The Kato–Temple inequality and eigenvalue concentration with applications to graph inference[*]

Joshua Cape, Minh Tang, and Carey E. Priebe

Department of Applied Mathematics and Statistics
Johns Hopkins University
3400 N. Charles St., Baltimore, MD 21218, USA
e-mail: joshua.cape@jhu.edu*;* mtang10@jhu.edu*;* cep@jhu.edu

Abstract: We present an adaptation of the Kato–Temple inequality for bounding perturbations of eigenvalues with applications to statistical inference for random graphs, specifically hypothesis testing and change-point detection. We obtain explicit high-probability bounds for the individual distances between certain signal eigenvalues of a graph's adjacency matrix and the corresponding eigenvalues of the model's edge probability matrix, even when the latter eigenvalues have multiplicity. Our results extend more broadly to the perturbation of singular values in the presence of quite general random matrix noise.

1. Introduction

1.1. Overview

Eigenvalues and eigenvectors are structurally fundamental quantities associated with matrices and are widely studied throughout mathematics, statistics, and engineering disciplines. For example, given an observed graph, the eigenvalues and eigenvectors of associated matrix representations (such as the adjacency matrix or Laplacian matrix) encode structural information about the graph (e.g. community structure, connectivity [8]). In the context of certain random graph models, the eigenvalues and eigenvectors associated with the underlying matrix-valued model parameter, the edge probability matrix, exhibit similar

information. It is therefore natural to study how "close" the eigenvalues and eigenvectors of a graph are to the underlying model quantities.

In this paper we consider simple, undirected random graphs on n vertices generated via the *inhomogeneous Erdős–Rényi model* (IERM) [7, 13], $\mathbb{G}(n, P)$, where $P := [P_{i,j}] \in [0,1]^{n \times n}$ denotes the (symmetric) edge probability matrix. This independent edge model generalizes numerous widely-studied random graph models including the classical Erdős–Rényi model [10], the stochastic block model [14], and the random dot product graph model [40]. For $G \sim \mathbb{G}(n, P)$, the (symmetric) adjacency matrix, $A \equiv A_G \in \{0,1\}^{n \times n}$, has entries that are independently distributed according to $A_{i,j} \sim$ Bernoulli$(P_{i,j})$ for all $i \leq j$. This yields $P \equiv \mathbb{E}[A]$, where $\mathbb{E}[\cdot]$ denotes probabilistic expectation.

We focus our attention on the eigenvalues of A and P. Specifically, we consider the eigenvalues in pairs (e.g. the largest eigenvalues of A and of P form a pair, as do the second-largest eigenvalues of each matrix, etc.). We obtain bounds on the distance between eigenvalues in certain "signal pairs", thereby and therein demonstrating a local sense in which random graphs concentrate. Note that in the random graph literature, the term *concentration* is primarily used to describe global, uniform behavior via the spectral norm quantity $\|A - \mathbb{E}[A]\|_2$.

The following description provides an overview of our results for the IERM setting. Given a collection of consecutive, ordered eigenvalues of P which are sufficiently separated from the remainder of the spectrum and conditional on that the corresponding eigenvalues of A are not near the remainder of the spectrum of either A or P, then Theorems 3.3 and 3.7 yield high-probability bounds on the distances between the eigenvalues in each pair. The individual, pair-specific (i.e. local) bounds we obtain stand in contrast to weaker bounds which hold uniformly for all eigenvalue pairs (for example, bounds implied by Weyl's inequality [15]). Our results hold even in the presence of eigenvalue multiplicity.

We demonstrate that when the matrix P has low rank, our results compare favorably with the recent study of low rank matrices undergoing random perturbation in [28] (see our Example 3.14). We also demonstrate that our results can lead to meaningful estimation in high rank settings (see Example 3.15).

After presenting our main theoretical results, we then apply the theory in this paper to both hypothesis testing and change-point detection for random graphs. Moreover, we generalize our results beyond the IERM setting to obtain high-probability bounds for perturbations of singular values of rectangular matrices in a quite general random matrix noise setting.

Broadly speaking, we adapt the original, deterministic setting in a paper by T. Kato [18] to a new setting involving randomness, and this approach is novel in the context of random graphs, random matrix theory, and statistical inference for random graphs. We further detail the key modifications and differences between our work and [18] in our subsequent remarks and proofs. The present paper also stands in contrast to a deterministic generalization of the Kato–Temple inequality in [16].

1.2. Inhomogeneous random graphs

In the inhomogeneous random graph literature, concentration bounds have been known for some time for each eigenvalue of A, denoted $\lambda_i(A)$, both around its median and around its expectation, $\mathbb{E}[\lambda_i(A)]$ [1]. Unfortunately, since the latter quantities are inaccessible in practice, such bounds are of limited practical use. Moreover, in general $\mathbb{E}[\lambda_i(A)] \neq \lambda_i(\mathbb{E}[A])$.

By way of contrast, numerous results in the literature bound the spectral norm matrix difference $\|A - \mathbb{E}[A]\|_2$, thereby immediately and uniformly bounding each of the eigenvalue differences $|\lambda_i(A) - \lambda_i(\mathbb{E}[A])|$ via an application of Weyl's inequality. For example, [27] proved an asymptotically almost surely spectral norm bound of $\|A - \mathbb{E}[A]\|_2 = O(\sqrt{\Delta \log n})$ for $\Delta = \Omega(\log(n))$ where $\Delta \equiv \Delta(n)$ denotes the maximum expected degree of a graph. In [23] the above bound is improved to $\|A - \mathbb{E}[A]\|_2 \leq (2 + o(1))\sqrt{\Delta}$ under the stronger assumption that $\Delta = \omega(\log^4 n)$ with further refinement being subsequently obtained in [22]. We on the other hand show that under certain conditions, for particular eigenvalue pairs one can obtain tighter and non-uniform high probability bounds of the form $|\lambda_i(A) - \lambda_i(\mathbb{E}[A])| = O(\log^\delta n)$ for small $\delta > 0$.

Spectral theory for random graphs overlaps with the random matrix theory literature. There, asymptotic analysis includes proving, for example, convergence of the empirical spectral distribution to a limiting measure [9]. Related approaches to studying the spectrum of random graphs consider normalized versions of the adjacency matrix [20] and employ standard random matrix theory techniques such as the Stieltjes transform method [4, 41]. In contrast, we do not study normalized versions of the adjacency or the edge probability matrix.

Indeed, much of the existing literature focuses on properties of eigenvectors corresponding to random graphs [11, 22, 35] given, among other reasons, the success of spectral clustering methods for graph inference [38]. We do not consider eigenvectors since our aim is to demonstrate the usefulness of adapting and applying the eigenvalue-centric Kato–Temple framework.

The stochastic block model (SBM) offers an example of an inhomogeneous random graph model which is widely (and wildly) popular in the literature [6, 17, 21, 22, 42] and in which our results apply to the top (signal) eigenvalues of A and P. Previously, the authors in [3] obtained a collective deviation bound on the top eigenvalues of A and P for certain stochastic block model graphs in order to prove the main limit theorem therein. Our Theorem 3.3 improves upon Lemma 2 in [3] by removing a distinct eigenvalue assumption and by yielding stronger high-probability deviation bounds for pairs of top eigenvalues of A and P which are of the same order. This implies a statistical hypothesis testing regime for random graphs which is discussed further in Section 4.

1.3. Organization

The remainder of this paper is organized as follows. In Section 2 we introduce notation and the Kato–Temple eigenvalue perturbation framework. In Section 3

we present our results for random graphs and more generally for matrix perturbation theory. There we also include illustrative examples together with comparative analysis involving recent results in the literature. In Section 4 we discuss applications of our results to problems involving graph inference. Sections 5 and 6 contain our acknowledgments and the proofs of our results, respectively.

2. Setup and notation

Let $\langle \cdot, \cdot \rangle$ denote the standard Euclidean inner (dot) product between two vectors, $\| \cdot \|$ denote the vector norm induced by the dot product, and $\| \cdot \|_2$ denote the spectral norm of a matrix. The identity matrix is implicitly understood when we write the difference of a matrix with a scalar. In this paper, $\mathcal{O}(\cdot)$, $\Omega(\cdot)$, and $\Theta(\cdot)$ denote standard big-O, big-Omega, and big-Theta notation, respectively, while $o(\cdot)$ and $\omega(\cdot)$ denote standard little-o and little-omega notation, respectively.

As prefaced in Section 1, we consider simple, undirected random graphs on n vertices generated by the inhomogeneous Erdős–Rényi model, $G \sim \mathbb{G}(n, P)$, via the corresponding (binary, symmetric) adjacency matrix $A \equiv A_G$. Given an open interval in the positive half of the real line, $(\alpha, \beta) \subset \mathbb{R}_{>0}$, we denote the d eigenvalues of P that lie in this interval (locally) by

$$\alpha < \lambda_1(P) \leq \lambda_2(P) \leq \cdots \leq \lambda_d(P) < \beta, \tag{2.1}$$

and similarly for A, noting that for A this amounts to a probabilistic statement. By symmetry one can just as well handle the case when the interval lies in the negative half of the real line. We are principally interested in eigenvalues that are large in magnitude, so we do not consider the case when the underlying interval contains the origin.

To highlight the Kato–Temple framework for bounding eigenvalues, we now reproduce two lemmas from [18] along with the Kato–Temple inequality as stated in [16] (see Theorem 2.3 below).[1] These results all hold in the following common setting.

