed derivations of Eqs. 4.1 through 4.3.

Proof of Eq. 4.1 Given $1 \leq i \leq d$, we have

$$\big(\hat{\lambda}_i \mathbf{I} - (\mathbf{A} - \mathbf{P})\big)\hat{\boldsymbol{u}}_i = \mathbf{A}\hat{\boldsymbol{u}}_i - (\mathbf{A} - \mathbf{P})\hat{\boldsymbol{u}}_i = \mathbf{P}\hat{\boldsymbol{u}}_i = \left(\sum_{j=1}^{r} \lambda_j \boldsymbol{u}_j \boldsymbol{u}_j^\top\right) \hat{\boldsymbol{u}}_i.$$

Now suppose that $\hat{\lambda}_i \mathbf{I} - (\mathbf{A} - \mathbf{P})$ is invertible; this holds with high probability for sufficiently large n. Then, multiplying both sides of the above display by $\boldsymbol{u}_i^\top (\hat{\lambda}_i \mathbf{I} - (\mathbf{A} - \mathbf{P}))^{-1}$ on the left and using the von Neumann identity, $(\mathbf{I} - \mathbf{X})^{-1} = \mathbf{I} + \sum_{k=1}^{\infty} \mathbf{X}^k$ for $\|\mathbf{X}\| < 1$, yields

$$\begin{aligned}
\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_i &= \sum_{j=1}^{d} \lambda_j \boldsymbol{u}_i^\top (\hat{\lambda}_i \mathbf{I} - (\mathbf{A} - \mathbf{P}))^{-1} \boldsymbol{u}_j \boldsymbol{u}_j^\top \hat{\boldsymbol{u}}_i \qquad (4.4) \\
&= \sum_{j=1}^{d} \lambda_j \hat{\lambda}_i^{-1} \boldsymbol{u}_i^\top (\mathbf{I} - \hat{\lambda}_i^{-1}(\mathbf{A} - \mathbf{P}))^{-1} \boldsymbol{u}_j \boldsymbol{u}_j^\top \hat{\boldsymbol{u}}_i \\
&= \sum_{j=1}^{d} \lambda_j \hat{\lambda}_i^{-1} \boldsymbol{u}_i^\top \left(\mathbf{I} + \sum_{k=1}^{\infty} \hat{\lambda}_i^{-k} (\mathbf{A} - \mathbf{P})^k\right) \boldsymbol{u}_j \boldsymbol{u}_j^\top \hat{\boldsymbol{u}}_i \\
&= \frac{\lambda_i}{\hat{\lambda}_i} \boldsymbol{u}_i^\top \left(\mathbf{I} + \sum_{k=1}^{\infty} \hat{\lambda}_i^{-k} (\mathbf{A} - \mathbf{P})^k\right) \boldsymbol{u}_i \boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_i \\
&\quad + \sum_{j \neq i} \frac{\lambda_j}{\hat{\lambda}_i} \boldsymbol{u}_i^\top \left(\sum_{k=1}^{\infty} \hat{\lambda}_i^{-k} (\mathbf{A} - \mathbf{P})^k\right) \boldsymbol{u}_j \boldsymbol{u}_j^\top \hat{\boldsymbol{u}}_i. \qquad (4.5)
\end{aligned}$$

First suppose that all of the eigenvalues of $\Delta \mathbf{I}_{p,q}$ are simple eigenvalues. The eigenvalues of $\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top$ are then well-separated, i.e., $\min_{j \neq i} |\lambda_i - \lambda_j| = \Omega_{\mathbb{P}}(n)$ for $1 \leq i \neq j \leq d$. The Davis–Kahan $\sin\Theta$ theorem (Davis and Kahan, 1970; Yu et al., 2015) then implies that for some positive constant C,

$$1 - \boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_i = \tfrac{1}{2}\|\boldsymbol{u}_i - \hat{\boldsymbol{u}}_i\|^2 \leq \frac{C^2 \|\mathbf{A} - \mathbf{P}\|^2}{\min\{|\lambda_i - \lambda_{i+1}|^2, |\lambda_{i-1} - \lambda_i|^2\}} = O_{\mathbb{P}}(n^{-1}), \qquad (4.6)$$

$$|\boldsymbol{u}_j^\top \hat{\boldsymbol{u}}_i| = \frac{|u_i^\top(\mathbf{A}-\mathbf{P})\hat{u}_j|}{|\hat{\lambda}_j-\lambda_i|} = \frac{|\boldsymbol{u}_j^\top(\mathbf{A}-\mathbf{P})\hat{\boldsymbol{u}}_i}{|\lambda_j-\hat{\lambda}_i|} + \frac{|\boldsymbol{u}_j^\top(\mathbf{A}-\mathbf{P})(\mathbf{I}-\boldsymbol{u}_j\boldsymbol{u}_j^\top)\hat{\boldsymbol{u}}_i|}{|\lambda_j-\hat{\lambda}_i|}$$
$$= O_{\mathbb{P}}(n^{-1}). \tag{4.7}$$

Dividing both sides of the above display equation by $\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_i$ yields

$$1 = \frac{\lambda_i}{\hat{\lambda}_i} + \frac{\lambda_i}{\hat{\lambda}_i}\boldsymbol{u}_i^\top\Big(\sum_{k=1}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\Big)\boldsymbol{u}_i + \sum_{j\neq i}\frac{\lambda_j}{\hat{\lambda}_i}\boldsymbol{u}_i^\top\Big(\sum_{k=1}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\Big)\boldsymbol{u}_j\left(\frac{\boldsymbol{u}_j^\top\hat{\boldsymbol{u}}_i}{\boldsymbol{u}_i^\top\hat{\boldsymbol{u}}_i}\right).$$

