Graph Theory

Stochastic block model

Instructor: *Emmanuel Esposito* version of May 26, 2024

Most of the material in this handout is inspired or directly taken from: Roman Vershynin, "High-Dimensional Probability: An Introduction with Applications in Data Science", 2018.

1 Community detection in the stochastic block model

Consider you want to study the structure of a network, which could be a biological network (e.g., protein-protein networks or connectomes) or a social network. A network is naturally modeled by a (undirected) graph $G = (V, E)$. In particular, we are interested in detecting *communities* in networks, that is, clusters in the network that are tightly connected (or dense). Solving the *community detection problem* (or community recovery) efficiently is one of the fundamental problems in network analysis.

Intuitively, the community structure of a network is inherently dependent on the nature of the graph. A natural way to represent it is by assuming that G is generated by some reasonable random model. In the previous lectures we focused our attention on Erdős-Rényi random graphs $\mathcal{G}(n, p)$, where n is the fixed number of vertices and $p \in [0, 1]$ is the probability that any edge appears (independently) in the graph. This random graph model is very simple and elegant, but fails in capturing communities because of its intrinsic homogeneity.

A more effective probabilistic model of graphs that is able to provide a community structure to the graphs is the *stochastic block model* (SBM). We consider the special case of 2 communities over n vertices with equal size $n/2$; we assume n is an even positive integer throughout these notes. Given two probabilities $p, q \in (0, 1]$, the SBM $\mathcal{G}(n, p, q)$ partitions the n vertices into the two communities of size $n/2$ and constructs a random graph $G \sim \mathcal{G}(n, p, q)$ by inserting an edge between any pair of vertices independently with probability p if the vertices belong to the same community, and with probability q if they do not. The partition into the two communities

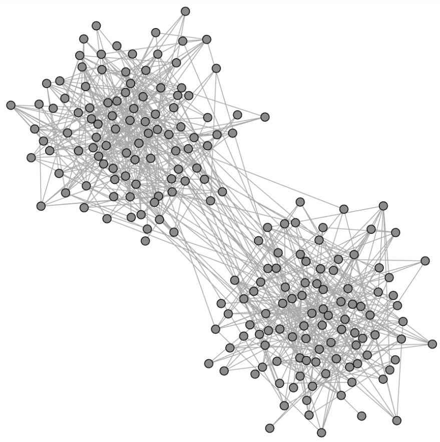

Figure 1: A random graph generated according to the stochastic block model $\mathcal{G}(n, p, q)$ with $n = 200$, $p = 1/20$ and $q = 1/200$. Figure taken from Vershynin [2018].

is, of course, unknown and our objective is trying to recover it. Without loss of generality, we associate vertices V with the first n positive integers $[n] = \{1, \dots, n\}$ so that the first $n/2$ vertices are exactly the ones belonging to one of the two communities, and the remaining $n/2$ belong to the other community.

Observe that when $p = q$ we recover exactly the Erdős-Rényi model $\mathcal{G}(n, p, p) = \mathcal{G}(n, p)$ and thus no community structure will be expected in the random graphs. Therefore, we assume that $q < p$ so that edges will occur more likely within each community rather than across the two.

2 Preliminaries

2.1 Linear algebra

For any vector $\mathbf{x} := (x_1, \dots, x_n) \in \mathbb{R}^n$, we denote by $\|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{\mathbf{x}^\top \mathbf{x}}$ its Euclidean norm and, more generally, by $\|\mathbf{x}\|_p := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ its p-norm for $p \in [1, \infty)$. We say that $\mathbf{x}$ is a unit vector, or normal, if $\|\mathbf{x}\|_2 = 1$. Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are said to be orthogonal, and we write $\mathbf{x} \perp \mathbf{y}$, whenever $\mathbf{x}^\top \mathbf{y} = 0$.

For any symmetric matrix $M \in \mathbb{R}^{n \times n}$ we denote by $\lambda_1(M), \dots, \lambda_n(M) \in \mathbb{R}$ its real eigenvalues so that $\lambda_1(M) \geq \dots \geq \lambda_n(M)$, and by $\mathbf{v}_1(M), \dots, \mathbf{v}_n(M) \in \mathbb{R}^n$ their corresponding orthonormal eigenvectors. We denote by

$$\|M\| := \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\|M\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 = 1} \|M\mathbf{x}\|_2$$

the *spectral norm* (also known as operator norm or induced 2-norm) of M. We remark that $\|M\| = \max_{i \in [n]} |\lambda_i(M)| = \max\{|\lambda_1(M)|, |\lambda_n(M)|\}$.

Exercise 1. Prove that $\|M\| = \max_{i \in [n]} |\lambda_i(M)|$ for any symmetric matrix $M \in \mathbb{R}^{n \times n}$.

