S&DS 684: Statistical Inference on Graphs

Fall 2018

Lecture 8: Correlated recovery Lecturer: Yihong Wu

Scribe: Soham Jana, October 31, 2018

Recall model

$$G \sim SBM(n, p, q)$$

$$\sigma = (\sigma_1, \dots, \sigma_n) \in \{\pm 1\}^n$$

$$\mathbb{P}[i \sim j] = \begin{cases} p \text{ if } \sigma_i = \sigma_j \\ q \text{ if } \sigma_i \neq \sigma_j. \end{cases}$$

Goal: As described in Section ??, an estimator $\hat{\sigma} = \hat{\sigma}(G)$ achieves correlated recovery if the overlap is strictly better than random guessing, that is,

$$|\langle \hat{\sigma}, \sigma \rangle| \ge \Omega(n) \text{ as } n \to \infty \Leftrightarrow \min_{\pm} \|\hat{\sigma} \pm \sigma\|_1 \le (\frac{1}{2} - \Omega(1)) \cdot n.$$

8.1 Impossibility

We start with an information theoretic characterization of correlated recovery:

Theorem 8.1 (Mutual information characterization). Correlated recovery is possible $\Leftrightarrow I(\sigma_1, \sigma_2; G) = \Omega(1)$ as $n \to \infty$.

Remark 8.1 (Mutual information and probability of error). Note that for all $x_1, x_2 \in \{\pm\}$,

$$Law(G|\sigma_1 = x_1, \sigma_2 = x_2) = Law(G|\sigma_1 = -x_1, \sigma_2 = -x_2)$$

This means the product $\sigma_1 \sigma_2$ is a sufficient statistic of the pair (σ_1, σ_2) for G and hence

$$I(\sigma_1, \sigma_2; G) = I(\sigma_1 \sigma_2; G)$$

The condition $I(\sigma_1 \sigma_2; G) = \Omega(1)$ means that G offers some nontrivial information so that one can decide whether a (or any) pair of vertices have the same label better than chance. This can be quantified as follows.

Aside: mutual information vs probability of error. Suppose we have two random variables $X \sim \overline{\text{Rad}(\frac{1}{2})}$ and Y. Then

$$\min_{\hat{X}(\cdot)} \mathbb{P}(X \neq \hat{X}(Y)) = \frac{1}{2} [1 - \mathrm{TV}(P_+, P_-)].$$
(8.1)

where

$$P_{+} \triangleq \mathcal{L}(Y|X = +) =$$
$$P_{-} \triangleq \mathcal{L}(Y|X = -)$$

So no better than random guess $\Leftrightarrow TV(P_+, P_-) = o(1)$. We claim this is equivalent to $I(X; Y) \to 0$.

Indeed,

$$\begin{split} I(X;Y) &= \mathbb{E}_X \left[D(P_{Y|X} \| P_Y \right]) \\ &= \frac{1}{2} \left[D(P_+ \| \bar{P}) + D(P_- \| \bar{P}) \right] \\ & \bar{P} = \frac{P_+ + P_-}{2} \\ & \stackrel{Pinsker}{\geq} \mathrm{TV}^2(P_+, \bar{P}) + \mathrm{TV}^2(P_-, \bar{P}) \\ &= \frac{1}{2} \mathrm{TV}^2(P_+, P_-). \end{split}$$

On the other hand, from the inequality $D \leq \chi^2$ we get

$$I(X;Y) \leq \frac{1}{2} \left[\chi^2(P_+ \|\bar{P}) + \chi^2(P_- \|\bar{P}) \right]$$

= $\frac{1}{2} \left[\int \frac{(P_+ - \bar{P})^2}{\bar{P}} + \int \frac{(P_- - \bar{P})^2}{2} \right]$
= $\int \frac{(P_- - P_+)^2}{2(P_+ + P_-)} \leq \frac{1}{2} \int |P_+ - P_-| = \mathrm{TV}(P_+, P_-).$

Remark 8.2. Mutual information characterization in Theorem 8.1 holds under much more general conditions, e.g., *k*-community SBM. See [WX18, Appendix A].