Equivalently,

$$\hat{\lambda}_i-\lambda_i = \lambda_i\boldsymbol{u}_i^\top\Big(\sum_{k=1}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\Big)\boldsymbol{u}_i + \sum_{j\neq i}\frac{\lambda_j}{\hat{\lambda}_i}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_j\left(\frac{\boldsymbol{u}_j^\top\hat{\boldsymbol{u}}_i}{\boldsymbol{u}_i^\top\hat{\boldsymbol{u}}_i}\right)$$
$$+\sum_{j\neq i}\lambda_j\boldsymbol{u}_i^\top\Big(\sum_{k=2}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\Big)\boldsymbol{u}_j\left(\frac{\boldsymbol{u}_j^\top\hat{\boldsymbol{u}}_i}{\boldsymbol{u}_i^\top\hat{\boldsymbol{u}}_i}\right). \tag{4.8}$$

Now $\lambda_i^{-1}\hat{\lambda}_j = O_{\mathbb{P}}(1)$,
and by Hoeffding's inequality, $\boldsymbol{u}_j^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i = O_{\mathbb{P}}(1)$.
Since $\boldsymbol{u}_j^\top\hat{\boldsymbol{u}}_i = O_{\mathbb{P}}(n^{-1})$, we have

$$\sum_{j\neq i}\frac{\lambda_j}{\hat{\lambda}_i}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_j\left(\frac{\boldsymbol{u}_j^\top\hat{\boldsymbol{u}}_i}{\boldsymbol{u}_i^\top\hat{\boldsymbol{u}}_i}\right) = O_{\mathbb{P}}(n^{-1}).$$

Next we note that $\|\sum_{k=2}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\|$ can be bounded as

$$\left\|\sum_{k=2}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\right\| \leq \frac{\|\hat{\lambda}_i^{-2}(\mathbf{A}-\mathbf{P})^2\|}{1-\|\hat{\lambda}_i^{-1}(\mathbf{A}-\mathbf{P})\|} = O_{\mathbb{P}}(\hat{\lambda}_i^{-1}). \tag{4.9}$$

We thus have

$$\sum_{j\neq i}\lambda_j\boldsymbol{u}_i^\top\Big(\sum_{k=2}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\Big)\boldsymbol{u}_j\left(\frac{\boldsymbol{u}_j^\top\hat{\boldsymbol{u}}_i}{\boldsymbol{u}_i^\top\hat{\boldsymbol{u}}_i}\right) = O_{\mathbb{P}}(n^{-1}).$$

The above bounds together imply

$$\hat{\lambda}_i-\lambda_i = \lambda_i\boldsymbol{u}_i^\top\Big(\sum_{k=1}^{\infty}\hat{\lambda}_i^{-k}(\mathbf{A}-\mathbf{P})^k\Big)\boldsymbol{u}_i + O_{\mathbb{P}}(n^{-1}). \tag{4.10}$$

Similar to the derivation of Eq. 4.9, we also show that

$$\left\| \sum_{k=3}^{\infty} \hat{\lambda}_i^{-k} (\mathbf{A} - \mathbf{P})^k \right\| \leq C \| \hat{\lambda}_i^{-3} (\mathbf{A} - \mathbf{P})^3 \| \leq C \hat{\lambda}_i^{-3/2} \tag{4.11}$$

with high probability and thus Eqs. 4.10 and 4.11 imply

$$\hat{\lambda}_i - \lambda_i = \frac{\lambda_i}{\hat{\lambda}_i} \boldsymbol{u}_i^\top (\mathbf{A} - \mathbf{P}) \boldsymbol{u}_i + \frac{\lambda_i}{\hat{\lambda}_i^2} \boldsymbol{u}_i^\top (\mathbf{A} - \mathbf{P})^2 \boldsymbol{u}_i + O_{\mathbb{P}}(n^{-1}). \tag{4.12}$$

This establishes (4.1).

We now consider the case where the i-th eigenvalue of $\Delta \mathbf{I}_{p,q}$ has multiplicity $r_i \geq 2$. Let $\mathcal{S}_i$ be the index set of the r_i eigenvalues λ_j of $\mathbf{P} = \mathbf{X} \mathbf{I}_{p,q} \mathbf{X}^\top$ that are closest to $n\lambda_i(\Delta \mathbf{I}_{p,q})$, i.e.,

$$\max_{j \in \mathcal{S}_i} |\lambda_j - n\lambda_i(\Delta \mathbf{I}_{p,q})| \leq \min_{k \notin \mathcal{S}_i} |\lambda_k - n\lambda_i(\Delta \mathbf{I}_{p,q})|; \quad |\mathcal{S}_i| = r_i.$$