2.2 Perturbation theory

Theorem 1 (Courant-Fischer — proof omitted)**.** *Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix with real eigenvalues $\lambda_1(M) \geq \dots \geq \lambda_n(M)$. Then, for any $k \in [n]$,*

$$\lambda_k(M) = \max_{\substack{S \subseteq \mathbb{R}^n : \\ \dim(S) = k}} \min_{\mathbf{x} \in S : \|\mathbf{x}\|_2 = 1} \mathbf{x}^\top M \mathbf{x} = \min_{\substack{S \subseteq \mathbb{R}^n : \\ \dim(S) = n-k+1}} \max_{\mathbf{x} \in S : \|\mathbf{x}\|_2 = 1} \mathbf{x}^\top M \mathbf{x} .$$

The following theorem shows that all eigenvalues of any two symmetric matrices A and B are close to each other whenever the spectral norm $\|A - B\|$ of their difference is small.

Theorem 2 (Weyl's inequality)**.** *Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric matrices with real eigenvalues. Then,*

$$\max_{i \in [n]} |\lambda_i(A) - \lambda_i(B)| \leq \|A - B\| .$$

Proof. Without loss of generality, we show that $\lambda_i(A) - \lambda_i(B) \leq \|A - B\|$ holds for any $i \in [n]$; the other inequality $\lambda_i(A) - \lambda_i(B) \geq -\|A - B\|$ follows by a similar reasoning.

Let $C = A - B$ and fix any $i \in [n]$. Then, we can show that

$$\begin{aligned}
\lambda_i(A) &= \min_{\substack{S \subseteq \mathbb{R}^n: \\ \dim(S)=n-i+1}} \max_{\mathbf{x} \in S: \|\mathbf{x}\|_2=1} \mathbf{x}^\top A \mathbf{x} && \text{by Courant-Fischer} \\
&= \min_{\substack{S \subseteq \mathbb{R}^n: \\ \dim(S)=n-i+1}} \max_{\mathbf{x} \in S: \|\mathbf{x}\|_2=1} \mathbf{x}^\top (B+C) \mathbf{x} && \text{by definition of } C \\
&\leq \min_{\substack{S \subseteq \mathbb{R}^n: \\ \dim(S)=n-i+1}} \left(\max_{\mathbf{x} \in S: \|\mathbf{x}\|_2=1} \mathbf{x}^\top B \mathbf{x} + \max_{\mathbf{x} \in S: \|\mathbf{x}\|_2=1} \mathbf{x}^\top C \mathbf{x} \right) \\
&\leq \min_{\substack{S \subseteq \mathbb{R}^n: \\ \dim(S)=n-i+1}} \left(\max_{\mathbf{x} \in S: \|\mathbf{x}\|_2=1} \mathbf{x}^\top B \mathbf{x} + \max_{\mathbf{x} \in \mathbb{R}^n: \|\mathbf{x}\|_2=1} \mathbf{x}^\top C \mathbf{x} \right) && \text{since } S \subseteq \mathbb{R}^n \\
&= \lambda_1(C) + \min_{\substack{S \subseteq \mathbb{R}^n: \\ \dim(S)=n-i+1}} \max_{\mathbf{x} \in S: \|\mathbf{x}\|_2=1} \mathbf{x}^\top B \mathbf{x} && \text{var. characterization} \\
&= \lambda_1(C) + \lambda_i(B) && \text{by Courant-Fischer} \\
&\leq \|C\| + \lambda_i(B) && \text{since } \lambda_1(C) \leq \|C\| \ .
\end{aligned}$$

By rearranging, we obtain that $\lambda_i(A) - \lambda_i(B) \leq \|C\| = \|A - B\|$ by definition of C. ☐

Exercise 2. Prove the inequality in the opposite direction for any $i \in [n]$, which is currently missing in the proof of Weyl's inequality (Hint: use again Courant-Fischer).

While Weyl's inequality allows us to control the distance between the eigenvalues of two symmetric matrices, it does not allow us to conclude that the eigenvectors will be close to each other too. However, the following theorem actually states that it is indeed the case under some reasonable assumption of separability between eigenvalues (sometimes called eigengap, or spectral gap, assumption).