Proof of Theorem 8.1. (" \Leftarrow ") Suppose that $I(\sigma_1, \sigma_2; G) \ge \epsilon$. Then by symmetry $I(\sigma_i, \sigma_j; G) \ge \epsilon$ for all $i \ne j$. Therefore,

by Remark 8.1 and (8.1), for all $i \neq j$, $\exists \hat{T}_{ij} = \hat{T}_{ij}(G)$, such that

$$\mathbb{P}\{\hat{T}_{ij} = \underbrace{\sigma_i \sigma_j}_{T_{ij}}\} \ge \frac{1}{2} + \delta.$$

for some $\delta = \delta(\epsilon)$. Then we can define an estimator of the labels $\hat{\sigma} = (\hat{\sigma}_1, \dots, \hat{\sigma}_n)$ by

$$\hat{\sigma}_1 = +, \hat{\sigma}_i = \hat{T}_{1i}; \ i = 2, \dots, n.$$

Then the expected number of correctly classified nodes is

$$\max_{\pm} \sum_{i \in [n]} \mathbb{P}\left[\sigma_i = \pm \hat{\sigma}_i\right] = \sum_{i \in [n]} \mathbb{P}\left[T_{1i} = \hat{T}_{1i}\right] \ge (1/2 + \delta)n.$$

(" \Rightarrow ") Suppose $I(\sigma_i, \sigma_j; G) = o(1)$. Then $\forall \hat{T}_{ij}, \mathbb{P}[\hat{T}_{ij} = \sigma_i \sigma_j] = \frac{1}{2} + o(1)$. This means given $\hat{\sigma} = (\hat{\sigma}_1, \dots, \hat{\sigma}_n)$, we have

$$2n^{2} - \mathbb{E}|\langle \sigma, \hat{\sigma} \rangle|^{2} = \mathbb{E}||\sigma\sigma^{\top} - \hat{\sigma}\hat{\sigma}^{T}||_{F}^{2}$$
$$= 4 \cdot \sum_{i \neq j} \mathbb{P}(\sigma_{i}\sigma_{j} \neq \hat{\sigma}_{i}\hat{\sigma}_{j})$$
$$= 2n^{2} - o(n^{2}),$$

which means $\mathbb{E}|\langle\hat{\sigma},\sigma\rangle|^2 = o(n^2)$, or $|\langle\sigma,\hat{\sigma}\rangle| = o_P(n)$.

Next we show that

$$\tau = \frac{(a-b)^2}{2(a+b)} < 1 \implies I(\sigma_1, \sigma_2; G) = o(1) \implies \text{ Correlation recovery impossible.}$$

First note the following variational representation of total variation:

$$TV(P_+, P_-) = \frac{1}{2} \inf_Q \sqrt{\int \frac{(P_+ - P_-)^2}{Q}}.$$
(8.2)

Proof. By C-S, $\int \frac{(P_+ - P_-)^2}{Q} = \int (\frac{P_+ - P_-}{\sqrt{Q}})^2 \int (\sqrt{Q})^2 \ge (\int |P_+ - P_-|)^2 = 4\text{TV}^2$, with equality if $Q = |P_+ - P_-| / \int |P_+ - P_-|$.

To apply this variational representation, take $Q = \text{Law of } G(n, \frac{d}{n})$. To show $I(\sigma_1 \sigma_2; G) = o(1)$, it suffices to show $\int \frac{(P_+ - P_-)^2}{Q} = o(1)$. This is a second-moment calculation similar to what we did in Lecture ?? for detection. The difference is that here there is no null model. Write

$$\int \frac{(P_+ - P_-)^2}{Q} = \int \frac{P_+^2}{Q} + \int \frac{P_-^2}{Q} - 2 \int \frac{P_+ P_-}{Q}$$

Next we show that $\int \frac{P_z P_{\tilde{z}}}{Q} = constant + o(1), \ z, \tilde{z} \in \{\pm 1\}$. Consider the case of iid labels. By the same argument in Section ??, we have

$$\begin{split} \int \frac{P_z P_{\tilde{z}}}{Q} & F_{\underset{i < j}{=} i} \mathbb{E} \left[\exp(\frac{\tau + o(1)}{n} \underbrace{\sum_{i < j} \sigma_i \sigma_j \tilde{\sigma}_i \tilde{\sigma}_j}) \middle| \sigma_1 \sigma_2 = z, \tilde{\sigma}_1 \tilde{\sigma}_2 = \tilde{z} \right] \\ &= (1 + o(1)) \mathbb{E} \left[\exp(\frac{\tau + o(1)}{2} \underbrace{\underbrace{\frac{1}{n} \langle \sigma, \tilde{\sigma} \rangle^2}}_{N(0,1)^2}) \middle| \sigma_1 \sigma_2 = z, \tilde{\sigma}_1 \tilde{\sigma}_2 = \tilde{z} \right] \\ &\to \mathbb{E} \exp\left(\frac{\tau + o(1)}{2} N(0,1)^2\right) \triangleq C(\tau). \end{split}$$