Let $\mathbf{U}_{\mathcal{S}_i}$ denote the $n \times r_i$ matrix whose columns are the eigenvectors corresponding to the λ_j with $j \in \mathcal{S}_i$. We have $i \in \mathcal{S}_i$ with high probability for sufficiently large n. Furthermore, $|\lambda_j - \lambda_k| = \Omega_{\mathbb{P}}(n)$ for all $j \in \mathcal{S}_i$ and $k \notin \mathcal{S}_i$. A similar argument to that for Eq. 4.7 establishes $\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_k = O_{\mathbb{P}}(n^{-1})$ for all $k \notin \mathcal{S}_i$. We now consider $\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_j$ for $j \in \mathcal{S}_i, j \neq i$. We have

$$\begin{aligned}
\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_j &= \frac{\boldsymbol{u}_i^\top \mathbf{A} \hat{\boldsymbol{u}}_j - \boldsymbol{u}_i^\top \mathbf{P} \hat{\boldsymbol{u}}_j}{\hat{\lambda}_j - \lambda_i} \\
&= \frac{\boldsymbol{u}_i^\top (\mathbf{A} - \mathbf{P}) \mathbf{U}_{\mathcal{S}_i} \mathbf{U}_{\mathcal{S}_i}^\top \hat{\boldsymbol{u}}_j}{\hat{\lambda}_j - \lambda_i} + \frac{\boldsymbol{u}_i^\top (\mathbf{A} - \mathbf{P})(\mathbf{I} - \mathbf{U}_{\mathcal{S}_i} \mathbf{U}_{\mathcal{S}_i}^\top) \hat{\boldsymbol{u}}_j}{\hat{\lambda}_j - \lambda_i} \\
&= \frac{n^{-1/2} \boldsymbol{u}_i^\top (\mathbf{A} - \mathbf{P}) \mathbf{U}_{\mathcal{S}_i} \mathbf{U}_{\mathcal{S}_i}^\top \hat{\boldsymbol{u}}_j}{n^{-1/2}(\hat{\lambda}_j - \lambda_j) + n^{-1/2}(\lambda_j - \lambda_i)} \\
&\quad + \frac{n^{-1/2} \boldsymbol{u}_i^\top (\mathbf{A} - \mathbf{P})(\mathbf{I} - \mathbf{U}_{\mathcal{S}_i} \mathbf{U}_{\mathcal{S}_i}^\top) \hat{\boldsymbol{u}}_j}{n^{-1/2}(\hat{\lambda}_j - \lambda_j) + n^{-1/2}(\lambda_j - \lambda_i)}.
\end{aligned} \tag{4.13}$$

By Hoeffding's inequality (or Bernstein's inequality), $\boldsymbol{u}_i^\top (\mathbf{A} - \mathbf{P}) \mathbf{U}_{\mathcal{S}_i} = O_{\mathbb{P}}(1)$. Since $j \in \mathcal{S}_i$, we have $\|(\mathbf{I} - \mathbf{U}_{\mathcal{S}_i} \mathbf{U}_{\mathcal{S}_i}^\top) \hat{\boldsymbol{u}}_j\| = O_{\mathbb{P}}(n^{-1/2})$ by the Davis–Kahan theorem. We then bound $\hat{\lambda}_j - \lambda_j$ using the following result (Cape et al., 2017, Theorem 3.7) (see also O'Rourke and Vu, 2018, Theorem 23).

THEOREM 4. *Let* $\mathbf{A}$ *and be a* $n \times n$ *symmetric random matrix with* $\mathbf{A}_{ij} \sim$ Bernoulli$(\mathbf{P}_{ij})$ *for* $i \leq j$ *and the entries* $\{\mathbf{A}_{ij}\}$ *are independent. Denote the* $d+1$ *largest singular values of* $\mathbf{A}$ *by* $0 \leq \hat{\sigma}_{d+1} < \hat{\sigma}_d \leq \hat{\sigma}_{d-1} \leq \cdots \leq \hat{\sigma}_1$, *and*

denote the $d+1$ *largest singular values of* $\mathbf{P}$ *by* $0 \le \sigma_{d+1} < \sigma_d \le \sigma_{d-1} \le \cdots \le \sigma_1$. *Suppose that* $\Upsilon = \max_i \sum_j \mathbf{P}_{ij} = \omega(\log^4 n)$, $\sigma_1 \ge C\Upsilon$, $\sigma_{d+1} \le c\Upsilon$ *for some absolute constants* $C > c > 0$. *Then for each* $k \in \{1, 2, \ldots, d\}$, *there exists some positive constant* $c_{k,d}$ *such that as* $n \to \infty$, *with probability at least* $1 - n^{-3}$, *we have*

$$|\hat{\sigma}_k - \sigma_k| \le c_{k,d} \log n.$$

We thus have

$$|\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_j| = \frac{n^{-1/2} \cdot O_{\mathbb{P}}(1)}{n^{-1/2} \cdot O_{\mathbb{P}}(\log n) + n^{-1/2}(\lambda_j - \lambda_i)}.$$

The next step is to analyze $n^{-1/2}(\lambda_j - \lambda_i)$. In the present setting, $\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top$ can be viewed as a kernel matrix with symmetric kernel $h(X_i, X_j) = X_i^\top \mathbf{I}_{p,q} X_j$ where $X_i, X_j \sim F$. As h is finite-rank, let $(\phi_i, \lambda_i(\mathbf{I}_{p,q}\Delta))_{i=1}^d$ denote the eigenvalues and associated eigenfunctions of the integral operator $\mathcal{K}_h \colon L_2(F) \to L_2(F)$, i.e.,

$$\mathcal{K}_h \phi_i(x) \coloneqq \int h(x, y)\phi_i(y)\, \mathrm{d}F(y) = \lambda_i(\mathbf{I}_{p,q}\Delta)\phi_i(x).$$