Theorem 3 (Davis-Kahan — proof omitted)**.** *Let $A, B \in \mathbb{R}^{n \times n}$ be symmetric matrices. Assume that there exists some $i \in [n]$ such that the i-th largest eigenvalue of A is well separated from the other eigenvalues of A:*

$$\Delta_i(A) \coloneqq \min_{j \neq i} |\lambda_i(A) - \lambda_j(A)| > 0 \ .$$

Then, the angle $\angle(\mathbf{v}_i(A), \mathbf{v}_i(B)) \in [0, \pi/2]$ between the eigenvectors $\mathbf{v}_i(A)$ and $\mathbf{v}_i(B)$ corresponding to the i-th largest eigenvalues satisfies

$$\sin \angle(\mathbf{v}_i(A), \mathbf{v}_i(B)) \leq \frac{2\|A - B\|}{\Delta_i(A)} \ .$$

A consequence of Davis-Kahan theorem is that the (unit) eigenvectors $\mathbf{v}_i(A)$ and $\mathbf{v}_i(B)$ are close to each other in Euclidean distance up to a sign whenever the spectral norm $\|A - B\|$ is small. This is stated in the corollary below.

Corollary 4. *Under the assumptions of Davis-Kahan theorem,*

$$\min_{\xi \in \{-1,1\}} \|\mathbf{v}_i(A) - \xi \mathbf{v}_i(B)\|_2 \leq \frac{2\sqrt{2}\|A - B\|}{\Delta_i(A)} \ .$$

Proof. For any $\xi \in \{-1, 1\}$ it holds that

$$\begin{aligned}\|\mathbf{v}_i(A) - \xi \mathbf{v}_i(B)\|_2^2 &= \big(\mathbf{v}_i(A) - \xi \mathbf{v}_i(B)\big)^\top \big(\mathbf{v}_i(A) - \xi \mathbf{v}_i(B)\big) \\ &= \|\mathbf{v}_i(A)\|_2^2 + \xi^2 \|\mathbf{v}_i(B)\|_2^2 - 2\xi \mathbf{v}_i(A)^\top \mathbf{v}_i(B) \\ &= 2 - 2\xi \mathbf{v}_i(A)^\top \mathbf{v}_i(B) \ .\end{aligned}$$

Consequently, by observing that $|\mathbf{v}_i(A)^\top \mathbf{v}_i(B)| \leq \|\mathbf{v}_i(A)\|_2 \|\mathbf{v}_i(B)\|_2 = 1$ by Cauchy-Schwarz, we have that

$$\begin{aligned}\min_{\xi \in \{-1,1\}} \|\mathbf{v}_i(A) - \xi \mathbf{v}_i(B)\|_2^2 &= 2 - 2 \max_{\xi \in \{-1,1\}} \xi \mathbf{v}_i(A)^\top \mathbf{v}_i(B) \\ &= 2 - 2|\mathbf{v}_i(A)^\top \mathbf{v}_i(B)| \\ &\leq 2 - 2\big(\mathbf{v}_i(A)^\top \mathbf{v}_i(B)\big)^2 && \text{since } x^2 \leq x \text{ for } x \in [0,1] \\ &= 2\Big(1 - \cos^2 \angle(\mathbf{v}_i(A), \mathbf{v}_i(B))\Big) && \text{since } \mathbf{x}^\top \mathbf{y} = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \cos \angle(\mathbf{x}, \mathbf{y}) \\ &= 2 \sin^2 \angle(\mathbf{v}_i(A), \mathbf{v}_i(B)) && \text{since } \sin^2 \theta = 1 - \cos^2 \theta \ .\end{aligned}$$

Therefore, by Davis-Kahan and monotonicity of the square root, this implies that

$$\min_{\xi \in \{-1,1\}} \|\mathbf{v}_i(A) - \xi \mathbf{v}_i(B)\|_2 \leq \sqrt{2} \sin \angle(\mathbf{v}_i(A), \mathbf{v}_i(B)) \leq \frac{2\sqrt{2}\|A - B\|}{\Delta_i(A)} \ . \qquad \square$$

2.3 Random matrix theory

In this part, we simply state the matrix concentration bound that will be needed to derive our community recovery result.

Theorem 5 (Matrix Bernstein's inequality)**.** *Let $K \geq 0$ be any constant. Let $X_1, \ldots, X_m \in \mathbb{R}^{n \times n}$ be independent zero-mean symmetric random matrices, such that $\|X_i\| \leq K$ almost surely for all $i \in [m]$. Then, for every $t > 0$, we have*

$$\mathbb{P}\left(\left\|\sum_{i=1}^m X_i\right\| \geq t\right) \leq 2n \exp\Big(-\frac{t^2/2}{\sigma^2 + Kt/3}\Big) \ ,$$

where $\sigma^2 := \big\|\sum_{i=1}^m \mathbb{E}\big[X_i^2\big]\big\|$ is the norm of the matrix variance of the sum.

