Statistical inference on graphs: Selected Topics[*]

Yihong Wu[†] Jiaming Xu[‡]

[*] Lecture notes created for S&DS 684 at Yale University and BA 990 at Duke University. This version: February 21, 2023

[†] Department of Statistics and Data Science, Yale University, New Haven, USA, yihong.wu@yale.edu

[‡] The Fuqua School of Business, Duke University, Durham, USA, jx77@duke.edu.

Part I

Clique problem

§ 1. Introduction & Max Clique in Erdős-Rényi graphs

1.1 Introduction

1.1.1 Basic Definitions

A **graph** $G = (V, E)$ consists of

- A **vertex set** V. Without loss of generality (WLOG), we shall assume $V = [n] \equiv \{1, \ldots, n\}$ for some positive integer n.
- An **edge set** $E \subset \binom{V}{2}$. Each element of E is an edge $e = (i, j)$ (unordered pair). We say i and j are connected and write $i \sim j$ if $(i, j) \in E$.

For the most part, we will be focusing on graphs that are *undirected* (i.e., edges do not have orientation) and *simple* (i.e., no multi-edges or self-loops).

Alternatively, one can also represent a graph as an **adjacency matrix** $A = (A_{ij})_{i,j\in[n]}$, which is an $n \times n$ symmetric binary matrix with zero diagonal. In particular, for a simple and undirected graph $G = (V, E)$, the entries A_{ij} are defined as:

$$A_{ij} = \mathbb{1}\{i \sim j\} = \begin{cases} 1 & (i,j) \in E \\ 0 & \text{o.w.} \end{cases} .$$

Some basic concepts of graphs are defined as follows:

- The **neighborhood** of a given vertex $v \in V$ is defined as $N(v) = \{u \in V : u \sim v\}$, i.e., it is the set of vertices (neighbors) that are connected with v.
- The **degree** of v is defined as $d_v = |N(v)|$, i.e., the number of neighbors of v.
- **Induced subgraph**: For any $S \subset V$, the subgraph induced by S is defined as the graph $G[S] = (S, E_S)$, where $E_S \triangleq \{(u, v) \in E : u, v \in S\}$.
- A **clique** is a **complete** subgraph. A graph is complete iff all pairs of vertices in the graph are connected.

1.1.2 Sample topics

The goal of statistical inference is to using data to make informed decisions (hypotheses testing, estimation, etc). The usual framework of statistical inference is the following:

$$\underbrace{\theta \in \Theta}_{\text{parameter}} \mapsto \underbrace{X}_{\text{data}} \mapsto \underbrace{\widehat{\theta}}_{\text{estimate}} .$$

The theoretical objectives of this class are two-fold:

1. Understand and characterize the fundamental (statistical) limits: What is possible/impossible information-theoretically?
2. Can statistical limits be attained computationally efficiently, e.g., in polynomial time? If yes, how? If not, why?

In this course,

- Data = graphs;
- Parameter = hidden (latent, or planted) structure;
- We will focus on large-graph limit (number of vertices $\to \infty$).

As a preview, we briefly describe two models that we will study below: the **Planted Clique Model** and the **Stochastic Block Model**.

The Planted Clique Model Let V be a vertex set and $n = |V|$, and let $k \le n$ be a given positive integer. The edge set E in a graph $G = (V, E)$ is generated in the following manner:

1. A set S of k vertices is selected out of n vertices to form a clique (all possible edges between them are added to E).
2. Remaining edges are added independently with probability $\frac{1}{2}$.

Given the resulting graph $G = (V, E)$, the goal is to find the planted (hidden) clique S.

To start, notice that this set up follows a classical statistical framework: a sample (here, the graph G) is generated from a distribution (i.e., the random process described above), and we want to estimate a parameter of that distribution (here, the set S) via the sample (here, G).