Here the justification of the CLT steps is almost the same as before: $\frac{1}{\sqrt{n}}\langle \sigma, \tilde{\sigma} \rangle = \frac{1}{\sqrt{n}} \sum_{j=3}^{n} \sigma_j \tilde{\sigma}_j + \frac{1}{\sqrt{n}} (\sigma_1 \tilde{\sigma}_1 + \sigma_2 \tilde{\sigma}_2)$, where the first term is asymptotically N(0, 1) and independent of $\sigma_1, \sigma_2, \tilde{\sigma}_1, \tilde{\sigma}_2$, and the last term is negligible.

More generally,

- For exact bisection the same statement holds true, except one should be more careful with the conditioning.
- For the spiked Wigner model (??), the same calculation shows that $\lambda < 1 \implies$ correlated recovery is impossible.
- In fact, for the spiked Wigner model, one can directly prove (by a sample splittin reduction) that impossibility of detection ⇒ impossibility of correlated recovery (Homework).

8.2 Correlated recovery via spectral methods

Next we explain how to achieve the sharp threshold of correlated recovery via suitable versions of spectral methods. We only provide the main ideas and some proof sketch.

Spiked Wigner model: Let's rewrite (??) as follows:

$$W = \frac{\mu}{n} \sigma \sigma^\top + Z.$$

where the entries of Z is $N(0, \frac{1}{n})$, so that its eigenvalues are between [-2, 2] with high probability.

Consider the following spectral method for estimation σ : take the top eigenvector $\hat{u} = u_1(W)$ of the matrix W corresponding to the largest eigenvalue λ_1 , and report $\operatorname{sign}(u_1)$ as the estimate $\hat{\sigma}$. Let $u = \frac{1}{\sqrt{n}}\sigma$. This method succeeds in correlated recovery if and only (why?) if $|\langle u, \hat{u} \rangle|$ is bounded away from 0.

The well-known BBP phase transition [BBAP05] states that

$$\lambda_1(W) \to \begin{cases} \mu + \frac{1}{\mu} & \text{if } \mu > 1\\ 2 & \text{if } \mu \le 1 \end{cases}$$

and correspondingly, \hat{u} is correlated with u if and only if $\lambda_1(W)$ escapes the bulk of the spectrum, namely,

$$|\langle u, \hat{u} \rangle| \to \begin{cases} 1 - \frac{1}{\mu^2} & \text{if } \mu > 1\\ 0 & \text{if } \mu \le 1, \end{cases}$$

SBM(n, p, q) **model:** Suppose that the adjacency matrix of the graph G is given by A. Mimicking the above Gaussian result, the "wishful thinking" on our part is to view

$$A = \mathbb{E}A + A - \mathbb{E}A$$

where

$$\mathbb{E}A = \boxed{\begin{array}{c|c} p & q \\ \hline q & p \end{array}} = \frac{p+q}{2} \ast \boxed{\begin{array}{c|c} 1 \\ \hline - \end{array}} + \frac{p-q}{2} \ast \boxed{\begin{array}{c|c} + & - \\ \hline - & + \end{array}}$$

and $Var(A_{ij} - \mathbb{E}A_{ij}) = \frac{d+o(1)}{n}$, with d = (a+b)/2. The first eigenvector of $\mathbb{E}A$ is uninformative, and the second is exactly the label. So we can consider taking the signs of the second eigenvector of A. If we pretend the entries of the perturbation $A - \mathbb{E}A$ are iid $N(0, \frac{d}{n})$, then making analogy to the Gaussian result shows that the sharp thresholding is given by $s = \frac{a-b}{2} > \sqrt{d}$, which is the exactly the sharp threshold we want to show.