3 Spectral methods for strong recovery

We finally introduce the recovery problem we want to address in the (symmetric) stochastic block model $\mathcal{G}(n, p, q)$ with 2 balanced communities and we provide an analysis of a simple spectral approach that solves it under reasonable assumptions. The problem in question in that of *strong recovery*, also called almost exact recovery. It consists of recovering the partition of the nodes while in high probability while allowing only a vanishing fraction of misclassified vertices. More formally, an algorithm $\mathcal{A}$ given $G \sim \mathcal{G}(n, p, q)$ is said to perform strong recovery if its accuracy (or its normalized agreement) is approaching 1 with high probability (w.h.p.), that is,

$$\mathbb{P}\big(\text{accuracy of } \mathcal{A}(G) = 1 - o(1)\big) = 1 - o(1) \ .$$

Here the accuracy of the recovery algorithm is measured as the fraction of vertices correctly classified in the partition. Note that the accuracy does not take into account whether the vertices are correctly labeled into community 1 or 2, as what matters here is whether the vertices are split as consistently to the underlying partition as possible.

Recall the definition of *adjacency matrix* A of an undirected graph $G = (V, E)$:

$$A \coloneqq [A_{ij}]_{i,j\in V} \in \{0,1\}^{n\times n} \quad \text{where} \quad A_{ij} \coloneqq \mathbb{I}\{(i,j)\in E\} = \begin{cases} 1 & \text{if } (i,j)\in E \\ 0 & \text{otherwise} \end{cases} .$$

Note that A is symmetric because G is undirected. In our case, the graph $G \sim \mathcal{G}(n,p,q)$ is random and so must be A. More precisely, A is a random matrix whose entries are Bernoulli random variables and those above (or below) the diagonal are independent. Each entry A_{ij} is thus $A_{ij} \sim \mathrm{Ber}(p)$ if i and j belong to the same community, whereas $A_{ij} \sim \mathrm{Ber}(q)$ otherwise.

For simplicity, we assume that also the diagonal entries are Bernoulli random variables $A_{ii} \sim \mathrm{Ber}(p)$, which correspond to self-loops (i,i) being drawn independently with probability p. This is a slight change in the model to simplify the presentation of the results, and it does not negatively affect the stated results.

The expectation of A is thus a block matrix:

$$\bar{A} \coloneqq \mathbb{E}[A] = \left(\begin{array}{ccc|ccc} p & \cdots & p & q & \cdots & q \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p & \cdots & p & q & \cdots & q \\ \hline q & \cdots & q & p & \cdots & p \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ q & \cdots & q & p & \cdots & p \end{array}\right) = \begin{pmatrix} p\mathbf{1}\mathbf{1}^\top & q\mathbf{1}\mathbf{1}^\top \\ q\mathbf{1}\mathbf{1}^\top & p\mathbf{1}\mathbf{1}^\top \end{pmatrix} .$$

By matrix concentration bounds, we can show that the adjacency matrix A is close to its expectation $\bar{A}$ in spectral norm, i.e., $\|A - \bar{A}\|$ is sufficiently small with high probability. Before that, let us observe some properties of the matrix $\bar{A}$:

- The matrix $\bar{A}$ has $\mathrm{rank}(\bar{A}) = 2$ since $p \neq q$, hence it has only two non-zero eigenvalues.
- The eigenvalues of $\bar{A}$ can be derived by finding the roots of its characteristic polynomial

$$\begin{aligned} \det\left(\lambda I - \bar{A}\right) &= \det\begin{pmatrix} \lambda I - p\mathbf{1}\mathbf{1}^\top & q\mathbf{1}\mathbf{1}^\top \\ q\mathbf{1}\mathbf{1}^\top & \lambda I - p\mathbf{1}\mathbf{1}^\top \end{pmatrix} \\ &= \underbrace{\det\left(\lambda I - (p+q)\mathbf{1}\mathbf{1}^\top\right)}_{=:(\star)} \cdot \underbrace{\det\left(\lambda I - (p-q)\mathbf{1}\mathbf{1}^\top\right)}_{=:(\circ)} , \end{aligned}$$

  where the second equality follows from the fact that $\det\begin{pmatrix} A & B \\ B & A \end{pmatrix} = \det(A-B)\det(A+B)$. Note that $(\star)$ is the characteristic polynomial of the $\frac{n}{2}\times\frac{n}{2}$ matrix $(p+q)\mathbf{1}\mathbf{1}^\top$, whereas $(\circ)$ is the one of the $\frac{n}{2}\times\frac{n}{2}$ matrix $(p-q)\mathbf{1}\mathbf{1}^\top$. Both matrices have rank 1 and their only non-zero eigenvalue is, respectively, $\mathrm{tr}\big((p+q)\mathbf{1}\mathbf{1}^\top\big) = \big(\frac{p+q}{2}\big)n$ and $\mathrm{tr}\big((p-q)\mathbf{1}\mathbf{1}^\top\big) = \big(\frac{p-q}{2}\big)n$. Thus, we conclude that $\lambda_1(\bar{A}) = \big(\frac{p+q}{2}\big)n$ and $\lambda_2(\bar{A}) = \big(\frac{p-q}{2}\big)n$, while $\lambda_3(\bar{A}) = \cdots = \lambda_n(\bar{A}) = 0$.