Following Koltchinskii and Giné (2000), let Ψ_i denote the $r_i \times r_i$ random symmetric matrix whose upper triangular entries (including the diagonal) are (jointly) distributed as multivariate normal with mean $\mathbf{0}$ and $r_i(r_i+1)/2 \times r_i(r_i+1)/2$ covariance matrix with entries of the form

$$\begin{aligned}\mathrm{Cov}(\Psi_i(s,t), \Psi_i(u,v)) = &\int \phi_s(y)\phi_t(y)\phi_u(y)\phi_v(y)\mathrm{d}F(y) \\ &- \int \phi_s(y)\phi_t(y)\mathrm{d}F(y) \int \phi_u(y)\phi_v(y)\mathrm{d}F(y) \quad (4.14)\end{aligned}$$

for $1 \le s \le t \le r_i, 1 \le u \le v \le r_i$. With a slight abuse of notation, the collection $\{\phi_s\}_{s \le r_i}$ denotes the r_i eigenfunctions of $\mathcal{K}_h$ associated with the eigenvalue $\lambda_i(\mathbf{I}_{p,q}\Delta)$. A simplification of the statement of Theorem 5.1 in Koltchinskii and Giné (2000) to the setting where h is a finite-rank kernel yields

$$n^{1/2}(\lambda_j/n - \lambda_i(\mathbf{I}_{p,q}\Delta))_{j \in \mathcal{S}_i} \to \lambda_i(\mathbf{I}_{p,q}\Delta) \times (\lambda_s(\Psi_i))_{1 \le s \le r_i}$$

as $n \to \infty$; here we use the notation $\lambda_s(\mathbf{M})$ to denote the s-th largest eigenvalue, in modulus, of the matrix $\mathbf{M}$.

The joint distribution of $\{n^{-1/2}(\lambda_s - n\lambda_i(\Delta \mathbf{I}_{p,q}))\}_{s\in\mathcal{S}_i}$ converges to a non-degenerate limiting distribution and hence the limiting distribution of $n^{-1/2}(\lambda_i - \lambda_j)$ is also non-degenerate. We therefore have

$$|\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_j| = \frac{n^{-1/2}\cdot O_{\mathbb{P}}(1)}{n^{-1/2}\cdot O_{\mathbb{P}}(\log n) + n^{-1/2}(\lambda_j - \lambda_i)} = o_{\mathbb{P}}(1); \quad j \in \mathcal{S}_i, \quad j \neq i.$$

Finally, we note that there exists an orthogonal matrix $\mathbf{W}$ such that $\|\mathbf{U}^\top \hat{\mathbf{U}} - \mathbf{W}\| = O_{\mathbb{P}}(n^{-1})$. Hence, for any $1 \le i \le d$, $\sum_{j=1}^{d}(\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_j)^2 = 1 + O_{\mathbb{P}}(n^{-1})$. From the earlier bound for $\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_j$ when $j \neq i$, we have $\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_i = 1 - o_{\mathbb{P}}(1)$. In summary, when the eigenvalues of $\mathbf{I}_{p,q}\Delta$ are not all simple eigenvalues, we have (in place of Eqs. 4.6 and 4.7) the bounds

$$\boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_i = 1 - o_{\mathbb{P}}(1); \quad \boldsymbol{u}_i^\top \hat{\boldsymbol{u}}_j = o_{\mathbb{P}}(1). \tag{4.15}$$

Consequently, Eq. 4.8 still holds and the remaining steps in the derivation of Eq. 4.12 can be adapted to yield

$$\hat{\lambda}_i - \lambda_i = \frac{\lambda_i}{\hat{\lambda}_i}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i + \frac{\lambda_i}{\hat{\lambda}_i^2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i + o_{\mathbb{P}}(1).$$

Proof of Eq. 4.2 Let $Z = \lambda_i^{-1}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i$. To derive (4.2), we show the concentration of Z around $\mathbb{E}[Z]$, where the expectation is taken with respect to $\mathbf{A}$, conditional on $\mathbf{P}$, using a log-Sobolev concentration inequality from Boucheron et al. (2003). More specifically, let $\mathbf{A}' = (a'_{rs})$ be an independent copy of $\mathbf{A}$, i.e., the upper triangular entries of $\mathbf{A}'$ are independent Bernoulli random variables with mean parameters $\{p_{rs}\}_{r\le s}$. For any pair of indices (r,s), let $\mathbf{A}^{(rs)}$ be the matrix obtained by replacing the (r,s) and (s,r) entries of $\mathbf{A}$ by a'_{ij} and let $Z^{(rs)} = \lambda_i^{-1}\boldsymbol{u}_i^\top(\mathbf{A}^{(rs)}-\mathbf{P})^2\boldsymbol{u}_i$. Theorem 5 below, based on Chapter 6 of Boucheron et al. (2003), states the following.

THEOREM 5. *Assume that there exist positive constants a and b such that*

$$\sum_{r\le s}(Z - Z^{(rs)})^2 \le aZ + b.$$

Then for all $t > 0$,

$$\mathbb{P}[Z - \mathbb{E}[Z] \ge t] \le \exp\Big(\frac{-t^2}{4a\mathbb{E}[Z] + 4b + 2at}\Big), \tag{4.16}$$

$$\mathbb{P}[Z - \mathbb{E}[Z] \le -t] \le \exp\Big(\frac{-t^2}{4a\mathbb{E}[Z]}\Big). \tag{4.17}$$

The main technical step is then to bound $\sum_{r\leq s}(Z-Z^{(rs)})^2$. An identical argument to the one provided in Lemma A.6 of Tang et al. (2017) yields

$$\sum_{r\leq s}(Z-Z^{(rs)})^2 \leq a\lambda_i^{-1}Z+b$$

for some positive constants a and b. Theorem 5 therefore implies

$$|Z-\mathbb{E}[Z]| \leq \sqrt{\mathbb{E}[Z]}\times O_{\mathbb{P}}(n^{-1/2}) = O_{\mathbb{P}}(n^{-1/2}),$$

as desired.