However, applying spectral method to A itself does not work, as sparse graphs are plagued by high degree vertices. Indeed, for $G(n, \frac{d}{n})$ with constant d, it is known [KS03]

$$\lambda_1(A) = ||A|| = \sqrt{d_{\max}}(1 + o(1)), \quad d_{\max} = \Theta\left(\frac{\log n}{\log \log n}\right).$$
(8.3)

In fact, not only the top eigenvalue, $\lambda_i(A) = \lambda_1(1 - o(1))$ for an unbounded many of i [KS03, Sec. 4]. Suppose that $d_i = d_{\max}$, e_i is the *i*-th coordinate vector. Then $||A|| \ge \frac{||Ae_i||}{||e_i||} = \sqrt{d_{\max}}$. As the matrix has all non-negative entries, by Perron frobeneous theorem we can say that $||A|| = \lambda_1(A)$, which concludes the proof.

To see the effect of high-degree vertices, let's look at power iteration: say $d_i = d_{\text{max}}$. Then

$$(A^{2k})_{ii} = \sum_{i_2,\dots,i_{2k}} A_{ii_2} A_{i_2i_3} \cdots A_{i_{2k}i}$$

$$= \text{ number of closed walks from } i \text{ to } i \text{ of length } 2k$$

$$\geq d_{\max}^k,$$
(8.4)

where the last inequality follows by restricting to those backtracking paths that goes from i to one of its neighbors and immediately goes back. Thus

$$||A||^{2k} \ge ||Ae_i||_2^2 = e_i^\top A^{2k} e_i \ge d_{\max}^k$$

Thus $\lambda_1(A) = ||A|| \ge \sqrt{d_{\max}}$, where the first inequality follows from Perron-Frobenius theorem applied to the nonnegative matrix A. The other side can be shown by arguing that most of the contribution in the moment calculation comes from those backtracking paths. Thus the top eigenvalue $\lambda_1(A)$ is not bounded. In fact, correspondingly, the limiting spectral distribution of the bulk has unbounded support.

The fact that d_{max} is unbounded even when the average degree d is bounded is because of the following: for each v,

$$d_v \sim \operatorname{Binom}(n, \frac{d}{n}) \approx \operatorname{Poi}(d)$$

Pretending they are independent, the maximum of n iid Poisson is given by the $\frac{1}{n}$ -quantile, namely, $\frac{e^{-d}d^k}{k!} \approx \frac{1}{n}$, that is, $k \approx \frac{\log n}{\log \log n}$.

In summary: Adjacency matrix of sparse graphs is plagued by high-degree vertices, and the top eigenvector is localized on those vertices and not informative.

Solutions:

- 1. Regularize, e.g., remove high-degree vertices then apply spectral methods. However, it is unclear whether this achieves the sharp thresholds of $s^2 \ge d$. In [CO10] a sufficient condition of $s^2 \gtrsim d \log d$ is shown.
- 2. Turn to other matrices, e.g., the non-backtracking matrix, which we briefly explain next. The motivation comes from the above moment calculation (8.4), wherein the pathological behavior is due to backtracking in the neighborhood of the high-degree vertices, so we remove those.

8.3 Spectrum of non-backtracking matrices

Given a simple undirected graph G = (V, E). Denote the set of oriented edges (ordered pairs) by $\vec{E} = \{(u, v) : \{u, v\} \in E\}$. The non-backtracking matrix $B \in \{0, 1\}^{\vec{E} \times \vec{E}}$ is defined as follows: for $e = (e_1, e_2), f = (f_1, f_2) \in \vec{E}$,

$$B_{ef} = 1_{\{e_2 = f_1\}} 1_{\{e_1 \neq f_2\}}$$

Properties of NB matrix: Let n = |V|, m = |E|.

1. B is a $2m \times 2m$ matrix, and can be partitioned into four $m \times m$ blocks:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \qquad \begin{array}{c} B_{11} = B_{22}^T \\ B_{12}, B_{21} \text{ symmetric,} \end{array}$$

- 2. Row sum: $\forall e = (u, v), \sum_{e' \in \vec{E}} B_{ee'} = d_v 1.$
- 3. Singular values of B are $\{d_v 1 : v \in V\} \cup \{1\}$ and thus not informative. (Why? Consier BB^{\top}).
- 4. Spectrum (eigenvalues) of B: $det(I \lambda B) = (1 \lambda^2)^{m-n} det(I \lambda A + (D I)\lambda^2).$
- 5. Ihara-Bass identity [Ter10, p. 89]:

$$\det(\mathbf{I} - \lambda B) = (1 - \lambda^2)^{m-n} \det(\mathbf{I} - \lambda A + \lambda^2 (D - \mathbf{I})),$$
(8.5)

where $D = \text{diag}(d_v)$. This means B has 2(m - n) useless eigenvalues that are equal to ± 1 , and the rest of the 2n eigenvalues are useful.