- The orthonormal eigenvectors corresponding to the largest two eigenvalues of $\bar{A}$ are

$$\mathbf{v}_1(\bar{A}) = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \hline 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{v}_2(\bar{A}) = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \hline -1 \\ \vdots \\ -1 \end{pmatrix}$$

and thus $\bar{A} = \frac{p+q}{2}\mathbf{v}_1(\bar{A})\mathbf{v}_1(\bar{A})^\top + \frac{p-q}{2}\mathbf{v}_2(\bar{A})\mathbf{v}_2(\bar{A})^\top = \frac{p+q}{2}\mathbf{1}\mathbf{1}^\top + \frac{p-q}{2}\mathbf{v}_2(\bar{A})\mathbf{v}_2(\bar{A})^\top$.

Now, we can rewrite the matrix A as $A = \bar{A} + W$ where $W := A - \bar{A}$. In other words, we split A into two parts: a deterministic part $\bar{A}$ that denotes the expectation and constitutes the *informative* part (or "signal"), and a random part W which constitutes its *perturbation* (or "noise"). By the above properties we can indeed conclude that $\bar{A}$ is the informative part. It particularly reveals information about the community partition we want to recover through its spectrum. The crucial property we observe is that the *signs* of the entries of the second eigenvector $\mathbf{v}_2(\bar{A})$ exactly distinguish the vertices of the two communities into "positives" and "negatives". However, we cannot directly compute $\mathbf{v}_2(\bar{A})$ as we do not have direct access to the matrix $\bar{A}$ because of the additive noise W. We have instead access to the random adjacency matrix A of the (random) graph G.

Hopefully, A is sufficiently close to $\bar{A}$, say, in spectral norm and we can thus use *perturbation theory* to conclude that the eigenstructure of the two matrices is sufficiently similar to allow us to recover the partition by directly working with A. Moreover, we can prove as a matter of fact that $\|A - \bar{A}\|$ is small enough with high probability by adopting concentration bounds from *random matrix theory.*

We will show the following theorem on the recovery of the two communities in the SBM.

Theorem 6 (Strong recovery)**.** *Let $p := a/n$ and $q := b/n$ for some appropriate $a > b > 0$. Assume that there exist sufficiently large $\alpha > 0$ and $c \gg (1+\alpha)\ln n$ such that*

(i) $\frac{a+b}{2} \geq (1+\alpha)\ln n$,

(ii) $(a-b)^2 \geq c(a+b)$,

(iii) $a \leq 3b$.

Then, with probability $1 - o(1)$, the signs of the coefficients of $\mathbf{v}_2(A)$ recover the communities with accuracy $1 - o(1)$.

We begin by proving that the accuracy of our simple spectral method is controlled under the eigengap assumption.

Lemma 7. *Let A be the adjacency matrix of $G \sim \mathcal{G}(n,p,q)$ and let $\bar{A} = \mathbb{E}[A]$ be its expectation. If the eigengap assumption $\Delta_2(\bar{A}) > 0$ holds, then the accuracy of the algorithm that outputs the community partition according to the signs of the coefficients of $\mathbf{v}_2(A)$ is at least*

$$1 - \frac{2\sqrt{2}\|A - \bar{A}\|}{\Delta_2(\bar{A})} \,.$$

Proof. The number of mistakes for the algorithm that outputs the community partition provided by the signs of the coefficients of $\mathbf{v}_2(A)$ is

$$\min_{\xi\in\{-1,1\}} \sum_{i=1}^n \mathbb{I}\left\{\xi \operatorname{sign}(\mathbf{v}_2(A)_i) \neq \operatorname{sign}(\mathbf{v}_2(\bar{A})_i)\right\} = \min_{\xi\in\{-1,1\}} \sum_{i=1}^n Y_i(\xi) \ ,$$

where we let $Y_i(\xi) \coloneqq \mathbb{I}\left\{\xi \operatorname{sign}(\mathbf{v}_2(A)_i) \neq \operatorname{sign}(\mathbf{v}_2(\bar{A})_i)\right\}$. Observe that whenever $\xi \operatorname{sign}(\mathbf{v}_2(A)_i) \neq \operatorname{sign}(\mathbf{v}_2(\bar{A})_i)$ for some $i \in [n]$, that is, the signs of $\xi\mathbf{v}_2(A)_i$ and $\mathbf{v}_2(\bar{A})_i$ disagree for vertex i and thus $Y_i(\xi) = 1$, we have that