We now evaluate $\mathbb{E}[Z]=\mathbb{E}[\lambda_i^{-1}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i]$. Let ζ_{rs} denote the rs-th entry of $\mathbb{E}[(\mathbf{A}-\mathbf{P})^2]$. Then ζ_{rs} is of the form

$$\zeta_{rs}=\sum_t \mathbb{E}[(a_{rt}-p_{rt})(a_{st}-p_{st})] = \begin{cases} 0 & \text{if } r\neq s,\\ \sum_t p_{rt}(1-p_{rt}) & \text{if } r=s. \end{cases} \tag{4.18}$$

We therefore have

$$\begin{aligned}\mathbb{E}[Z] &= \lambda_i^{-1}\boldsymbol{u}_i^\top\mathbb{E}[(\mathbf{A}-\mathbf{P})^2]\boldsymbol{u}_i = \lambda_i^{-1}\sum_{s=1}^n u_{is}^2\sum_{t=1}^n p_{st}(1-p_{st})\\ &= \lambda_i^{-1}\sum_{s=1}^n u_{is}^2\sum_{t=1}^n X_s^\top\mathbf{I}_{p,q}X_t(1-X_s^\top\mathbf{I}_{p,q}X_t).\end{aligned} \tag{4.19}$$

Let $\tilde{\lambda}_i$ and $\boldsymbol{v}_i$ be an eigenvalue–eigenvector pair for the eigenvalue problem

$$(\mathbf{X}^\top\mathbf{X})^{1/2}\mathbf{I}_{p,q}(\mathbf{X}^\top\mathbf{X})^{1/2}\,\boldsymbol{v}=\tilde{\lambda}\boldsymbol{v}. \tag{4.20}$$

If $\tilde{\lambda}_i$ and $\boldsymbol{v}_i$ satisfy (4.20) then $\tilde{\lambda}_i$ and $\boldsymbol{u}_i=\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i$ are an eigenvalue–eigenvector pair for the eigenvalue problem $\mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top\boldsymbol{u}=\tilde{\lambda}\boldsymbol{u}$. Conversely, if $\tilde{\lambda}_i\neq 0$ and $\boldsymbol{u}_i$ are an eigenvalue–eigenvector pair for $\mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^\top$, then $\tilde{\lambda}_i$ and $\boldsymbol{v}_i=(\mathbf{X}^\top\mathbf{X})^{-1/2}\mathbf{X}^\top\boldsymbol{u}_i$ satisfies (4.20).

In addition, if the vectors $\{\boldsymbol{v}_i\}_{i=1}^d$ are mutually orthonormal, then the vectors $\{\boldsymbol{u}_i\}_{i=1}^d$ are also mutually orthonormal. Hence,

$$\begin{aligned}\mathbb{E}[Z] &= \tfrac{1}{\lambda_i}\sum_{s=1}^n(\boldsymbol{v}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}X_s)^2\sum_{t=1}^n X_s^\top\mathbf{I}_{p,q}X_t(1-X_s^\top\mathbf{I}_{p,q}X_t)\\ &= \tfrac{1}{\lambda_i}\sum_{s=1}^n\boldsymbol{v}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}X_sX_s^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_iX_s^\top\mathbf{I}_{p,q}\big(\sum_{t=1}^n(X_t-X_tX_t^\top\mathbf{I}_{p,q}X_s)\big)\\ &= \tfrac{n}{\lambda_i}\Big(\tfrac{1}{n}\sum_{s=1}^n\boldsymbol{v}_i^\top(\tfrac{\mathbf{X}^\top\mathbf{X}}{n})^{-1/2}X_sX_s^\top(\tfrac{\mathbf{X}^\top\mathbf{X}}{n})^{-1/2}\boldsymbol{v}_iX_s^\top\mathbf{I}_{p,q}\big(\tfrac{1}{n}\sum_{t=1}^n(X_t-X_tX_t^\top\mathbf{I}_{p,q}X_s)\big)\Big)\\ &= \tilde{\eta}_i.\end{aligned}$$

By the strong law of large numbers

$$\frac{1}{n}\sum_t X_t \to \mu, \qquad \frac{\mathbf{X}^\top\mathbf{X}}{n} = \frac{1}{n}\sum_t X_t X_t^\top \to \Delta, \qquad \frac{\lambda_i}{n} \to \lambda_i(\Delta\mathbf{I}_{p,q})$$

almost surely as $n \to \infty$. Furthermore, if $\lambda_i(\mathbf{I}_{p,q}\Delta)$ is a simple eigenvalue, then also $\boldsymbol{v}_i \to \boldsymbol{\xi}_i$ as $n \to \infty$. Therefore, when $\lambda_i(\Delta\mathbf{I}_{p,q})$ is a simple eigenvalue, we have

$$\mathbb{E}[Z] \to \frac{1}{\lambda_i(\Delta\mathbf{I}_{p,q})}\mathbb{E}\left[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_i(X^\top\mathbf{I}_{p,q}\mu - X^\top\mathbf{I}_{p,q}\Delta\mathbf{I}_{p,q}X)\right] \tag{4.21}$$

as desired.