6. *B* is not symmetric, but satisfies the following symmetry: Given $e = (e_1, e_2)$, let $e^{-1} = (e_2, e_1)$ denote its reversal. Then

$$(B^{\top})_{ef} = B_{e^{-1}f^{-1}}.$$
(8.6)

In matrix notation, let $P = (\mathbb{1} \{ e = f^{-1} \})$ denote the involution that maps a vector $(x_e : e \in \vec{E})$ to $(x_{e^{-1}} : e \in \vec{E})$ such that $P^{\top} = P$ and $P^2 = \mathbf{I}$. Then

 $B^{\top} = PBP$

(in other words, BP is a symmetric) and consequently $B^k = PB^kP$.

For sparse random graphs, the spectrum of the NB matrix looks like the following for $G(n, \frac{d}{n})$ and $SBM(n, \frac{a}{n}, \frac{b}{n})$: [BLM18]



In addition, the following result gives a spectral method based on B that achieves the optimal threshold:

Theorem 8.2 ([BLM18]). Let $s = \frac{a-b}{2}$, $d = \frac{a+b}{2}$. Let $u_2 = u_2(B)$ be the second largest eigenvector of B. Define

$$\hat{\sigma}_u = \operatorname{sign}\left(\sum_{e:e_1=u} (u_2)_e\right).$$

Then $\hat{\sigma}$ achieves correlated recovery if $s^2 > d$.

Proving this result is outside the scope here. We explain some intuitions:

Why is *B* not hindered by high-degree vertices? This applies to both Erdős-Rényi and SBM. Here we consider the former. In the previous section, we see for $G(n, \frac{d}{n})$, the outlier eigenvalues of *A* exist due to high-degree vertices. This no longer occurs for *B*. To explain some intuition, we apply the trace method to $B^k(B^{\top})^k$ for some large *k*. Claim that for each oriented edge *e*,

$$(B^k(B^{\top})^k)_{ee} \tag{8.7}$$

= # NB walks starting with e in k steps then reversing the last step and returning to e in k steps such as



Indeed, using the symmetry property,

$$(B^{k}(B^{\top})^{k})_{ee} = \sum_{e_{2}\dots e_{2k}} B_{e_{1}e_{2}}B_{e_{2}e_{3}}\dots B_{e_{k}e_{k+1}}B_{e_{k+1}e_{k+2}}^{\top}\dots B_{e_{2k}e_{2k+1}}^{\top} \qquad e_{1} = e, e_{2k+1} = e$$

$$\stackrel{(8.6)}{=} \sum B_{ee_{2}}B_{e_{2}e_{3}}\dots B_{e_{k}e_{k+1}}B_{e_{k+1}e_{k+2}}^{\top}\dots B_{e_{2k}e_{2k+1}}^{\top}$$

To simplify the counting in (8.7), crucially, recall the *locally tree-like structure* of sparse graphs: with high probability, for each vertex u, its k-hop neighborhood $N_k(u)$ is a tree, provided that k is not too big, e.g. $k = o(\log n)$. If $N_k(v)$ is a tree, then for each summand in (8.7), the path must reverse itself (otherwise there will be a cycle). Thus, on the event that locally tree-like structure holds, we have

$$(B^k(B^{\top})^k)_{ee} = k$$
th generation descendents of $u \approx d^k$,

even if the degree of u is as large as $\frac{\log n}{\log \log n}$! To justify the last step,

- For $G(n, \frac{d}{n})$, the local neighborhood behaves as (can be coupled to) a Galton Watson tree with offspring distribution Poi(d).
- For $SBM(n, \frac{a}{n}, \frac{b}{n})$, the local neighborhood behaves as a two-type Galton Watson tree, where the total offspring distribution is still Poi(d), and each + has $Poi(\frac{a}{2})$ children of type + and $Poi(\frac{b}{2})$ children of type -, and vice versa. This can be encoded into the following matrix:

$$M = \begin{bmatrix} \frac{a}{2} & \frac{b}{2} \\ \frac{b}{2} & \frac{a}{2} \end{bmatrix}.$$
(8.8)

Basic results in branching process states that the total number of kth-gen children grows exponentially as d^k .