$$Y_i(\xi)\left|\mathbf{v}_2(\bar{A})_i - \xi\mathbf{v}_2(A)_i\right| = Y_i(\xi)\Big(\left|\mathbf{v}_2(\bar{A})_i\right| + |\xi\mathbf{v}_2(A)_i|\Big) \geq Y_i(\xi)\left|\mathbf{v}_2(\bar{A})_i\right| = \frac{Y_i(\xi)}{\sqrt{n}} \ ,$$

and the same inequality holds in the other case of $Y_i(\xi) = 0$. Hence, the number of mistakes is

$$\begin{aligned}
\min_{\xi\in\{-1,1\}} \sum_{i=1}^n Y_i(\xi) &\leq \sqrt{n} \min_{\xi\in\{-1,1\}} \sum_{i=1}^n \left|\mathbf{v}_2(\bar{A})_i - \xi\mathbf{v}_2(A)_i\right| Y_i(\xi) \\
&\leq \sqrt{n} \min_{\xi\in\{-1,1\}} \sum_{i=1}^n \left|\mathbf{v}_2(\bar{A})_i - \xi\mathbf{v}_2(A)_i\right| \\
&= n^{3/2} \min_{\xi\in\{-1,1\}} \sum_{i=1}^n \frac{1}{n}\sqrt{\left(\mathbf{v}_2(\bar{A})_i - \xi\mathbf{v}_2(A)_i\right)^2} \\
&\leq n \min_{\xi\in\{-1,1\}} \sqrt{\sum_{i=1}^n \left(\mathbf{v}_2(\bar{A})_i - \xi\mathbf{v}_2(A)_i\right)^2} \\
&= n \min_{\xi\in\{-1,1\}} \|\mathbf{v}_2(\bar{A}) - \xi\mathbf{v}_2(A)\|_2 \ ,
\end{aligned}$$

where the third inequality follows by Jensen's inequality with respect to the square root. Alternatively, we could have noticed earlier that the sum after the second equality in the above math display corresponds to $\|\mathbf{v}_2(\bar{A}) - \xi\mathbf{v}_2(A)\|_1$ and used the property that $\|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$ for any $\mathbf{x} \in \mathbb{R}^n$. At this point, it immediately follows that

$$\min_{\xi\in\{-1,1\}} \|\mathbf{v}_2(\bar{A}) - \xi\mathbf{v}_2(A)\|_2 \leq \frac{2\sqrt{2}\|A - \bar{A}\|}{\Delta_2(\bar{A})}$$

by Davis-Kahan (Corollary 4), since we can apply it thanks to the spectral gap assumption $\Delta_2(\bar{A}) > 0$. Then, recalling that we define the accuracy as the fraction of correctly classified vertices, by combining the above inequalities we conclude that it is

$$1 - \frac{1}{n} \min_{\xi\in\{-1,1\}} \sum_{i=1}^n Y_i(\xi) \geq 1 - \frac{2\sqrt{2}\|A - \bar{A}\|}{\Delta_2(\bar{A})} \ . \qquad \square$$

Exercise 3. Without explicitly using Jensen's inequality, prove that $\|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$ for any $\mathbf{x} \in \mathbb{R}^n$ (Hint: use the Cauchy-Schwarz inequality).

The next step is showing that the spectral norm of the perturbation matrix is sufficiently small w.h.p. via matrix concentration bounds.

Lemma 8. *Let A be the adjacency matrix of $G \sim \mathcal{G}(n, a/n, b/n)$. Let $\bar{A} = \mathbb{E}[A]$ be its expectation and let $W = A - \bar{A}$ be the zero-mean "perturbation" matrix. Consider any $\delta > 0$ such that $\ln(2n/\delta) \leq a + b$. Then,*

$$\mathbb{P}\left(\|W\| \geq \frac{5}{3}\sqrt{(a+b)\ln\frac{2n}{\delta}}\right) \leq \delta \ .$$

Proof. We want to somehow control the perturbation matrix W. Let us decompose it as follows. For any $1 \leq i \leq j \leq n$, define a binary matrix $Z_{ij} \in \{0,1\}^{n \times n}$ such that its only 1-entries are at coordinates (i,j) and (j,i); in other words,

$$Z_{ij} = \begin{cases} \mathbf{e}_i \mathbf{e}_j^\top + \mathbf{e}_j \mathbf{e}_i^\top & \text{if } i < j \\ \mathbf{e}_i \mathbf{e}_i^\top & \text{if } i = j \end{cases}$$

where $\mathbf{e}_i \in \{0,1\}^n$ is the i-th vector in the canonical basis of $\mathbb{R}^n$, for each $i \in [n]$. These matrices allow us to rewrite W as $W = \sum_{1 \leq i \leq j \leq n} W_{ij} Z_{ij}$, where we recall that $W_{ij} = A_{ij} - \bar{A}_{ij}$. The useful property about this way of writing W is that we expressed it as a sum of independent matrices with mean zero; in fact, $\mathbb{E}[W_{ij} Z_{ij}] = Z_{ij} \mathbb{E}[A_{ij} - \bar{A}_{ij}] = 0$ while the independence comes from the idea of defining Z_{ij} with the purpose of isolating dependent components within each one of those matrices so as to guarantee independence across them.