A decision-theoretic setting is to consider the minimax framework for the worst-case analysis, in which the goal is to find an estimator $\widehat{S} = \widehat{S}(G)$ that correctly recovers S with probability close to 1, regardless of the true set S used to generate the graph G. In other words,

$$\min_{S \in \binom{[n]}{k}} \mathbb{P}_S\left[\widehat{S}(G) = S\right] \approx 1,$$

where $\mathbb{P}_S$ denotes the law of G conditioned on the location of the planted clique S. Alternatively, one can consider the more relaxed Bayesian setting, assuming S is drawn uniformly at random. Equivalently, this amounts to finding an $\widehat{S}$ that preforms well on average:

$$\mathbb{E}_{S \sim \mathrm{Unif}\left(\binom{[n]}{k}\right)} \mathbb{P}_S\left[\widehat{S}(G) = S\right] \approx 1.$$

Remark 1.1. For problems with symmetry, these two formulations are often equivalent, in the sense that

$$\sup_{\widehat{S}} \min_{S \in \binom{[n]}{k}} \mathbb{P}_S\left[\widehat{S}(G) = S\right] = \sup_{\widehat{S}} \mathbb{E}_{S \sim \mathrm{Unif}\left(\binom{[n]}{k}\right)} \mathbb{P}_S\left[\widehat{S}(G) = S\right].$$

This follows from the permutation invariance of the model, which implies the least favorable prior is uniform.

The Stochastic Block Model (SBM) Given a vertex set V, suppose V can be partitioned into two "communities" of equal size. Community membership is represented by a vector

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

where $\sigma_i = \sigma_j$ means that i and j belong to the same community, and $\sum_{i=1}^n \sigma_i = 0$ because the size of the two communities are equal. An edge between two vertices $i, j \in V$ is added to E according to the following probabilities:

$$\mathbb{P}\big[(i,j) \in E\big] = \begin{cases} p & \sigma_i = \sigma_j \\ q & \sigma_i \neq \sigma_j \end{cases},$$

where $0 \leq p, q \leq 1$ (note that p, q need not sum to 1). Thus, in this model, in-group ties and out-group ties have a different probability of forming. There are also several different statistical inference tasks associated with this problem that SBMs address. For example, if p and q are known, then our goal could be to estimate the parameter σ. Or, if p and q are unknown, then we may be interested in jointly estimating $p, q,$ *and* σ.

Note that when $p = q$, the SBM reduces to the classical Erdős-Rényi random graph $G(n, p)$.

1.2 Asymptotic Behavior of Max Clique in $G(n, \frac{1}{2})$

We start with the ensemble of the Erdős-Rényi graph: $G \sim G(n, p)$ is a graph on n vertices where each pair of vertices is connected independently with probability p. Next, as a warmup, we will focus on the behavior of the maximum size of a clique in $G(n, \frac{1}{2})$.

In particular, let $G \sim G(n, \frac{1}{2})$. Define its clique number $\omega(G) \triangleq$ size of the maximum clique in G. We will show that $\omega(G) \approx 2 \log_2 n$ for large n:

Theorem 1.1. *For any $\epsilon > 0$, with high probability (whp) as $n \to \infty$,*

$$\omega(G) \leq (2 + \epsilon) \log_2 n, \tag{1.1}$$

$$\omega(G) \geq (2 - \epsilon) \log_2 n. \tag{1.2}$$

In other words, $\frac{\omega(G)}{\log_2 n} \to 2$ in probability.

The statistical consequence of this computation will come in in the next lecture.

1.2.1 Proof of (1.1): First moment method

Let $\epsilon > 0$ be any given constant. We will show that $\mathbb{P}\big[\omega(G) \geq (2 + \epsilon) \log_2 n\big] \to 0$.

To start, consider any positive integer k, as well as any $S \subset [n]$ where $|S| = k$. Notice that there are $\binom{k}{2}$ possible edges that can form between the k vertices in S, meaning that:

$$P(G[S] \text{ is a } k\text{-clique}) = 2^{-\binom{k}{2}}.$$

And, there are $\binom{n}{k}$ different sets of k vertices in a graph with n vertices. So, by the union bound,

$$\mathbb{P}(\exists S \subset [n] : G[S] \text{ is a } k\text{-clique}) \leq \binom{n}{k} 2^{-\binom{k}{2}}.$$