Finally,

$$\sum_{k=1}^{2m} |\lambda_k(B)|^{2k} = ||B^k||_F^2 = \mathsf{Tr}(B^k(B^k)^T) \approx 2md^m$$

which implies that the *bulk* of the eigenvalues belong to the disk of radius \sqrt{d} .

Why is the eigenvector of *B* informative? This applies to SBM.

Let $\xi \in \mathbb{R}^{\vec{E}}$ denote the 2nd eigenvector of B. Let $\xi^* \in \mathbb{R}^{\vec{E}}$ be defined by $\xi_e^* = \sigma(e_2)$, where $e = (e_1, e_2)$ as usual. For each node u, we estimate its label $\sigma(u)$ by $\hat{\sigma}_u = \operatorname{sign}(\sum_{e:e_1=u} \xi_e)$.

To gain some insight, let's proceed with the following *wishful thinking*: Suppose we can apply power method to study the behavior of the eigenvectors. Since ξ^* is orthogonal to the all-one vector, the 1st eigenvector of B in the population case, let's hope we can gain some insight about the 2nd eigenvector ξ by studying $B^k \xi^*$ for some large k.¹ Then for each node u,

$$\sum_{e:e_1=u} (B^k \xi^*)_e = \sum_{e:e_1=u} \sum_f (B^k)_{ef} \xi_f^*$$

$$= \underbrace{\sum_{v:\sigma(v)=+} \sum_{e:e_1=u} \sum_{f:f_2=v} (B^k)_{ef}}_{\# \text{ of } k\text{th-gen children of type} + \triangleq Z_k^+} - \underbrace{\sum_{v:\sigma(v)=-} \sum_{e:e_1=u} \sum_{f:f_2=v} (B^k)_{ef}}_{\# \text{ of } k\text{th-gen children of type} - \triangleq Z_k^-}$$

where in the last step follows again from the tree structure of $N_u(k)$.

The celebrated result of Kesten-Stigum [KS66] says that the behavior of this number is governed by the matrix M in (8.8), whose eigenvalues are $\lambda_1 = d$ and $\lambda_2 = s$. If $\lambda_2^2 > \lambda_1$, then $(Z_k^+ - Z_k^-)/\lambda_2^k$ converges to some non-degenerate limit X, where X is correlated with the label of the root. This means that $\sum_{e:e_1=u} (B^k \xi^*)_e$ has non-trivial correlation with u, and correlated recovery can be achieved by majority vote.

Nevertheless, the above plan is too simplistic as B is asymmetric so straightforward power method does not work. In reality, to apply the power method properly, one needs to study the SVD of B by considering $B^k(B^k)^{\top}$. But as opposed to the above calculation for B^k which only involves the number of children at the kth generation, the same calculation with $B^k(B^k)^{\top}$ will involve the number of children of all generations up to k.² For details, see [BLM18, Sec 8].

References

- [BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.
 - [BLM18] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular ramanujan graphs. *The Annals* of Probability, 46(1):1–71, 2018.
 - [CO10] A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. Comb. Probab. Comput., 19(2):227–284, 2010.
 - [KS66] Harry Kesten and Bernt P Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. The Annals of Mathematical Statistics, pages 1463–1481, 1966.

¹The rationale of the power method is that $\frac{1}{\|B^k\xi^*\|}B^k\xi^*$ will converge to ξ , but since the matrix B is not symmetric, this does not quite work.

²To see this, note that $B^2(B^2)^{\top}$ will involve paths like u, u_1, u_2, u_1, u_2 , where u_i is a *i*th-gen children

- [KS03] Michael Krivelevich and Benny Sudakov. The largest eigenvalue of sparse random graphs. Combinatorics, Probability and Computing, 12(1):61–72, 2003.
- [Ter10] Audrey Terras. Zeta functions of graphs: a stroll through the garden. Cambridge University Press, 2010.
- [WX18] Yihong Wu and Jiaming Xu. Statistical problems with planted structures: Informationtheoretical and computational limits. In Yonina Eldar and Miguel Rodrigues, editors, *Information-Theoretic Methods in Data Science*. Cambridge University Press, forthcoming, 2018.