Since we also have that $\|W_{ij} Z_{ij}\| \leq 1$ holds almost surely, we can apply matrix Bernstein's inequality (Theorem 5) with $K = 1$ to control $\|W\| = \|A - \bar{A}\|$. As a consequence, we obtain for any $t > 0$ that

$$\mathbb{P}\big(\|W\| \geq t\big) \leq 2n \exp\left(-\frac{t^2/2}{\sigma^2 + t/3}\right) ,$$

where $\sigma^2 := \left\| \mathbb{E}\left[\sum_{i \leq j} (W_{ij} Z_{ij})^2\right] \right\|$. Therefore, if we set t as

$$t = \frac{1}{3}\left(\ln \frac{2n}{\delta} + \sqrt{\ln \frac{2n}{\delta}\left(\ln \frac{2n}{\delta} + 18\sigma^2\right)}\right) ,$$

we immediately obtain by explicit calculations that

$$\mathbb{P}\left(\|W\| \geq \frac{1}{3}\left(\ln \frac{2n}{\delta} + \sqrt{\ln \frac{2n}{\delta}\left(\ln \frac{2n}{\delta} + 18\sigma^2\right)}\right)\right) \leq 2n \exp\left(-\frac{t^2}{\sigma^2 + t/3}\right) = \delta .$$

Observe that we can use the inequality $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for any $x, y \geq 0$ to show that

$$\frac{1}{3}\left(\ln \frac{2n}{\delta} + \sqrt{\ln \frac{2n}{\delta}\left(\ln \frac{2n}{\delta} + 18\sigma^2\right)}\right) \leq \sigma \sqrt{2 \ln \frac{2n}{\delta}} + \frac{2}{3} \ln \frac{2n}{\delta} ,$$

hence we obtain that

$$\|W\| < \sigma \sqrt{2 \ln \frac{2n}{\delta}} + \frac{2}{3} \ln \frac{2n}{\delta} \tag{1}$$

holds with probability at least $1 - \delta$.

The next step is explicitly computing σ or, as we will do, an upper bound to it. To do so, we leverage the structure of the matrices derived from the decomposition of W into a sum of "simpler" matrices. First, notice that $Z_{ij}^2 = (\mathbf{e}_i \mathbf{e}_j^\top + \mathbf{e}_j \mathbf{e}_i^\top)^2 = \mathbf{e}_i \mathbf{e}_i^\top + \mathbf{e}_j \mathbf{e}_j^\top$ if $i < j$ while

$Z_{ii}^2 = Z_{ii} = \mathbf{e}_i\mathbf{e}_i^\top$. This implies that

$$\begin{aligned}
\sum_{1\le i<j\le n} \big(W_{ij}Z_{ij}\big)^2 &= \sum_{1\le i<j\le n} W_{ij}^2(\mathbf{e}_i\mathbf{e}_i^\top + \mathbf{e}_j\mathbf{e}_j^\top) \\
&= \sum_{1\le i<j\le n} W_{ij}^2\mathbf{e}_i\mathbf{e}_i^\top + \sum_{1\le i<j\le n} W_{ij}^2\mathbf{e}_j\mathbf{e}_j^\top \\
&= \sum_{1\le i<j\le n} W_{ij}^2\mathbf{e}_i\mathbf{e}_i^\top + \sum_{1\le j<i\le n} W_{ji}^2\mathbf{e}_i\mathbf{e}_i^\top \quad \text{renaming } i \leftrightarrow j \text{ in the second sum} \\
&= \sum_{1\le i<j\le n} W_{ij}^2\mathbf{e}_i\mathbf{e}_i^\top + \sum_{1\le j<i\le n} W_{ij}^2\mathbf{e}_i\mathbf{e}_i^\top \quad \text{since } W_{ij} = W_{ji} \\
&= \sum_{i\ne j} W_{ij}^2\mathbf{e}_i\mathbf{e}_i^\top \ ,
\end{aligned}$$

and therefore

$$\begin{aligned}
\sum_{1\le i\le j\le n} \big(W_{ij}Z_{ij}\big)^2 &= \sum_{i=1}^n \big(W_{ii}Z_{ii}\big)^2 + \sum_{1\le i<j\le n} \big(W_{ij}Z_{ij}\big)^2 \\
&= \sum_{i=1}^n W_{ii}^2\mathbf{e}_i\mathbf{e}_i^\top + \sum_{i\ne j} W_{ij}^2\mathbf{e}_i\mathbf{e}_i^\top \\
&= \sum_{i,j=1}^n W_{ij}^2\mathbf{e}_i\mathbf{e}_i^\top \\
&= \sum_{i=1}^n \underbrace{\left(\sum_{j=1}^n W_{ij}^2\right)}_{=:Z_i} \mathbf{e}_i\mathbf{e}_i^\top \\
&= \operatorname{diag}(Z_1,\dots,Z_n) \ .
\end{aligned}$$