Let H be a self-adjoint operator on a Hilbert space. Assume a unit vector w is in the domain of H and define $\eta := \langle Hw, w \rangle$ along with $\epsilon := \|(H - \eta)w\|$, noting that $\eta^2 + \epsilon^2 = \|Hw\|^2$. The quantity η may be viewed as an "approximate eigenvalue" of H corresponding to the "approximate eigenvector" w, while ϵ represents a scalar residual term.

Lemma 2.1 ([18], Lemma 1). *For every α such that $\alpha < \eta$ (where $\alpha = -\infty$ is permitted), the interval $(\alpha, \eta + \frac{\epsilon^2}{\eta - \alpha}]$ contains a point in the spectrum of H.*

Lemma 2.2 ([18], Lemma 2). *For every β such that $\beta > \eta$ (where $\beta = \infty$ is permitted), the interval $[\eta - \frac{\epsilon^2}{\beta - \eta}, \beta)$ contains a point in the spectrum of H.*

---

[1]Of primary importance in this paper is the extension of Theorem 2.3 to multiple eigenvalues as presented in [18]. The original statement of the extension to multiple eigenvalues is more involved and therefore omitted for simplicity.

Theorem 2.3 (Kato–Temple inequality; [16], Theorem 2). *Suppose that $\epsilon^2 < (\beta - \eta)(\eta - \alpha)$ where $\alpha < \beta$. Then* $\mathrm{spectrum}(H) \cap (\alpha, \beta) \neq \emptyset$. *Moreover, if the only point of the spectrum of H in the interval (α, β) is the eigenvalue $\lambda(H)$, then*

$$-\frac{\epsilon^2}{\beta - \eta} \leq \lambda(H) - \eta \leq \frac{\epsilon^2}{\eta - \alpha}.$$

Remark 2.4 (Hermitian dilation). Given an $m \times n$ real matrix M, it will be useful to consider the corresponding real symmetric $(m+n) \times (m+n)$ Hermitian dilation matrix $\tilde{M}$ given by

$$\tilde{M} := \begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}.$$

It is well-known that the non-zero eigenvalues of $\tilde{M}$ correspond to the signed singular values of M (see Theorem 7.3.3 in [15]). This correspondence between the singular values of arbitrary matrices and the eigenvalues of Hermitian matrices allows our results to generalize beyond the IERM setting to the more general study of matrix perturbation theory for singular values in a straightforward manner.

3. Results

3.1. Results for random graphs

In the IERM setting, a graph's adjacency matrix can be written as $A = P + E$ where $E := A - P$ is a random matrix and P is the (deterministic) expectation of A. We begin with a preliminary observation concerning the tail behavior of $A - P$ which will subsequently be invoked for the purpose of obtaining standard union bounds. The proof follows from a straightforward application of Hoeffding's inequality.

Proposition 3.1 (General IERM concentration). *Let $u, v \in \mathbb{R}^n$ denote (non-random) unit vectors. Then for any $t > 0$,*

$$\mathbb{P}[|\langle (A - P)u, v \rangle| > t] \leq 2 \exp(-t^2). \tag{3.1}$$

It is indeed possible to invoke more refined concentration inequalities than Proposition 3.1 in the presence of additional structure (e.g. when all entries of P have uniformly very small magnitude). Doing so is particularly useful when it is simultaneously possible to obtain a strong bound on $\|A - P\|_2$. This observation will be made clearer in the context of Theorem 3.3 below. Furthermore, consideration of Proposition 3.1 will facilitate the subsequent presentation of our generalized results which extend beyond the IERM setting.

Remark 3.2. In this paper the main diagonal elements of P are allowed to be strictly positive, in which case realizations of A need not necessarily be hollow (i.e. observed graphs may have self-loops). To avoid graphs with self-loops, one

may either condition on the event that A is hollow or set the main diagonal of P to be zero. In the former case, note that $P \equiv \mathbb{E}[A]$ no longer holds on the main diagonal. In the latter case, a modified version of Proposition 3.1 holds.

We now present our main results for the IERM setting. The proofs, which are located in Section 6, also formulate a bound for the special case when the upper bound threshold β may be chosen to be infinity. This special case is particularly useful in applications.

Theorem 3.3 (IERM eigenvalue perturbation bounds, conditional version)**.** *Let the matrices $A \in \{0,1\}^{n \times n}$ and $P \in [0,1]^{n \times n}$ correspond to the IERM setting described in Section 2. Suppose the interval $(\alpha, \beta) \subset \mathbb{R}_{>0}$ contains precisely d eigenvalues of P, $\lambda_1(P) \leq \lambda_2(P) \leq \cdots \leq \lambda_d(P)$ (possibly with multiplicity). Condition on the event that (α, β) contains precisely d eigenvalues of A, $\{\lambda_i(A)\}_{i=1}^d$, as well as the set $\{\langle Aw_i, w_i \rangle\}_{i=1}^d$ where $\{w_i\}_{i=1}^d$ is an orthonormal collection of eigenvectors of P corresponding to the eigenvalues $\{\lambda_i(P)\}_{i=1}^d$. Fix $k \in [d]$. Define $l := (d - k + 1)$. Then, for $t > 0$,*

$$\lambda_k(A) \geq \lambda_k(P) - t - \zeta^-, \tag{3.2}$$

where $\zeta^- := \frac{l\|E\|_2^2 + ((\beta - \lambda_k(P)) + (\lambda_d(P) - \lambda_k(P)) + 3t)l(l-1)t}{\beta - \lambda_d(P) - (l(l-1)+1)t}$ with probability at least $1 - \left(l + \binom{l}{2}\right) 2 \exp(-t^2)$. Also, for $t > 0$,

$$\lambda_k(A) \leq \lambda_k(P) + t + \zeta^+, \tag{3.3}$$

where $\zeta^+ := \frac{k\|E\|_2^2 + (3\lambda_k(P) - \alpha + 3t)k(k-1)t}{\lambda_1(P) - \alpha - (k(k-1)+1)t}$ with probability at least $1 - \left(k + \binom{k}{2}\right) 2 \exp(-t^2)$. Moreover, the upper and lower bounds hold collectively with probability at least $1 - \left(d + \binom{d}{2}\right) 2 \exp(-t^2)$.

Remark 3.4. Our proof depends upon several new observations with respect to Kato's original argument. In particular, for w_i as defined above, the matrix $[\langle Aw_i, w_j \rangle]_{i,j=1}^d$ need not be diagonal, so $\{w_i\}_{i=1}^d$ need not constitute an orthonormal collection of "approximate eigenvectors" of A in the sense of [18]. Instead, here the notion of "approximate" may be interpreted via Proposition 3.1 as the source of randomness which allows for Kato–Temple methodology to be adapted beyond the original deterministic setting. Of additional note is that the vectors w_i as defined in this paper agree in function and notation with Kato's original paper, the operational distinction being that our setting provides a canonical choice for these vectors.

Remark 3.5. We note that the term $\|E\|_2^2$ in the formulation of both ζ^+ and ζ^- can be replaced by an appropriate maximum over quantities of the form $\|Ew_i\|^2$ (see Equation (6.19)). That is to say, in the presence of additional local structure and knowledge, one can refine the above bounds in Theorem 3.3.

Remark 3.6. In settings wherein the eigenvalues of interest have disparate orders of magnitude, Kato–Temple methodology is not guaranteed to yield useful

bounds. This can be seen in the bounds' dependence on the ratio of the eigenvalues of P in Theorem 3.3. Moreover, within the Kato–Temple framework, poor separation from the remainder of the spectrum also deteriorates the bounds, as is evident in the denominators' dependence on the interval endpoints α and β along with the smallest and largest local eigenvalues of P. On the other hand, by further localizing, i.e. by restricting to a subset of $d' < d$ eigenvalues in a particular interval, applying Theorem 3.3 to said fewer eigenvalue pairs may yield improved bounds (see Example 3.14 and Remark 3.5).

Next, we formulate an unconditional version of Theorem 3.3. For both simplicity and the purpose of applications, Theorem 3.7 is stated in terms of the largest singular values in the IERM setting.

Theorem 3.7 (IERM singular value perturbation bounds, unconditional version). *Let the matrices $A \in \{0,1\}^{n \times n}$ and $P \in [0,1]^{n \times n}$ correspond to the IERM setting described in Section 2 with maximum expected degree (via P) given by $\Delta \equiv \Delta(n)$. Denote the $d+1$ largest singular values of A by $0 \leq \hat{\sigma}_0 < \hat{\sigma}_1 \leq \cdots \leq \hat{\sigma}_d$, and denote the $d+1$ largest singular values of P by $0 \leq \sigma_0 < \sigma_1 \leq \cdots \leq \sigma_d$. Suppose that $\Delta = \omega(\log^4 n)$, $\sigma_1 \geq C\Delta$, and $\sigma_0 \leq c\Delta$ for some absolute constants $C > c > 0$. Let $\delta \in (0,1]$. Then for each $k \in [d]$, there exists some positive constant $c_{k,d}$ such that as $n \to \infty$, with probability $1 - o(1)$ involving δ,*

$$|\hat{\sigma}_k - \sigma_k| \leq c_{k,d} \left(\log^\delta n \right). \tag{3.4}$$

A similar version of Theorem 3.12 holds when $\Delta = \Omega(\log n)$ under slightly different assumptions on the entries of P for which one still has $\|A - P\|_2 = O(\sqrt{\Delta})$ with high probability [22]. On a related yet different note, see [20] for discussion of the sparsity regime $\Delta = O(1)$ in which graphs fail to concentrate in the classical sense.