Proof of Eq. 4.3 First recall that, conditional on $\mathbf{P}$, $\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i$ is a sum of independent mean zero random variables. Therefore, by the Lindeberg–Feller central limit theorem for triangular arrays, we have that $\tilde{\Gamma}_{ii}^{-1/2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i$ converges in distribution to a standard normal random variable, where here $\tilde{\Gamma}_{ii} = \mathrm{Var}[\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i]$. All that remains is to evaluate $\tilde{\Gamma}_{ii}$. Since $\mathbf{A}-\mathbf{P}$ is symmetric, we have

$$\begin{aligned}
\tilde{\Gamma}_{ii} &= \mathrm{Var}\Big[\sum_{k<\ell}(a_{k\ell}-p_{k\ell})(u_{ik}u_{i\ell}+u_{i\ell}u_{ik}) + \sum_k (a_{kk}-p_{kk})u_{ik}^2\Big] \\
&= \sum_{k<\ell} 4u_{ik}^2u_{i\ell}^2p_{k\ell}(1-p_{k\ell}) + \sum_k p_{kk}(1-p_{kk})u_{ik}^4 \\
&= 2\sum_k\sum_\ell u_{ik}^2u_{i\ell}^2p_{k\ell}(1-p_{k\ell}) - \sum_k p_{kk}(1-p_{kk})u_{ik}^4 \\
&= 2\sum_k\sum_\ell (X_k^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i)^2(X_\ell^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i)^2X_k^\top\mathbf{I}_{p,q}X_\ell(1-X_k^\top\mathbf{I}_{p,q}X_\ell) + o_{\mathbb{P}}(1). \\
&= 2\Big(\sum_k (X_k^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i)^2X_k^\top\Big)\mathbf{I}_{p,q}\Big(\sum_\ell (X_\ell^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i)^2\Big) + o_{\mathbb{P}}(1) \\
&- 2\mathrm{tr}\Big(\sum_k (X_k^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i)^2X_kX_k^\top\Big)\mathbf{I}_{p,q}\Big(\sum_\ell (X_\ell^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i)^2X_\ell X_\ell^\top\Big)\mathbf{I}_{p,q}.
\end{aligned}$$

When $\lambda_i(\mathbf{I}_{p,q}\Delta)$ is a simple eigenvalue, the strong law of large numbers and Slutsky's theorem together imply the almost sure convergence

$$\begin{aligned}
\tilde{\Gamma}_{ii} &\longrightarrow (\mathbb{E}[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_iX]^\top\mathbf{I}_{p,q}\mathbb{E}[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_iX]) \\
&- \mathrm{tr}\big(\mathbb{E}[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_iXX^\top]\mathbf{I}_{p,q}\mathbb{E}[\boldsymbol{\xi}_i^\top\Delta^{-1/2}XX^\top\Delta^{-1/2}\boldsymbol{\xi}_iXX^\top]\mathbf{I}_{p,q}\big) = \Gamma_{ii}.
\end{aligned}$$

One final application of Slutsky's theorem yields $\frac{\lambda_i}{\hat{\lambda}_i}\boldsymbol{u}_i^\top(\mathbf{A}–\mathbf{P})\boldsymbol{u}_i \xrightarrow{\mathrm{d}} \mathcal{N}(0,\Gamma_{ii})$ as desired.

Finally, we show that if the eigenvalues of $\mathbf{I}_{p,q}\Delta$ are all distinct then $(\hat{\lambda}_i - \lambda_i)_{i=1}^d \xrightarrow{\mathrm{d}} \mathrm{MVN}(\boldsymbol{\mu},\boldsymbol{\Gamma})$. More specifically, for any vector of constants $\boldsymbol{s} = (s_1, s_2, \ldots, s_d)$ in $\mathbb{R}^d$, we have

$$\begin{aligned}\sum_i s_i(\hat{\lambda}_i - \lambda_i) &= \sum_i \frac{s_i\lambda_i}{\hat{\lambda}_i}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i + \sum_i \frac{s_i\lambda_i}{\hat{\lambda}_i^2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i + o_{\mathbb{P}}(1)\\ &= \sum_i s_i\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i + \sum_i s_i\mu_i + o_{\mathbb{P}}(1)\\ &= \mathrm{tr}\Big((\mathbf{A}-\mathbf{P})\big(\sum_i s_i\boldsymbol{u}_i\boldsymbol{u}_i^\top\big)\Big) + \sum_i s_i\mu_i + o_{\mathbb{P}}(1).\end{aligned}$$

Now let $\mathbf{H} = \sum_i s_i\boldsymbol{u}_i\boldsymbol{u}_i^\top$. Then, conditional on $\mathbf{P}$, the matrix trace term $\mathrm{tr}\big((\mathbf{A}-\mathbf{P})\mathbf{H}\big)$ is once again a sum of independent mean zero random variables.

Letting h_{ij} denote the ij-th entry of $\mathbf{H}$, we have

$$\begin{aligned}\mathrm{Var}\Big(\mathrm{tr}\Big((\mathbf{A}-\mathbf{P})\mathbf{H}\Big)\Big) &= 2\sum_k\sum_l p_{kl}(1-p_{kl})h_{kl}^2 + o_{\mathbb{P}}(1)\\ &= 2\sum_k\sum_l p_{kl}(1-p_{kl})\big(\sum_i s_i u_{ik}u_{il}\big)^2 + o_{\mathbb{P}}(1)\\ &= 2\sum_i\sum_j s_is_j\sum_k\sum_l p_{kl}(1-p_{kl})u_{ik}u_{il}u_{jk}u_{jl}\\ &\quad + o_{\mathbb{P}}(1),\end{aligned}$$