Second, observe that we end up with a diagonal matrix whose entries on the diagonal are $Z_1,\dots,Z_n \ge 0$ and its expectation $\mathbb{E}[\operatorname{diag}(Z_1,\dots,Z_n)] = \operatorname{diag}\big(\mathbb{E}[Z_1],\dots,\mathbb{E}[Z_n]\big)$ is also a diagonal matrix. This allows us to bound σ^2 from above as follows:

$$\begin{aligned}
\sigma^2 &= \big\|\mathbb{E}\big[\operatorname{diag}(Z_1,\dots,Z_n)\big]\big\| \\
&= \max_{i\in[n]} \mathbb{E}[Z_i] && \text{as } \|M\| = \max_{i\in[n]}|\lambda_i(M)| \text{ for symmetric } M \in \mathbb{R}^{n\times n} \\
&= \max_{i\in[n]} \sum_{j=1}^n \mathbb{E}\big[W_{ij}^2\big] && \text{by linearity of expectation} \\
&= \frac{n}{2}\operatorname{Var}(\operatorname{Ber}(p)) + \frac{n}{2}\operatorname{Var}(\operatorname{Ber}(q)) && \text{sum of variances of entries in any row of } A \\
&= \frac{np(1-p)}{2} + \frac{nq(1-q)}{2} && \text{variance of Bernoulli random variables} \\
&\le \Big(\frac{p+q}{2}\Big)n \\
&= \frac{a+b}{2} && \text{by definition of } p \text{ and } q \ .
\end{aligned}$$

Going back to the previous bound on $\|W\|$ provided by Equation (1) and applying our upper bound on σ^2, we see that

$$\|W\| < \sqrt{(a+b)\ln\frac{2n}{\delta}} + \frac{2}{3}\ln\frac{2n}{\delta} \le \frac{5}{3}\sqrt{(a+b)\ln\frac{2n}{\delta}}$$

holds with probability at least $1-\delta$, where the second inequality follows by assumption on δ. $\square$

Finally, we have all the results we need in order to prove our main theorem.

Proof of Theorem 6. First, we observe that the eigengap $\Delta_2(\bar{A})$ is

$$\Delta_2(\bar{A}) = \min\left\{\lambda_1(\bar{A}) - \lambda_2(\bar{A}), \lambda_2(\bar{A})\right\} = n \min\left\{q, \frac{p-q}{2}\right\} = \min\left\{b, \frac{a-b}{2}\right\} = \frac{a-b}{2} ,$$

where the last equality follows by Assumption (iii). Second, since the spectral gap assumption $\Delta_2(\bar{A}) > 0$ is satisfied, we can apply Lemma 7 to obtain that the fraction of misclassified vertices is bounded from above by

$$\frac{2\sqrt{2}\|A-\bar{A}\|}{\Delta_2(\bar{A})} = \frac{4\sqrt{2}\|W\|}{a-b} .$$

Consider now setting $\delta \coloneqq n^{-\alpha}$ and notice that it satisfies

$$\ln(2n/\delta) = \ln(2n) + \alpha \ln n \leq 2(1+\alpha)\ln n , \tag{2}$$

which is not greater than $a+b$ by Assumption (i). Then, we have with probability at least $1-n^{-\alpha} = 1-o(1)$ that

$$\begin{aligned} \frac{4\sqrt{2}\|W\|}{a-b} &< \frac{20\sqrt{2}}{3(a-b)}\sqrt{(a+b)\ln\frac{2n}{\delta}} && \text{by Lemma 8} \\ &\leq \frac{20}{3}\sqrt{\frac{2}{c}\ln\frac{2n}{\delta}} && \text{by Assumption (ii)} \\ &\leq \frac{40}{3}\sqrt{\frac{1+\alpha}{c}\ln n} && \text{by (2)} \end{aligned}$$

where the last inequality follows by Assumption (ii). Therefore, whenever $c \gg (1+\alpha)\ln n$, meaning that $\frac{1}{c} = \frac{1}{1+\alpha} o\big(\frac{1}{\ln n}\big)$, we have that the accuracy is $1-o(1)$ with probability $1-o(1)$. $\square$

References

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018. URL `http://jmlr.org/papers/v18/16-480.html`.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018. URL `https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html`.