Remark 3.8 (Random dot product graph model). When the edge probability matrix P can be written as $P = XX^\top$ for some matrix $X \in \mathbb{R}^{n \times d}$ with $d \ll n$, then the IERM corresponds to the popular *random dot product graph (RDPG) model* [40]. In the random dot product graph model, the largest eigenvalues of A and P are of statistical interest in that they represent spectral "signal" in the model. These eigenvalues are separated from the remainder of their respective spectra and lie in an interval of the form (α, ∞), where α may be taken on the order of $\|A - P\|_2$.

Among its applications, the RDPG model has been used as a platform for modeling graphs with hierarchical and community structure [24]. A central limit theorem is known for the behavior of the top eigenvectors of adjacency matrices arising from the RDPG model [3], and this limit theorem relies upon a lemma which collectively bounds the differences between top eigenvalues of A and P but requires a stringent eigengap assumption. Namely, Lemma 2 in [3] states that with high probability,

$$\sqrt{\sum_{i=1}^{d} |\lambda_i(A) - \lambda_i(P)|^2} = O(\delta_{\text{gap}}^{-2} \log n). \tag{3.5}$$

In contrast, using Theorem 3.7 with $\sigma_0 := 0$, we do not require the gap assumption $\delta_{\mathrm{gap}} > 0$ and still obtain that with high probability,

$$\sqrt{\sum_{i=1}^{d} |\lambda_i(A) - \lambda_i(P)|^2} = O(\log n). \tag{3.6}$$

In practice, models involving repeated or arbitrarily close eigenvalues are prevalent and of interest (e.g. Section 4.2). As such, the above improvement is nontrivial and of practical significance.

Remark 3.9 (Latent position random graphs). Theorem 3.3 further extends to the more general setting of latent position random graphs. There, the matrix P is viewed as an operator $[\kappa(X_i, X_j)]_{i,j=1}^{n}$ where X_i and X_j are independent, identically distributed latent positions with distribution F, and the positive definite kernel, κ (viewed as an integral operator), is not necessarily of finite fixed rank as n increases [13, 36]. Note that for the RDPG model, the kernel κ is simply the standard Euclidean inner product between (latent position) vectors.

3.2. Results for matrix perturbation theory

The behavior of the random matrix $A - P$ (see Proposition 3.1) represents a specific instance of more general, widely-encountered probabilistic concentration as discussed in [28] and formulated in the following definition.

Definition 3.10 ([28]). An $m \times n$ random real ("error") matrix E is said to be (C, c, γ)-*concentrated* for a trio of positive constants $C, c, \gamma > 0$ if for all unit vectors $u \in \mathbb{R}^n, v \in \mathbb{R}^m$ and for every $t > 0$, then

$$\mathbb{P}[|\langle Eu, v \rangle| > t] \le C \exp(-ct^\gamma). \tag{3.7}$$

In particular, the IERM setting corresponds to (C, c, γ) concentration where $m = n$, $C = \gamma = 2$, and $c = 1$. For the Hermitian dilation discussed in Remark 2.4 one has the following correspondence between E and $\tilde{E}$.

Lemma 3.11 ([28]). *Let $E \in \mathbb{R}^{m \times n}$ be (C, c, γ)-concentrated. Define $\tilde{C} := 2C$ and $\tilde{c} := c/2^\gamma$. Then the matrix $\tilde{E} \in \mathbb{R}^{m+n \times m+n}$ is $(\tilde{C}, \tilde{c}, \gamma)$-concentrated.*

Definition 3.10 and Lemma 3.11 together with Remark 2.4 allow for Theorem 3.3 to be generalized in a straightforward manner. We frame the generalization in the context of a signal–plus–noise matrix model with tail probability bounds. In particular, replace A with $\hat{M} := M + E$, thought of as an observed data matrix. Also replace P with M, thought of as an underlying signal matrix, so that the matrix $A - P$ becomes E, thought of as an additive error matrix. We emphasize that the following generalization is in terms of the singular values of M and $\hat{M}$. This generalization resembles the formulation of a result obtained in [28] using different methods; however, unlike our Theorem 3.12, the bound in [28] depends upon the rank of M and assumes that the rank is known.

Given a matrix $M \in \mathbb{R}^{m \times n}$, write its singular value decomposition as $M \equiv U \Sigma V^\top$ where $M v_i = \sigma_i u_i$ holds for the normalized right (resp., left) singular vectors v_i (resp., u_i) and singular values $\sigma_i = \Sigma_{i,i}$. For each i such that $\sigma_i > 0$, define $\tilde{w}_i \in \mathbb{R}^{m+n}$ to be the concatenated unit vector $w_i := \frac{1}{\sqrt{2}}(u_i^\top, v_i^\top)^\top$. Note that $\tilde{w}_i$ is an eigenvector for $\tilde{M}$ with $\tilde{M}\tilde{w}_i = \sigma_i \tilde{w}_i$.

Theorem 3.12 (Singular value perturbation bounds, conditional version). *For matrices $M, E \in \mathbb{R}^{m \times n}$ and $\hat{M} := M + E$, suppose that E is (C, c, γ)-concentrated for positive constants $C, c, \gamma > 0$. Suppose the interval $(\alpha, \beta) \subset \mathbb{R}_{>0}$ contains the largest d singular values of M, denoted by $0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_d$. Condition on the event that the interval (α, β) contains precisely d singular values of $\hat{M}$, denoted $0 < \hat{\sigma}_1 \leq \hat{\sigma}_2 \leq \cdots \leq \hat{\sigma}_d$, as well as $\langle \tilde{\hat{M}}\tilde{w}_i, \tilde{w}_i \rangle$ for $1 \leq i \leq d$ and unit vector $\tilde{w}_i$ as defined above. Fix $k \in [d]$. Define $l := (d - k + 1)$. Then for $t > 0$,*

$$\hat{\sigma}_k \geq \sigma_k - t - \zeta^-, \tag{3.8}$$

where $\zeta^- := \frac{l\|E\|_2^2 + ((\beta - \sigma_k) + (\sigma_d - \sigma_k) + 3t)l(l-1)t}{\beta - \sigma_d - (l(l-1)+1)t}$ with probability at least $1 - \left(l + \binom{l}{2}\right)\tilde{C}\exp(-\tilde{c}t^\gamma)$. Also, for $t > 0$,

$$\hat{\sigma}_k \leq \sigma_k + t + \zeta^+, \tag{3.9}$$

where $\zeta^+ := \frac{k\|E\|_2^2 + (3\sigma_k - \alpha + 3t)k(k-1)t}{\sigma_1 - \alpha - (k(k-1)+1)t}$ with probability at least $1 - \left(k + \binom{k}{2}\right)\tilde{C}\exp(-\tilde{c}t^\gamma)$. Moreover, the upper and lower bound hold collectively with probability at least $1 - \left(d + \binom{d}{2}\right)\tilde{C}\exp(-\tilde{c}t^\gamma)$.

As with the results in Section 3.1, Theorem 3.12 can be formulated unconditionally and for collections of not–necessarily–the–largest singular values. Both of these aspects are explored in greater detail in Example 3.15. The following technical lemma will subsequently be employed in applications.

Lemma 3.13. *Let $E \in \mathbb{R}^{m \times n}$ be a (C, c, γ)-concentrated random matrix. Choose $\epsilon > 0$ such that $2 + \epsilon > 2\left(2\log(9)/c\right)^{1/\gamma}$ and define the quantity $c_{\epsilon,c,\gamma} := (c(1 + \epsilon/2)^\gamma - 2\log(9)) > 0$. Then,*

$$\mathbb{P}\left[\|E\|_2 > (2 + \epsilon)\max\{m, n\}^{1/\gamma}\right] \leq C\exp(-c_{\epsilon,c,\gamma}\max\{m, n\}). \tag{3.10}$$

If in addition $m = n$ and E is assumed to be symmetric, then the quantity $2\log(9)$ above may be replaced by $\log(9)$, an improvement.

3.3. Two illustrative examples

In the remainder of this section, we present two examples that highlight the usefulness and flexibility of Kato–Temple methodology. We begin with Example 3.14 which presents a simple stochastic block model setting wherein our

results compare favorably with those in the recent work of [28], noting that in general for similar settings, the corresponding results are often comparable.

Example 3.14 (Balanced affinity two block stochastic block model). Consider an n vertex realization from a two block stochastic block model in which $0 < q < p < 1$ where p and q denote the within–block and between–block edge probabilities, respectively. Suppose each block contains $n/2$ of the graph's vertices. The signal singular values and maximum expected degree of this rank two model are given by

$$\sigma_1(P) = \frac{n}{2}(p - q), \sigma_2(P) = \frac{n}{2}(p + q), \text{ and } \Delta = \sigma_2(P). \qquad (3.11)$$

For the purposes of large n comparison, view $\|E\|_2 \approx 2\sqrt{\Delta}$ from [23] and set the lower threshold α to be $\|E\|_2$. Define $r_{p,q}$ to be the edge probability-dependent parameter $r_{p,q} := (p+q)/(p-q)$. Then via Kato–Temple methodology applied jointly to $\sigma_1(P)$ and $\sigma_2(P)$, with probability approximately 0.99 when $t_{KT} \geq 2.55$, for each singular value, respectively,

$$-3t_{KT} \leq \hat{\sigma}_1(A) - \sigma_1(P) \leq 4r_{p,q} + t_{KT},$$
$$-t_{KT} \leq \hat{\sigma}_2(A) - \sigma_2(P) \leq (8 + 6t)r_{p,q} + t_{KT}.$$

By the same approach, the bounds obtained in [28] are given by

$$-t_{OVW} \leq \hat{\sigma}_1(A) - \sigma_1(P) \leq 8\sqrt{2}r_{p,q} + \sqrt{2}t_{OVW},$$
$$-t_{OVW} \leq \hat{\sigma}_2(A) - \sigma_2(P) \leq 8 + \sqrt{2}t_{OVW}.$$

Direct application of the results in [28] yields probability approximately at least 0.99 for $t_{OVW} \geq 11.6$, though it appears upon further inspection that this can be improved to, for example, $t_{OVW} \geq 5.6$. The above joint analysis demonstrates that our bounds are favorable for the pair $\{\hat{\sigma}_1(A), \sigma_1(P)\}$ whereas the opposite is true for the pair $\{\hat{\sigma}_2(A), \sigma_2(P)\}$.