which converges to $\sum_i\sum_j s_is_j\Gamma_{ij}$ as $n\to\infty$ for Γ_{ij} as defined in Eq. 2.3. Thus, for any vector $\boldsymbol{s}$ of fixed constants, $\sum_i s_i(\hat{\lambda}_i - \lambda_i) \xrightarrow{\mathrm{d}} \mathcal{N}(\boldsymbol{s}^\top\boldsymbol{\mu}, \boldsymbol{s}^\top\boldsymbol{\Gamma}\boldsymbol{s})$ and hence by the Cramer–Wold device, $(\hat{\lambda}_i - \lambda_i)_{i=1}^d \xrightarrow{\mathrm{d}} \mathrm{MVN}(\boldsymbol{\mu},\boldsymbol{\Gamma})$. The proof for the case where the eigenvalues of $\Delta\mathbf{I}_{p,q}$ are not distinct (see Theorem 3) follows the same argument. More specifically, we only need to show that for any vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_d)$,

$$\frac{1}{\boldsymbol{s}^\top\tilde{\boldsymbol{\Gamma}}\boldsymbol{s}}(\hat{\lambda}_1 - \lambda_1 - \tilde{\eta}_1, \ldots, \hat{\lambda}_d - \lambda_d - \tilde{\eta}_d) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

as $n\to\infty$. Here $\tilde{\boldsymbol{\Gamma}}$ is the matrix defined in Eq. 2.9. We omit the details.

REMARK 6. We now sketch a proof of Eq. 2.7. The expansion in Eq. 4.1 still holds when we disallow self-loops, i.e.,

$$\hat{\lambda}_i - \lambda_i = \frac{\lambda_i}{\hat{\lambda}_i}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i + \frac{\lambda_i}{\hat{\lambda}_i^2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i + O_{\mathbb{P}}(n^{-1}).$$

The quadratic form $\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i$ is now a function of the $n(n-1)/2$ independent random variables $\{a_{rs}-p_{rs}\}_{r<s}$; recall that if we allow self-loops then $\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i$ is a function of the $n(n+1)/2$ independent random variables $\{a_{rs}-p_{rs}\}_{r\le s}$. In either case, $\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i$ is concentrated around its expectation; see Theorem 5 and the sub-sequent discussion. In particular if there are no self-loops then letting ζ_{rs} denote the rsth entry of $\mathbb{E}[(\mathbf{A}-\mathbf{P})^2]$, we have

$$\zeta_{rs} = \sum_t \mathbb{E}[(a_{rt}-p_{rt})(a_{st}-p_{st})] = \begin{cases} 0 & \text{if } r \neq s, \\ \sum_{t\neq r} p_{rt}(1-p_{rt}) & \text{if } r = s. \end{cases}$$

Comparing the above expression with that in Eq. 4.18, we see that the only difference is in the indices of summation $t \neq r$ for the case when $r = s$. Plugging the above expression for ζ_{rs} into the derivations following Eq. 4.20, we still have

$$\frac{\lambda_i}{\hat{\lambda}_i^2}\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})^2\boldsymbol{u}_i \longrightarrow \eta_i.$$

We next consider the term $\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i$. We have

$$\boldsymbol{u}_i^\top(\mathbf{A}-\mathbf{P})\boldsymbol{u}_i = \sum_{r<s} 2u_{ir}u_{is}(a_{rs}-p_{rs}) + \sum_r u_{ir}^2(a_{rr}-p_{rr}).$$

When self-loops are allowed, the term $\sum_r u_{ir}^2(a_{rr}-p_{rr})$ is a sum of independent mean zero random variables and its contribution to the variance for $\hat{\lambda}_i - \lambda_i$ is negligible compared to the first term $\sum_{r<s} 2u_{ir}u_{is}(a_{rs}-p_{rs})$. However, when there are no self-loops, the term $\sum_r u_{ir}^2(a_{rr}-p_{rr}) = -\sum_r u_{ir}^2 p_{rr}$ is no longer a sum of mean zero random variables; this term now contributes to the bias in $\hat{\lambda}_i - \lambda_i$. Now recall Eq. 4.20; in particular $u_{ir} = X_r^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i$. We then have

$$\begin{aligned}
\sum_r u_{ir}^2 p_{rr} &= \sum_r \boldsymbol{v}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}X_rX_r^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_iX_r^\top\mathbf{I}_{p,q}X_r \\
&= \boldsymbol{v}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1/2}\Big(\sum_r X_rX_r^\top X_r^\top\mathbf{I}_{p,q}X_r\Big)(\mathbf{X}^\top\mathbf{X})^{-1/2}\boldsymbol{v}_i \\
&= \boldsymbol{v}_i^\top\big(\tfrac{\mathbf{X}^\top\mathbf{X}}{n}\big)^{-1/2}\Big(\frac{1}{n}\sum_r X_rX_r^\top X_r^\top\mathbf{I}_{p,q}X_r\Big)\big(\tfrac{\mathbf{X}^\top\mathbf{X}}{n}\big)^{-1/2}\boldsymbol{v}_i
\end{aligned}$$

$$\begin{aligned} &\to \quad \boldsymbol{\xi}_i^\top \Delta^{-1/2} \mathbb{E}\left[\left(X^\top \mathbf{I}_{p,q} X\right) X X^\top\right] \Delta^{-1/2} \boldsymbol{\xi}_i \\ &\equiv \quad \theta_i \end{aligned}$$

as desired.

Funding. The research leading to these results received funding from the Defense Advanced Research Projects Agency under Grant Agreement No. FA8750-12-2-0303.

Compliance with Ethical Standards.

The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* **9**, 1981–2014.

ARNOLD, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *Journal of Mathematical Analysis and Applications, 20*, 262–268.

AVRACHENKOV, K., COTTATELLUCCI, L. and KADAVANKANDY, A. (2015). Spectral properties of random matrices for stochastic block model. In *Proceedings of the* $4^{\text{th,}}$ *International Workshop on Physics-Inspired Paradigms in Wireless Communications and Networks* (pp. 537–544).