Now, let $k_0 = (2+\epsilon)\log_2 n$. Again by the union bound, we have that:

$$\begin{aligned}\mathbb{P}(\omega(G) \geq k_0) &\leq \sum_{k=k_0}^{n} \binom{n}{k} 2^{-\binom{k}{2}} \\ &\overset{(a)}{\leq} \sum_{k=k_0}^{n} \left(n 2^{-\frac{(k_0-1)}{2}}\right)^k \\ &\leq \sum_{k=k_0}^{\infty} \left(n 2^{-\frac{(k_0-1)}{2}}\right)^k \overset{(b)}{\leq} 2(n2^{-\frac{(k_0-1)}{2}})^{k_0},\end{aligned}$$

where (a) follows from $\binom{n}{k} \leq n^k$ and $k_0 \leq k$, (b) follows from $n2^{-\frac{k_0-1}{2}} = \sqrt{2}n^{-\epsilon/2} < 1/2$ for sufficiently large n.

1.2.2 Proof of (1.2): Second moment method

We will now show that $\lim_{n\to\infty} \mathbb{P}[\omega(G) \geq k] \to 1$, where $k \triangleq (2-\epsilon)\log_2 n$. Define:

$$T_k \triangleq \#\text{ of } k\text{-cliques in } G = \sum_{|S|=k} \mathbb{1}\left\{G[S] \text{ is a clique}\right\}. \tag{1.3}$$

Note that if a graph contains at least one clique of size k, then the max clique must be of size $\geq k$, implying that $\mathbb{P}[\omega(G) \geq k] \geq \mathbb{P}[T_k > 0]$. So, to show that $\mathbb{P}[\omega(G) \geq k] \to 1$ as $n \to \infty$, it suffices to show instead that $\mathbb{P}[T_k > 0] \to 1$.

Intuition

But, before trying to prove that $\mathbb{P}[T_k > 0] \to 1$, let's first build some intuition. What we computed in the union bound is in fact computing the *first moment* of $\mathbb{E}[T_k]$. By linearity of expectation, we have

$$\mathbb{E}[T_k] = \binom{n}{k} 2^{-\binom{k}{2}}. \tag{1.4}$$

Clearly, when $k = (2+\epsilon)\log_2 n$, $\mathbb{E}[T_k] \ll 0$, which implies that $\mathbb{P}[T_k > 0] \ll 0$ since T_k is integer-valued. As T_k is a positive random variable, it tempting to think that a sufficient condition for $\mathbb{P}[T_k > 0] \gg 0$ is $\mathbb{E}[T_k] \gg 0$. However, this direction is generally false: a counterexample would be a distribution that places almost *all* of its probability mass at zero, and the remaining very *small* amount of probability mass at, say, 10^{100}. Indeed, while the expected value of a random variable with this distribution would be very large, the probability that this random variable is non-zero would still be very small.

So, to show that $\mathbb{P}[T_k > 0]$ is large, it won't be enough to show that $\mathbb{E}[T_k]$ is large. What to do? Well, one way to characterize the distribution in the counterexample above is that it has very *high variance*. If we can show that the variance of T_k is not so large, then that would essentially show that T_k's distribution does not assign low probability to extremely high valued integers, essentially ruling out counterexamples like the one previously entertained. Will this be enough?

Second Moment Method

As it turns out, this approach works and is called the ***Second Moment Method***. Briefly, suppose X_n is a non-negative, integer-valued random variable. In this approach, one shows that $\mathbb{P}[X_n > 0] \to 1$

by showing that:

$$\mathrm{Var}[X_n] = o(\mathbb{E}^2[X_n]),$$

where Var stands for variance. Since we are going to apply the Second Moment Method to show that $\mathbb{P}[T_k > 0] \to 1$, let's first take a small detour to prove it works for the general random variable X_n describe above. And, the first step in doing so will be to prove the Paley-Zygmund inequality.