We emphasize that here the upper bounds are of primary importance and interest. Indeed, the (C, c, γ) property allows for straightforward lower bounds to be obtained by epsilon net techniques together with the Courant–Fisher–Weyl min-max principle. For example, note that a single application of (C, c, γ)-concentration yields that $\hat{\sigma}_2(A) - \sigma_2(P) \geq -t$ with probability at least $1 - C\exp(-ct^\gamma)$.

Among the advantages of Kato–Temple methodology is the ability to, in certain cases, refine one's initial analysis by further localizing the underlying interval (α, β). This is possible in the current example wherein we can "zoom in" further on the largest signal singular value. In particular, keeping the same indexing as above and setting α to be $\|E\|_2 + \sigma_1(P)$, then for n large and with probability approximately 0.99, we have

$$-t_{KT} \leq \hat{\sigma}_2(A) - \sigma_2(P) \leq 2\left(\tfrac{p}{q} + 1\right) + t_{KT}. \qquad \blacktriangle$$

In contrast to the low rank setting of Example 3.14, Example 3.15 below demonstrates how our results can be applied to the problem of estimating signal in a high rank matrix setting.

Example 3.15 (Estimating signal in a high rank spike model)**.** Let $m, n, p \in \mathbb{N}$ and set $q := m + n + p$. Let $M \in \mathbb{R}^{q \times q}$ be full rank with singular values given by the set

$$\{\underbrace{1, \ldots, 1}_{m \text{ times}}, \underbrace{\kappa + 1, \ldots, \kappa + 1}_{n \text{ times}}, \underbrace{\tau + \kappa + 1, \ldots, \tau + \kappa + 1}_{p \text{ times}}\},$$

where $\tau, \kappa > 0$. By slight abuse of notation, denote the singular values of M up to multiplicity by $\sigma_1 := 1$, $\sigma_2 := \kappa + 1$, and $\sigma_3 := \tau + \kappa + 1$.

Further suppose that $E \in \mathbb{R}^{q \times q}$ has entries which are independent, identically distributed standard normal random variables. It follows by Gaussian concentration that E is (C, c, γ)-concentrated with parameters $C = 2$, $c = \frac{1}{2}$, and $\gamma = 2$, and so by an application of Lemma 3.13 for $\epsilon = 4$, then

$$\mathbb{P}\left[\|E\|_2 > 6\sqrt{q}\right] \leq 2\exp\left(-\tfrac{1}{10}q\right).$$

Define $\hat{M} := M + E$ and organize the singular values of $\hat{M}$ in correspondence with the repeated singular values of M, namely write

$$\left\{\{\hat{\sigma}_{1,i_1}\}_{i_1=1}^{m}, \{\hat{\sigma}_{2,i_2}\}_{i_2=1}^{n}, \{\hat{\sigma}_{3,i_3}\}_{i_3=1}^{p}\right\}.$$

Suppose that $\tau, \kappa > (2 \times (6\sqrt{q}) + 1)$. Then we can use Weyl's inequality as a preliminary tool for selecting the threshold values α and β. In particular, such analysis yields that with high probability,

$$
\begin{aligned}
|\hat{\sigma}_{1,i_1} - 1| \leq \|E\|_2 &\implies & 0 \leq \hat{\sigma}_{1,i_1} \leq 6\sqrt{q} + 1, \\
|\hat{\sigma}_{2,i_2} - (\kappa + 1)| \leq \|E\|_2 &\implies & 6\sqrt{q} + 2 < \hat{\sigma}_{2,i_2} < \tau + \kappa - 6\sqrt{q}, \\
|\hat{\sigma}_{3,i_3} - (\tau + \kappa + 1)| \leq \|E\|_2 &\implies & \tau + \kappa + 1 - 6\sqrt{q} \leq \hat{\sigma}_{3,i_3} \leq \tau + \kappa + 1 + 6\sqrt{q}.
\end{aligned}
$$

For the choices $\alpha = 6\sqrt{q} + 2$ and $\beta = \tau + \kappa - 6\sqrt{q}$, observe that $\{\hat{\sigma}_{2,i_2}\}_{i_2=1}^{n} \subset (\alpha, \beta) \subset \mathbb{R}_{>0}$ while simultaneously $\{1, \kappa + 1, \tau + \kappa + 1\} \bigcap (\alpha, \beta) = \{\kappa + 1\}$. In this setting our perturbation theorems apply for κ sufficiently large. Namely, choosing $\delta \in (0, 1]$ and setting $t = \Theta(\log^{\delta} q)$ yields that for each $k \in [n]$ there exist positive constants c' and c'' such that with high probability,

$$|\hat{\sigma}_{2,k} - \sigma_2| \leq c't + c''.$$

To reiterate, this bound improves upon the bound implied by a naïve, terminal application of Weyl's inequality. Moreover, Example 3.15 demonstrates how Weyl's inequality may be invoked for the preliminary purpose of establishing threshold values when the paired singular values (eigenvalues) correspond to the same index after ordering. ▲

4. Applications to graph inference

4.1. Methods of graph inference

The field of statistical inference and modeling for graphs represents a burgeoning area of research with implications for the social and natural sciences among other disciplines [12, 19]. Within the current body of research, the pursuit of identifying and studying community structure within real–world networks continues to receive widespread attention [2, 5, 11, 25, 26, 37]. Still another area of investigation involves anomaly detection for time series of graphs by considering graph statistics such as the total degree, number of triangles, and various scan statistics [33, 39]. Here we apply our results to two such detection tasks.

4.2. Community detection via hypothesis testing

In this application we view the problem of community detection through the lens of hypothesis testing as in [2, 37]. We consider the simple setting of a balanced three block stochastic block model and the problem of detecting differences in between-block communication. Namely, consider the block edge probability matrix and block assignment vector given by

$$\text{Null model: } B_0 = \begin{pmatrix} p & q & q \\ q & p & q \\ q & q & p \end{pmatrix} \text{ and } \pi_0 = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3} \right), \tag{4.1}$$

where $p = 0.81$ and $q = 0.2025$. In this model, vertices have an equal probability of belonging to each of the three blocks. Vertices within the same block have probability p of being connected by an edge, whereas vertices in different blocks have probability q of being connected by an edge.

As an aside, we note that this SBM may be cast in the language of random dot product graphs for which the underlying distribution of latent positions F is a mixture of point masses. Specifically, take F to be the discrete uniform distribution on the vectors $x_1 \approx (0.55, 0.32, 0.64)$, $x_2 \approx (-0.55, 0.32, 0.64)$, and $x_3 \approx (0, -0.64, 0.64)$ in $\mathbb{R}^3$ (see Remarks 3.8 and 3.9).

For a graph on n vertices from this three block model, condition on the graph exhibiting equal block sizes, i.e. $n_1 = n_2 = n_3 = n/3$. For the corresponding P matrix, denoted $P_n(B_0)$, the non-trivial (signal) model eigenvalues themselves exhibit multiplicity (hence Equation (3.5) via [3] does not apply) and are

$$\lambda_1(P_n(B_0)) = \lambda_2(P_n(B_0)) = \frac{n}{3}(p - q) \text{ and } \lambda_3(P_n(B_0)) = \frac{n}{3}(p + 2q). \tag{4.2}$$

In contrast, consider an alternative model in which the first and second blocks exhibit stronger between-block communication. This stronger communication is represented by an additional additive factor $\epsilon \in (0, p - q)$ in the block edge

probability matrix B_ϵ, where ϵ is assumed to be bounded away from $p - q$ for convenience.

$$\text{Alternative model: } B_\epsilon = \begin{pmatrix} p & q + \epsilon & q \\ q + \epsilon & p & q \\ q & q & p \end{pmatrix} \text{ and } \pi_1 = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right). \quad (4.3)$$

Under B_ϵ, the signal eigenvalues of $P_n(B_\epsilon)$ (equiv., singular values) can be explicitly computed as functions of p, q, n, and ϵ. They are given by

$$\lambda_1(P_n(B_\epsilon)) = \frac{n}{3}(p - q - \epsilon), \quad \lambda_2(P_n(B_\epsilon)) = \frac{n}{6}(2p + q + \epsilon - \sqrt{9q^2 + 2q\epsilon + \epsilon^2}),$$
$$\lambda_3(P_n(B_\epsilon)) = \frac{n}{6}(2p + q + \epsilon + \sqrt{9q^2 + 2q\epsilon + \epsilon^2}).$$

Furthermore, the maximum expected degree of the model corresponding to B_ϵ is given by $\Delta_\epsilon = \frac{n}{3}(p + 2q + \epsilon)$.

For $\epsilon > 0$, now consider a simple null versus simple alternative hypothesis test written as

$$\mathbb{H}_0 : B = B_0 \text{ against } \mathbb{H}_A : B = B_\epsilon. \quad (4.4)$$

In what follows we choose the smallest signal eigenvalue as our test statistic and denote it by Λ_1. We compare our bounds obtained via Kato–Temple methodology with the large-sample approximation bounds implied by [23] for the specified values $n \in \{6000, 9000, 12000, 15000\}$. Similar comparison can be carried out with respect to the results in [28]. Our bounds compare favorably with those in [28] even for conservative choices of t therein.