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics* **227**, 494–521.

BORDENAVE, C. and CAPITAINE, M. (2016). Outlier eigenvalues for deformed i.i.d. random matrices. *Communications on Pure and Applied Mathematics* **69**, 2131–2194.

BOUCHERON, S., LUGOSI, G. and MASSART, P. (2003). Concentration inequalities using the entropy method. *Annals of Probability* **31**, 1583–1614.

CAPE, J., TANG, M. and PRIEBE, C. E. (2017). The Kato-Temple inequality and eigenvalue concentration. *Electronic Journal of Statistics* **11**, 3954–3978.

CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37**, 1–47.

CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2012). Central limit theorems for eigenvalues of deformations of Wigner matrices. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* **48**, 107–133.

CHAKRABARTY, A., CHAKRABORTY, S. and HAZRA, R. S. (2020). Eigenvalues outside the bulk of inhomogeneous Erdos-Renyi random graphs. *Journal of Statistical Physics* **181**, 1746–1780.

DAVIS, C. and KAHAN, W. (1970). The rotation of eigenvectors by a pertubation. III. *SIAM Journal on Numerical Analysis,* **7**, 1–46.

DING, X. and JIANG, T. (2010). Spectral distributions of adjacency and Laplacian matrices of random graphs. *Annals of Applied Probability* **20**, 2086–2117.

DONOHO, D., GAVISH, M. and JOHNSTONE, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model. *Annals of Statistics* **46**, 1742–1778.

ERDÖS, L., PÉCHÉ, S., RAMIREZ, J. A., SCHLEIN, B. and YAU, H.-T. (2010). Bulk universality for W,igner matrices. *Communications on Pure and Applied Mathematics* **63**, 895–925.

ERDÖS, L., SCHLEIN, B. and YAU, H. -T. (2009). Local semicircle law and complete delocalization for Wigner random matrices. *Communications in Mathematical Physics* **287**, 641–655.

FÜREDI, Z. and KOMLÓS, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1**, 233–241.

GHOSHDASTIDAR, D., GUTZEIT, M., CARPENTIER, A. and VON LUXBURG, U. (2017). Two-sample tests for large random graphs using network statistics. In *Proceedings of Machine Learning Research* (vol. 65, pp. 1–24).

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: first steps. *Social Networks* **5**, 109–137.

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

KNOWLES, A. and YIN, J. (2013). Eigenvector distribution of Wigner matrices. *Probability Theory and Related Fields* **155**, 543–582.

KNOWLES, A. and YIN, J. (2014). The outliers of a deformed Wigner matrix. *Annals of Probability* **42**, 1980–2031.

KOLTCHINSKII, V. and GINÉ, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* **6**, 113–167.

LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Annals of Statistics* **44**, 401–424.

NOROOZI, M., RIMAL, R. and PENSKY, M. (2019). Estimation and clustering in popularity adjusted stochastic blockmodel. arXiv:1902.00431

O'ROURKE, S. and RENFREW, D. (2014). Low rank perturbation of large elliptic random matrices. *Electronic Journal of Probability* **19**, 1–65.

O'ROURKE, S. and VU, V. (2018). K. Wang. Random perturbation of low rank matrices Improving classical bounds. *Linear Algebra and its Applications* **540**, 26–59.

PÉCHÉ, S. (2006). The largest eigenvalue of small-rank perturbations of Hermitean random matrices. *Probability Theory and Related Fields* **134**, 127–173.

PIZZO, A., RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformation of Wigner matrices. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques* **49**, 64–94.

RUBIN-DELANCHY, P., CAPE, J., PRIEBE, C. E. and TANG, M. (2017). A statistical interpretation of spectral embedding The random dot product graph. arXiv:1709.05506

SENGUPTA, S. and CHEN, Y. (2018). A blockmodel for node popularity in networks with community structure. *Journal of the Royal Statistical Society Series B*, 365–386.

SOSHNIKOV, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Communications in Mathematical Physics* **207**, 697–733.

TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V., PARK, Y. and PRIEBE, C. E. (2017). A semiparametric two-sample hypothesis testing problem for random dot product graphs. *Journal of Computational and Graphical Statistics* **26**, 344–354.

TAO, T. and VU, V. (2010). Random matrices: universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics* **298**, 549–572.

TAO, T. and VU, V. (2012). Random matrices: universal properties of eigenvectors. *Random Matrices: Theory and Applications*, 1.

TOKUDA, T. (2018). Statistical test for detecting community structure in real-valued edge-weighted graphs. *PLoS ONE* **13**, e0194079.

WIGNER, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics* **62**, 548–564.

YOUNG, S. and SCHEINERMAN, E. (2007). Random dot product graph models for social networks. In *Proceedings of the 5th international conference on algorithms and models for the web-graph* (pp. 138–149).

YU, Y., WANG, T. and SAMWORTH, R. J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika* **102**, 315–323.

ZHANG, X., NADAKUDITI, R. R. and NEWMAN, M. E. (2014). Spectra of random graphs with community structure and arbitrary degrees. *Physical Review E*, 89.

Avanti Athreya
Applied Mathematics and Statistics,
Johns Hopkins University, Baltimore,
MD, USA
E-mail: dathrey1@jhu.edu

Joshua Cape
Department of Statistics, University
of Pittsburgh, Pittsburgh, PA, USA
E-mail: joshua.cape@pitt.edu

Minh Tang
Department of Statistics, North
Carolina State University, Raleigh,
NC, USA
E-mail: mtang8@ncsu.edu

Paper received: 1 June 2020; accepted 17 September 2021.