By Lemma 3.13 and Proposition 3.1, irrespective of $\epsilon > 0$ above, we have the concentration inequality $\mathbb{P}\left[\|E\|_2 > 3\sqrt{n}\right] \leq 2\exp\left(-\frac{1}{20}n\right)$. This spectral norm bound allows us to invoke an unconditional version of Theorem 3.3. Specifically, for moderate choices of $t > 0$, the bounds in Theorem 3.3 hold with probability at least $1 - 12\exp(-t^2) - 2\exp\left(-\frac{1}{20}n\right)$. When $n \geq 6000$, the choice $t \approx 2.66$ yields probability at least 0.99.

Using these concentration inequality results, we determine confidence intervals which hold for Λ_1 with probability at least 0.99 under $\mathbb{H}_0$ and $\mathbb{H}_A$, respectively. We compute the value ϵ_n such that the confidence intervals under $\mathbb{H}_0$ and $\mathbb{H}_A$ no longer overlap for $\epsilon \in (\epsilon_n, 0.2]$, emphasizing that smaller values of ϵ_n indicate superior performance. This provides us with a region of the alternative in which the statistical test has power at least 0.99. Our results are summarized in the numerical table below.

TABLE 1

n	ϵ_n via [23]	ϵ_n via this paper
6000	0.1006	0.0407
9000	0.0818	0.0256
12000	0.0707	0.0187
15000	0.0631	0.0147

It is not too difficult to realize that the eigenvalue-based test considered here has asymptotic power equal to one as $n \to \infty$ for any choice of $0 < q < p < 1$

and $\epsilon \in (0, p - q)$. Moreover, as a consequence of Theorem 3.7 and subsequent discussion, we make the following observation.

Proposition 4.1. *Consider testing the hypothesis in Equation (4.4). Assume that $q \equiv q_n = \omega(\frac{\log n}{n})$ with $q_n < p_n$. Then for $n\epsilon_n = \omega(\log n)$ and $\epsilon_n < p_n - q_n$, the above test using Λ_1 has asymptotically full power.*

Note that the above analysis investigates testing performance as a function of ϵ for graphs with fixed block proportions. Next we investigate a setting wherein ϵ is fixed and the sizes of the graph communities change.

4.3. Change-point detection

We now consider a stylized example of change-point detection via hypothesis testing. Let $T^* \geq 1$ and suppose that $G_1, G_2, \ldots, G_T$ for $T < T^*$ are Erdős–Rényi graphs on n vertices, while for $T \geq T^*$ the graph G_T is sampled according to a two block stochastic block model with block edge probability matrix $B = \begin{bmatrix} p_\epsilon & p \\ p & p \end{bmatrix}$ for $p_\epsilon := p + \epsilon$ and $\epsilon > 0$, with m vertices assigned to the first block and $n - m$ vertices assigned to the second block. We note that B encapsulates a notion of chatter anomaly, i.e., a subset of the vertices in $[n]$ exhibit altered communication behavior in an otherwise stationary setting. For a given value of T, we are interested in testing the hypothesis that T is a change-point in the collection $\{G_1, G_2, \ldots, G_T\}$. Given two graphs with adjacency matrices, $A^{(T-1)}$ and $A^{(T)}$, this can be formulated as the problem of testing the two-sample hypotheses

$$\mathbb{H}_0 \colon A^{(T-1)} \sim \mathrm{ER}(n, p), A^{(T)} \sim \mathrm{ER}(n, p) \quad \text{against}$$
$$\mathbb{H}_A \colon A^{(T-1)} \sim \mathrm{ER}(n, p), A^{(T)} \sim \mathrm{SBM}(B, m, n - m).$$

We emphasize that in the above formulation, the parameter p in $\mathrm{ER}(n, p)$, the size m of the chatter community, and the associated communication probability p_ϵ are generally assumed to be unknown.

Many test statistics are available for this change-point detection problem, including those based on graph invariant statistics (such as number of edges or number of triangles) or those based on locality statistics (such as max degree or scan statistics). For a given graph with adjacency matrix A, let $N(i) = \{j \colon A_{i,j} = 1\}$ denote the collection of vertices adjacent to vertex i. Furthermore,

- let $\mathcal{T}_k$ count the number of k-cliques in A for $k \geq 2$;
- let $\delta(A) := \max_i \sum_j A_{i,j}$ be the max degree statistic of A;
- let $\Psi(A) := \max_i \sum_{j,k \in N(i)} A_{j,k}$ be the scan statistic of A.

We note that these test statistics are widely used in anomaly detection for time series of graphs; see [2, 29, 30, 39] and the references therein for a survey of results and applications.

One can then show [31, 34] that the test statistics based on $\mathcal{T}_2$ and $\mathcal{T}_3$ are consistent for the above hypothesis test when $m = \Omega(\sqrt{n})$. More precisely, under

the null hypothesis, one has

$$\frac{\mathcal{T}_2(A^{(T)}) - \mathcal{T}_2(A^{(T-1)})}{n\sqrt{p(1-p)}} \xrightarrow{d} N(0,1); \quad \frac{\mathcal{T}_3(A^{(T)}) - \mathcal{T}_3(A^{(T-1)})}{n^2 p^2 \sqrt{pp_\epsilon}} \xrightarrow{d} N(0,1),$$

as $n \to \infty$, while under the alternative hypothesis, one has

$$\frac{\mathcal{T}_2(A^{(T)}) - \mathcal{T}_2(A^{(T-1)})}{n\sqrt{p(1-p)}} \xrightarrow{d} N\Big(\frac{m(m-1)\epsilon}{n\sqrt{p(1-p)}}, C_1\Big);$$

$$\frac{\mathcal{T}_3(A^{(T)}) - \mathcal{T}_3(A^{(T-1)})}{n^2 p^2 \sqrt{pp_\epsilon}} \xrightarrow{d} N\Big(\frac{\mu_{n,m,p,\epsilon}}{n^2 p^2 \sqrt{pp_\epsilon}}, C_2\Big),$$

as $n \to \infty$ for some positive constants C_1 and C_2 together with $\mu_{n,m,p,\epsilon} := m^3 p_\epsilon^3/6 + m^2(n-m)p^2 p_\epsilon + (m(n-m)^2/2 + (n-m)^3/6)p^3 - n^3 p^3/6$. Now, if $m = \omega(\sqrt{n})$, then

$$\frac{m(m-1)\epsilon}{n\sqrt{p(1-p)}} \to \infty; \quad \frac{\mu_{n,m,p,\epsilon}}{n^2 p^2 \sqrt{pp_\epsilon}} \to \infty,$$

as $n \to \infty$, and thus both $\mathcal{T}_2$ and $\mathcal{T}_3$ are consistent for the above hypothesis test when $m = \Omega(\sqrt{n})$. Furthermore, Theorem 2 and Proposition 2 of [2] indicate that $\mathcal{T}_2$ is asymptotically optimal, i.e., if $m = o(\sqrt{n})$ then provided that

$$\lim_{n\to\infty} \mathcal{I}(m,n,p,\epsilon) := \lim_{n\to\infty} \frac{m\big(p_\epsilon \log\frac{p_\epsilon}{p} + (1-p_\epsilon)\log\frac{1-p_\epsilon}{1-p}\big)}{2\log(n/m)} < 1, \qquad (4.5)$$

no test statistic is consistent for testing the above hypotheses. Similarly, one can also show [32, 34] that the test statistics based on $\delta(A)$ and $\Psi(A)$ are consistent for the above hypothesis test when $m = \Omega(\sqrt{n\log n})$; in particular, the (normalized) limiting distributions of both $\delta(A^{(T)}) - \delta(A^{(T-1)})$ and $\Psi(A^{(T)}) - \Psi(A^{(T-1)})$ is the Gumbel distribution.

In the context of this paper, one could also use a test statistic based on the largest eigenvalue. Our earlier results indicate that, under the null hypothesis, with high probability the largest eigenvalues of $A^{(T)}$ and $A^{(T-1)}$ satisfy

$$|\lambda_{\max}(A^{(T)}) - \lambda_{\max}(P^{(T)})| = O(1) \text{ and } |\lambda_{\max}(A^{(T-1)}) - \lambda_{\max}(P^{(T-1)})| = O(1),$$

along with $|\lambda_{\max}(A^{(T)}) - \lambda_{\max}(A^{(T-1)})| = O(1)$. Meanwhile, under the alternative hypothesis, when $m = o(n)$, then with high probability

$$\left| \big|\lambda_{\max}(A^{(T)}) - \lambda_{\max}(A^{(T-1)})\big| - \frac{m^2 p\epsilon}{np - m\epsilon} \right| = O(1).$$

Thus the largest eigenvalue test statistic is also consistent when $m = \Omega(\sqrt{n})$.

The previous test statistics are all global test statistics in the sense that, if $\mathbb{H}_0$ is rejected, the resulting test procedures do not extract the subset of the vertices which exhibits anomalous behavior between $A^{(T)}$ and $A^{(T-1)}$. One can

construct related local test statistics which do extract the subset of anomalous vertices, although the resulting test procedure is computationally prohibitive. For example, assuming that m is known, we could replace $\Psi(A)$ with the (modified) scan statistic $\Upsilon_m(A) = \max_{|S|=m} \mathcal{T}_2(A_{|S})$ where $A_{|S}$ is the subgraph of A induced by the vertices in S and the maximum is taken over all subsets $S \subset [n]$ with $|S| = m$. Thus $\Upsilon_m(A)$ is the maximum number of edges in any subgraph induced by m vertices of A. By [2] the test statistic $\Upsilon_m(A^{(T)}) - \Upsilon_m(A^{(T-1)})$ is consistent for the hypothesis test considered in this section whenever

$$\lim_{n \to \infty} \mathcal{I}(m, n, p, \epsilon) > 1.$$

Thus, for any fixed p and ϵ, the (modified) scan statistic is consistent when $m = \Omega(\log n)$ as $n \to \infty$. Using a similar idea, one can define a local variant of the largest eigenvalue statistic as $\Lambda_m(A) = \max_{|S|=m} \lambda_{\max}(A_{|S})$. By Theorem 3.3 and a union bound over all $\binom{n}{m} = O(n^m)$ subsets $S \subseteq [n]$ with $|S| = m$, we have that there exists a constant $C > 0$ such that if $t = C\sqrt{m \log n}$, then with high probability

$$|\Lambda_m(A^{(T)}) - \Lambda_m(A^{(T-1)})| = O(\sqrt{m \log n})$$

under the null hypothesis, whereas under the alternative hypothesis, with high probability

$$\left| |\Lambda_m(A^{(T)}) - \Lambda_m(A^{(T-1)})| - m\epsilon \right| = O(\sqrt{m \log n}).$$

Thus for any fixed p and ϵ, the test statistic based on Λ_m is also consistent for the above hypothesis test whenever $m = \Omega(\log n)$ as $n \to \infty$.

In summary, the results in Section 3 facilitate eigenvalue-based test statistics for the change-point detection problem as presented in this section. Furthermore, the resulting procedure is consistent whenever the size of the chatter community m exceeds the threshold of detectability given in [2].

5. Acknowledgments

6. Appendix

6.1. Proof of Theorem 3.3

Proof. Let $P, E \in \mathbb{R}^{n \times n}$ be real symmetric matrices such that E satisfies Proposition 3.1. Denote the d largest eigenvalues of P and A by

$$0 < \lambda_1(P) \leq \lambda_2(P) \leq \cdots \leq \lambda_d(P),$$
$$0 < \lambda_1(A) \leq \lambda_2(A) \leq \cdots \leq \lambda_d(A).$$

Let $\{w_i\}_{i=1}^d$ denote a collection of orthonormal eigenvectors of P corresponding to the collection of eigenvalues $\{\lambda_i(P)\}_{i=1}^d$. Similarly, let $\{u_i\}_{i=1}^d$ denote a collection of orthonormal eigenvectors of A corresponding to the collection of eigenvalues $\{\lambda_i(A)\}_{i=1}^d$.

For each $i \in [d]$ define η_i to be an "approximate eigenvalue of A close to $\lambda_i(P)$" in the sense that

$$\eta_i := \langle Aw_i, w_i \rangle = \lambda_i(P) + \langle Ew_i, w_i \rangle, \tag{6.1}$$

and define a corresponding "residual quantity" ϵ_i as

$$\epsilon_i := \|(A - \eta_i)w_i\|. \tag{6.2}$$

6.1.1. Proof of Theorem 3.3: upper bound

Now for fixed $k \in [d]$ define the k-dimensional linear manifold $\mathcal{M}_k$ by

$$\mathcal{M}_k := \text{span}\{u_1, \ldots, u_k\}.$$

We now define a collection of "aggregate quantities":

- Define w to be an "aggregate approximate eigenvector of A" in the sense that $w := \sum_{i=1}^k r_i w_i$ for a collection of normalized coefficients $\{r_i\}_{i=1}^k$ such that $\|w\|^2 = \sum_{i=1}^k r_i w_i = 1$, and satisfying the under-determined linear system $\langle w, u_i \rangle = 0$ for $i = 1, 2, \ldots, k-1$.
- Define η to be an "aggregate approximate eigenvector of A" in the sense that $\eta := \langle Aw, w \rangle$.
- Define ϵ to be the "aggregate residual quantity" $\epsilon := \|(A - \eta)w\|$.

By Lemma 1 in [18], the interval $\left(\alpha, \eta + \frac{\epsilon^2}{\eta - \alpha}\right]$ contains a point in the spectrum of A. Note that by construction, $w \in \mathcal{M}_{k-1}^{\perp} =: \mathcal{N}_{k-1}$; moreover, $Aw \in \mathcal{N}_{k-1}$ as a function of $\{r_i\}_{i=1}^k$. In the Hilbert space $\mathcal{N}_{k-1}$, however, the spectrum of A does not contain $\lambda_1(A), \ldots, \lambda_{k-1}(A)$ since $u_1, \ldots, u_{k-1} \notin \mathcal{N}_{k-1}$. Thus, by another application of Lemma 1 in [18], the eigenvalue of A in the interval given by $\left(\alpha, \eta + \frac{\epsilon^2}{\eta - \alpha}\right]$ must be $\lambda_k(A)$ with associated unit eigenvector u_k. Hence,

$$\lambda_k(A) \leq \eta + \frac{\epsilon^2}{\eta - \alpha} = \frac{\eta^2 + \epsilon^2 - \alpha\eta}{\eta - \alpha}. \tag{6.3}$$

We pause briefly to make several computational observations. First,

$$\eta^2 + \epsilon^2 = \langle Aw, w \rangle^2 + \|(A - \langle Aw, w \rangle)w\|^2 \tag{6.4}$$

$$= \|Aw\|^2 = \sum_{i,j=1}^k r_i r_j \langle Aw_i, Aw_j \rangle. \tag{6.5}$$

Letting $\delta_{i,j} := \mathbb{I}\{i = j\}$ denote the Kronecker delta function, we have for each $i, j \in [d]$ that

$$\langle Aw_i, Aw_j \rangle = \langle (A - \eta_i)w_i, (A - \eta_j)w_j \rangle + (\eta_i + \eta_j)\langle Aw_i, w_j \rangle - \eta_i^2 \delta_{i,j}, \quad (6.6)$$

while

$$\langle Aw_i, w_j \rangle = \langle Ew_i, w_j \rangle \text{ for } i \neq j. \quad (6.7)$$

It will also prove useful to recognize the expansion

$$\eta = \langle Aw, w \rangle = \sum_{i=1}^{k} r_i^2 \eta_i + \sum_{1 \leq i < j \leq k} 2r_i r_j \langle Ew_i, w_j \rangle. \quad (6.8)$$

Combining these observations yields

$$
\begin{aligned}
\eta^2 + \epsilon^2 &= \sum_{i,j=1}^{k} r_i r_j \langle Aw_i, Aw_j \rangle \\
&= \left(\sum_{i=1}^{k} r_i^2 \langle Aw_i, Aw_i \rangle \right) + \left(\sum_{1 \leq i < j \leq k} 2r_i r_j \langle Aw_i, Aw_j \rangle \right) \\
&= \sum_{i,j=1}^{k} r_i r_j \langle (A - \eta_i)w_i, (A - \eta_j)w_j \rangle \\
&\quad + \sum_{i=1}^{k} r_i^2 \eta_i^2 + \sum_{1 \leq i < j \leq k} 2r_i r_j (\eta_i + \eta_j)\langle Aw_i, w_j \rangle.
\end{aligned}
$$

An application of the Cauchy–Schwarz inequality coupled with subsequent computation yields

$$
\sum_{i,j=1}^{k} r_i r_j \langle (A - \eta_i)w_i, (A - \eta_j)w_j \rangle \leq \sum_{i,j=1}^{k} r_i r_j \left(\|(A - \eta_i)w_i\| \|(A - \eta_j)w_j\| \right)
$$

$$
= \sum_{i,j=1}^{k} (r_i \epsilon_i)(r_j \epsilon_j) \leq \left(\sum_{i=1}^{k} \epsilon_i |r_i| \right)^2.
$$

Hence,

$$\eta^2 + \epsilon^2 \leq \left(\sum_{i=1}^{k} \epsilon_i |r_i| \right)^2 + \sum_{i=1}^{k} r_i^2 \eta_i^2 + \sum_{1 \leq i < j \leq k} 2r_i r_j (\eta_i + \eta_j)\langle Ew_i, w_j \rangle. \quad (6.9)$$

Returning to Eqn. (6.3), the numerator then becomes

$$\left(\sum_{i=1}^{k} \epsilon_i |r_i| \right)^2 + \sum_{i=1}^{k} r_i^2 \eta_i (\eta_i - \alpha) + \sum_{1 \leq i < j \leq k} 2r_i r_j (\eta_i + \eta_j - \alpha)\langle Ew_i, w_j \rangle \quad (6.10)$$

while the denominator becomes

$$\left(\sum_{i=1}^{k} r_i^2(\eta_i - \alpha)\right) + \left(\sum_{1 \le i < j \le k} 2r_i r_j \langle Ew_i, w_j \rangle\right). \tag{6.11}$$

By a simple union bound, observe that for $t > 0$,

$$\mathbb{P}\left[\max_{1 \le i \le j \le k} |\langle Ew_i, w_j \rangle| > t\right] \le \left(k + \binom{k}{2}\right) C \exp(-ct^\gamma), \tag{6.12}$$

in which case with high probability,

$$\left(\sum_{i=1}^{k} r_i^2(\eta_i - \alpha)\right) \ge \lambda_1(P) - \alpha - t, \tag{6.13}$$

$$\left(\sum_{1 \le i < j \le k} 2r_i r_j \langle Ew_i, w_j \rangle\right) \ge -k(k-1)t, \tag{6.14}$$

while with high probability,

$$\sum_{1 \le i < j \le k} 2r_i r_j(\eta_i + \eta_j - \alpha)\langle Ew_i, w_j \rangle \le (2\lambda_k(P) - \alpha + 2t)k(k-1)t, \tag{6.15}$$

$$\left(\sum_{i=1}^{k} r_i^2 \eta_i\right) k(k-1)t \le (\lambda_k(P) + t)k(k-1)t. \tag{6.16}$$

By adding and subtracting $\left(\sum_{i=1}^{k} r_i^2 \eta_i\right) k(k-1)t$ to the numerator of Eqn. (6.3) we obtain the following bound in which the first term on the right-hand side is the leading term while the second term on the right hand side corresponds to a residual term.

$$\lambda_k(A) \le \frac{\left(\sum_{i=1}^{k} \epsilon_i |r_i|\right)^2 + \left(\sum_{i=1}^{k} r_i^2 \eta_i(\eta_i - \alpha - k(k-1)t)\right)}{\left(\sum_{i=1}^{k} r_i^2(\eta_i - \alpha - k(k-1)t)\right)}$$
$$+ \frac{(3\lambda_k(P) - \alpha + 3t)k(k-1)t}{\lambda_1(P) - \alpha - (k(k-1) + 1)t}.$$

Now by the same arguments as in [18], Section 3, Eqns. (22–30), the constants $\{r_i\}_{i=1}^{k}$ can be removed. To this end, the quantity

$$\frac{\left(\sum_{i=1}^{k} \epsilon_i |r_i|\right)^2 + \left(\sum_{i=1}^{k} r_i^2 \eta_i(\eta_i - \alpha - k(k-1)t)\right)}{\left(\sum_{i=1}^{k} r_i^2(\eta_i - \alpha - k(k-1)t)\right)} \tag{6.17}$$

is bounded above by the quantity

$$\max_{1 \le i \le k} \eta_i + \left(\sum_{i=1}^{k} \frac{\epsilon_i^2}{\eta_i - \alpha - k(k-1)t}\right). \tag{6.18}$$

Note that $\max_{1 \leq i \leq k} \eta_i \leq \lambda_k(P) + t$, with high probability, while a simple computation reveals that for each $i \in [k]$,

$$\epsilon_i^2 = \|Ew_i\|^2 - |\langle Ew_i, w_i \rangle|^2 \leq \|Ew_i\|^2 \leq \|E\|_2^2. \tag{6.19}$$

Putting all these observations together finally produces an upper bound on $\lambda_k(A)$ of the form

$$\lambda_k(A) \leq \lambda_k(P) + t + \zeta^+, \tag{6.20}$$

where $\zeta^+ := \frac{k\|E\|_2^2 + (3\lambda_k(P) - \alpha + 3t)k(k-1)t}{\lambda_1(P) - \alpha - (k(k-1)+1)t}$.

6.1.2. Proof of Theorem 3.3: lower bound

Fix $k \in [d]$ and let $l := d - k + 1$. Define $\mathcal{M}_l$ to be the l-dimensional linear manifold given by

$$\mathcal{M}_l := \mathrm{span}\{u_k, \dots, u_d\}.$$

We now define a collection of "aggregate quantities" similar to the formulation in Section 6.1.1:

- Define w to be an "aggregate approximate eigenvector of A" in the sense that $w := \sum_{i=k}^d r_i w_i$ for a collection of normalized coefficients $\{r_i\}_{i=k}^d$ such that $\|w\|^2 = \sum_{i=k}^d r_i w_i = 1$, and satisfying the under-determined linear system $\langle w, u_i \rangle = 0$ for $i = k+1, \dots, d$.
- Define η to be an "aggregate approximate eigenvector of A" in the sense that $\eta := \langle Aw, w \rangle$.
- Define ϵ to be the "aggregate residual quantity" $\epsilon := \|(A - \eta)w\|$.

By Lemma 2 in [18], the interval $\left[\eta - \frac{\epsilon^2}{\beta - \eta}, \beta\right)$ contains a point in the spectrum of A. Note that by construction, $w \in \mathcal{M}_{l-1}^\perp =: \mathcal{N}_{l-1}$; moreover, $Aw \in \mathcal{N}_{l-1}$ as a function of $\{r_i\}_{i=k}^d$. In the Hilbert space $\mathcal{N}_{l-1}$, however, the spectrum of A does not contain $\lambda_{k+1}(A), \dots, \lambda_d(A)$ since $u_{k+1}, \dots, u_d \notin \mathcal{N}_{l-1}$. Thus, by another application of Lemma 2 in [18], the eigenvalue of A in the interval $\left[\eta - \frac{\epsilon^2}{\beta - \eta}, \beta\right)$ must be $\lambda_k(A)$ with associated unit eigenvector u_k.

Consider first the special case when $\beta = \infty$. By a simple union bound, observe that for $t > 0$,

$$\mathbb{P}\left[\max_{k \leq i \leq j \leq d} |\langle Ew_i, w_j \rangle| > t\right] \leq \left(l + \binom{l}{2}\right) C \exp(-ct^\gamma), \tag{6.21}$$

hence with high probability

$$\sum_{i=k}^d r_i^2 \eta_i \geq \min_{k \leq i \leq d} \eta_i \geq \lambda_k(P) - t \tag{6.22}$$

and

$$\lambda_k(A) \geq \eta = \sum_{i=k}^{d} r_i^2 \eta_i + \sum_{k \leq i < j \leq d} 2r_i r_j \langle Ew_i, w_j \rangle \tag{6.23}$$

$$\geq \lambda_k(P) - (l(l-1) + 1)t. \tag{6.24}$$

Now suppose that $\beta < \infty$. Then for the lower bound of the above interval, one has

$$\lambda_k(A) \geq \eta - \frac{\epsilon^2}{\beta - \eta} = \frac{\beta \eta - \eta^2 - \epsilon^2}{\beta - \eta} = \frac{-(\eta^2 + \epsilon^2) + \beta \eta}{\beta - \eta}.$$

Reversing the direction of the previous application of the Cauchy–Schwarz inequality in Eqn. (6.9) permits the numerator to be bounded below by

$$-\left(\sum_{i=k}^{d} \epsilon_i |r_i|\right)^2 + \sum_{i=k}^{d} r_i^2 \eta_i (\beta - \eta_i) + \sum_{k \leq i < j \leq d} 2r_i r_j (\beta - \eta_i - \eta_j) \langle Ew_i, w_j \rangle,$$

whereas the denominator has the expansion

$$\sum_{i=k}^{d} r_i^2 (\beta - \eta_i) + \sum_{k \leq i < j \leq d} 2r_i r_j \langle Ew_i, w_j \rangle.$$

For the denominator terms, note that with high probability

$$\sum_{i=k}^{d} r_i^2 (\beta - \eta_i) \geq \beta - \lambda_d(P) - t,$$

$$\sum_{k \leq i < j \leq d} 2r_i r_j \langle Ew_i, w_j \rangle \geq -l(l-1)t,$$

while in the numerator, with high probability,

$$\sum_{k \leq i < j \leq d} 2r_i r_j (\beta - \eta_i - \eta_j) \langle Ew_i, w_j \rangle \geq -(\beta - \lambda_k(P) + \lambda_d(P) + 2t)l(l-1)t.$$

In the numerator of Eqn. (6.23), add and subtract the quantity $\left(\sum_{i=k}^{d} r_i^2 \eta_i\right) l(l-1)t$ which is bounded below by $(\lambda_k(P) - t)l(l-1)t$. Combining these observations yields

$$\lambda_k(A) \geq \frac{-\left(\sum_{i=k}^{d} \epsilon_i |r_i|\right)^2 + \sum_{i=k}^{d} r_i^2 \eta_i (\beta - \eta_i - l(l-1)t)}{\sum_{i=k}^{d} r_i^2 (\beta - \eta_i - l(l-1)t)}$$
$$+ \frac{-(\beta - \lambda_k(P) + \lambda_d(P) - \lambda_k + 3t)l(l-1)t}{\beta - \lambda_d(P) - (l(l-1) + 1)t}.$$

By employing the same approach used to obtain the upper bound and taking negatives when necessary (thereby reversing the direction in which bounds hold), we obtain the lower bound for $\lambda_k(A)$ of the form

$$\lambda_k(A) \geq \lambda_k(P) - t - \zeta^-, \tag{6.25}$$

where $\zeta^- := \frac{l\|E\|_2^2 + ((\beta - \lambda_k(P)) + (\lambda_d(P) - \lambda_k(P)) + 3t)l(l-1)t}{\beta - \lambda_d(P) - (l(l-1)+1)t}$. □

6.2. Proof of Theorem 3.7

Proof. The hypotheses imply by [23] that $\|A - P\|_2 = O(\sqrt{\Delta})$ with probability $1 - o(1)$ as $n \to \infty$. Set $\alpha = (C - c)\Delta/2$ and $\beta = \infty$ as Kato–Temple threshold values. Choose $\delta \in (0, 1]$ and set $t = \Theta(\log^\delta n)$. Then in Theorem 3.3, for sufficiently large n, one has $\zeta^+, \zeta^- = O(t)$ where the underlying constant depends upon k, d, as well as underlying (unspecified) constants. So for $n \geq n_0$, then $|\hat{\sigma}_k - \sigma_k| \leq c_{k,d}t$ with probability $1 - o(1)$ as claimed. □

6.3. Proof of Theorem 3.12

Proof. The proof follows essentially *mutatis mutandis* as in Theorem 3.3 via Remark 2.4, Definition 3.10, and Lemma 3.11. In particular, observe that one has $\langle \tilde{M}\tilde{w}_i, \tilde{w}_j \rangle = \sigma_i \delta_{i,j} + \langle \tilde{E}\tilde{w}_i, \tilde{w}_j \rangle$ for each pair i, j, while at the same time $\|\tilde{E}\|_2 = \|E\|_2$. □

6.4. Proof of Lemma 3.13

Proof. Let $E \in \mathbb{R}^{m \times n}$ be a (C, c, γ)-concentrated random matrix. Take $\mathcal{X}$ and $\mathcal{Y}$ to be $\frac{1}{4}$-nets of the spheres S^{n-1} and S^{m-1}, respectively, with cardinalities at most 9^n and 9^m, respectively. Then a standard net argument yields that for $t > 0$,

$$\begin{aligned}
\mathbb{P}\left[\|E\|_2 > t\right] &\leq \mathbb{P}\left[2 \max_{x \in \mathcal{X}, y \in \mathcal{Y}} |\langle Ex, y \rangle| > t\right] \\
&\leq 9^{m+n} \mathbb{P}\left[|\langle Ex, y \rangle| > t/2\right] \\
&\leq C \exp((m + n)\log(9) - c(t/2)^\gamma) \\
&\leq C \exp(2\log(9)\max\{m, n\} - c(t/2)^\gamma).
\end{aligned}$$

Choose $\epsilon > 0$ such that $2 + \epsilon > 2 (2\log(9)/c)^{1/\gamma}$ and set $t = (2 + \epsilon)\max\{m, n\}^{1/\gamma}$. Then for $c_{\epsilon, c, \gamma} := (c(1 + \epsilon/2)^\gamma - 2\log(9)) > 0$, we have

$$\mathbb{P}\left[\|E\|_2 > (2 + \epsilon)\max\{m, n\}^{1/\gamma}\right] \leq C \exp(-c_{\epsilon, c, \gamma}\max\{m, n\}).$$

If in addition $m = n$ and E is assumed to be symmetric, then since $\|E\|_2 \equiv \sup_{\|x\|_2 = 1} |\langle Ex, x \rangle|$, one need only consider the $\frac{1}{4}$-net $\mathcal{X}$ for the purposes of a union bound. □

References

[1] Noga Alon, Michael Krivelevich, and Van Vu, *On the concentration of eigenvalues of random symmetric matrices*, Israel Journal of Mathematics **131** (2002), no. 1, 259–267. MR1942311

[2] Ery Arias-Castro and Nicolas Verzelen, *Community detection in dense random networks*, The Annals of Statistics **42** (2014), no. 3, 940–969. MR3210992

[3] Avanti Athreya, Carey E. Priebe, Minh Tang, Vince Lyzinski, David J. Marchette, and Daniel L. Sussman, *A limit theorem for scaled eigenvectors of random dot product graphs*, Sankhya A (2016), 1–18. MR3494576

[4] Konstantin Avrachenkov, Laura Cottatellucci, and Arun Kadavankandy, *Spectral properties of random matrices for stochastic block model*, Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on, IEEE, 2015, pp. 537–544.

[5] Peter J. Bickel and Aiyou Chen, *A nonparametric view of network models and Newman–Girvan and other modularities*, Proceedings of the National Academy of Sciences **106** (2009), no. 50, 21068–21073.

[6] Peter J. Bickel and Purnamrita Sarkar, *Hypothesis testing for automated community detection in networks*, preprint arXiv:1311.2694 (2013). MR3453655

[7] Béla Bollobás, Svante Janson, and Oliver Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures and Algorithms **31** (2007), no. 1, 3–122. MR2337396

[8] Fan R. K. Chung, *Spectral graph theory*, vol. 92, American Mathematical Soc., 1997. MR1421568

[9] Xue Ding and Tiefeng Jiang, *Spectral distributions of adjacency and Laplacian matrices of random graphs*, The Annals of Applied Probability **20** (2010), no. 6, 2086–2117. MR2759729

[10] Paul Erdős and Alfréd Rényi, *On random graphs*, Publicationes Mathematicae (Debrecen) **6** (1959), 290–297. MR0120167

[11] Santo Fortunato, *Community detection in graphs*, Physics Reports **486** (2010), no. 3, 75–174. MR2580414

[12] Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi, *A survey of statistical network models*, Foundations and Trends in Machine Learning **2** (2010), no. 2, 129–233.

[13] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock, *Latent space approaches to social network analysis*, Journal of the American Statistical Association **97** (2002), no. 460, 1090–1098. MR1951262

[14] Paul W. Holland, Kathryn B. Laskey, and Samuel Leinhardt, *Stochastic blockmodels: first steps*, Social Networks **5** (1983), no. 2, 109–137. MR0718088

[15] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, 2012. MR2978290

[16] Evans M. Harrell II, *Generalizations of Temple's inequality*, Proceedings of the American Mathematical Society **69** (1978), no. 2, 271–276. MR0487733

[17] Brian Karrer and Mark E. J. Newman, *Stochastic blockmodels and community structure in networks*, Physical Review E **83** (2011). MR2788206

[18] Tosio Kato, *On the upper and lower bounds of eigenvalues*, Physical Review Letters **77** (1950), 334–339. MR0038738

[19] Eric D. Kolaczyk, *Statistical analysis of network data*, Springer Series in Statistics, Springer, New York, 2009. MR2724362

[20] Can M. Le and Roman Vershynin, *Concentration and regularization of random graphs*, arXiv:1506.00669 (2015).

[21] Jing Lei, *A goodness-of-fit test for stochastic block models*, The Annals of Statistics **44** (2016), no. 1, 401–424. MR3449773

[22] Jing Lei and Alessandro Rinaldo, *Consistency of spectral clustering in stochastic block models*, The Annals of Statistics **43** (2015), no. 1, 215–237. MR3285605

[23] Linyuan Lu and Xing Peng, *Spectra of edge-independent random graphs*, The Electronic Journal of Combinatorics **20** (2013), no. 4, 1–18. MR3158266

[24] Vince Lyzinski, Minh Tang, Avanti Athreya, Youngser Park, and Carey E. Priebe, *Community detection and classification in hierarchical stochastic blockmodels*, IEEE Transactions on Network Science and Engineering **4** (2017), no. 1, 13–26. MR3625952

[25] Mark E. J. Newman, *Modularity and community structure in networks*, Proceedings of the National Academy of Sciences **103** (2006), no. 23, 8577–8582.

[26] Mark E. J. Newman and Michelle Girvan, *Finding and evaluating community structure in networks*, Physical Review E **69** (2004), no. 2, 026113.

[27] Roberto I. Oliveira, *Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges*, arXiv:0911.0600 (2010). MR2447295

[28] Sean O'Rourke, Van Vu, and Ke Wang, *Random perturbation of low rank matrices: improving classical bounds*, arXiv:1311.2657 (2013).

[29] Carey E. Priebe, John M. Conroy, David J. Marchette, and Youngser Park, *Scan statistics on Enron graphs*, Computational and Mathematical Organization Theory **11** (2005), 229–247.

[30] Stephen Ranshous, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, and Nagiza F. Samatova, *Anomaly detection in dynamic networks: a survey*, WIRES Computational Statistics **7** (2015), 223–247. MR3349301

[31] Andrey Rukhin and Carey E. Priebe, *A comparative power analysis of the maximum degree and size invariants for random graph inference*, Journal of Statistical Planning and Inference **141** (2011), 1041–1046. MR2732971

[32] Andrey Rukhin and Carey E. Priebe, *On the limiting distribution of a graph scan statistic*, Communication in Statistics: Theory and Methods **41** (2012), 1151–1170.

[33] Andrey Rukhin, *Asymptotic analysis of various statistics for random graph inference*, ProQuest Dissertations and Theses (2009). MR2713849

[34] Minh Tang, Youngser Park, Nam H. Lee, and Carey E. Priebe, *At-*

tribute fusion in a latent process model for time series of graphs, IEEE
Transactions on Signal Processing **61** (2013), 1721–1732. MR3038385

[35] Minh Tang and Carey E. Priebe, *Limit theorems for eigenvectors of the
normalized Laplacian for random graphs*, preprint arXiv:1607.08601 (2016).

[36] Minh Tang, Daniel L. Sussman, and Carey E. Priebe, *Universally consistent vertex classification for latent positions graphs*, The Annals of
Statistics **41** (2013), no. 3, 1406–1430. MR3113816

[37] Nicolas Verzelen and Ery Arias-Castro, *Community detection in sparse
random networks*, The Annals of Applied Probability **25** (2015), no. 6,
3465–3510. MR3404642

[38] Ulrike Von Luxburg, *A tutorial on spectral clustering*, Statistics and
Computing **17** (2007), no. 4, 395–416. MR2409803

[39] Heng Wang, Minh Tang, Youngser Park, and Carey E. Priebe, *Locality
statistics for anomaly detection in time series of graphs*, IEEE Transactions
on Signal Processing **62** (2014), no. 3, 703–717. MR3160307

[40] Stephen Young and Edward Scheinerman, *Random dot product graph
models for social networks*, Algorithms and Models for the Web-Graph
(2007), 138–149. MR2504912

[41] Xiao Zhang, Raj Rao Nadakuditi, and Mark E. J. Newman, *Spectra of
random graphs with community structure and arbitrary degrees*, Physical
Reivew E **89** (2014).

[42] Yunpeng Zhao, Elizaveta Levina, and Ji Zhu, *Consistency of community
detection in networks under degree-corrected stochastic block models*, The
Annals of Statistics **40** (2012), no. 4, 2266–2292. MR3059083