Lemma 1.1 (Paley-Zygmund Inequality)**.** *Let $X \geq 0$ be a random variable with $0 < \mathbb{E}[X^2] < \infty$. Then for any $0 \leq c \leq 1$,*

$$\mathbb{P}(X > c\mathbb{E}[X]) \geq (1-c)^2 \frac{\mathbb{E}^2[X]}{\mathbb{E}[X^2]} = (1-c)^2 \frac{\mathbb{E}^2[X]}{\mathbb{E}^2[X] + \mathrm{Var}[X]}. \tag{1.5}$$

Proof. First, note that:

$$\mathbb{E}[X] = \mathbb{E}[X \mathbb{1}\{X \leq c\mathbb{E}[X]\}] + \mathbb{E}[X \mathbb{1}\{X > c\mathbb{E}[X]\}] \leq c\mathbb{E}[X] + \mathbb{E}[X \mathbb{1}\{X > c\mathbb{E}[X]\}],$$

meaning that $(1-c)\mathbb{E}[X] \leq \mathbb{E}[X \mathbb{1}\{X > c\mathbb{E}[X]\}]$. Next, note that by Cauchy Swartz:

$$\mathbb{E}[X \mathbb{1}\{X > c\mathbb{E}[X]\}] \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{P}(X > c\mathbb{E}[X])}.$$

Thus:

$$(1-c)^2\mathbb{E}^2[X] \leq \mathbb{E}[X^2]\mathbb{P}(X > c\mathbb{E}[X]),$$

which implies the desired inequality. □

To show that the Second Moment Method works, notice that choosing $c = 0$ in the Paley Zygmund inequality gives us

$$\mathbb{P}(X_n > 0) \geq \frac{\mathbb{E}^2[X_n]}{\mathbb{E}^2[X_n] + \mathrm{Var}[X_n]} = \frac{1}{1 + \frac{\mathrm{Var}[X_n]}{\mathbb{E}^2[X_n]}},$$

so if $\mathrm{Var}[X_n] = o(\mathbb{E}^2[X_n])$, then $\mathbb{P}(X_n > 0) \to 1$, as desired.

Remark 1.2. In addition to Paley-Zygmund Inequality, we may also apply the Chebyshev inequality in the Second Moment Method:

$$\mathbb{P}(X_n > c\mathbb{E}[X_n]) = 1 - \mathbb{P}(X_n \leq c\mathbb{E}[X_n]) \geq 1 - \mathbb{P}\left(|X_n - \mathbb{E}[X_n]| \geq (1-c)\mathbb{E}[X_n]\right) \geq 1 - \frac{\mathrm{Var}[X_n]}{(1-c)^2\mathbb{E}^2[X_n]}.$$

Hence, if $\mathrm{Var}[X_n] = o(\mathbb{E}^2[X_n])$, then $\mathbb{P}(X_n > (1-o(1))\mathbb{E}[X_n]) \to 1$.

The advantage of Paley-Zygmund Inequality over the Chebyshev inequality shows up when $\mathrm{Var}[X_n] = \Theta(\mathbb{E}^2[X_n])$, for which we can still conclude from Paley-Zygmund Inequality that $\mathbb{P}(X_n \geq c\mathbb{E}[X_n]) = \Omega((1-c)^2)$.

Applying the Second Moment Method

We now return to our original goal of showing that $\mathbb{P}[T_k > 0] \to 1$, which we shall prove via the Second Moment Method. In particular, we need to show that $\mathrm{Var}[T_k] = o(\mathbb{E}^2[T_k])$. To start, notice that:

$$\begin{aligned}
\operatorname{Var}[T_k] &= \operatorname{Var}\Bigg[\sum_{|S|=k} \mathbb{1}\left\{G[S] \text{ is a } k \text{ clique}\right\}\Bigg] \\
&= \sum_{\substack{S,S' \\ |S|=|S'|=k}} \operatorname{Cov}\Big[\mathbb{1}\left\{G[S] \text{ is a } k \text{ clique}\right\}, \mathbb{1}\left\{G[S'] \text{ is a } k \text{ clique}\right\}\Big] \\
&\overset{(a)}{=} \sum_{\substack{|S\cap S'|\geq 2 \\ |S|=|S'|=k}} \operatorname{Cov}\Big[\mathbb{1}\left\{G[S] \text{ is a } k \text{ clique}\right\}, \mathbb{1}\left\{G[S'] \text{ is a } k \text{ clique}\right\}\Big] \\
&\leq \sum_{\substack{|S\cap S'|\geq 2 \\ |S|=|S'|=k}} \mathbb{P}\Big[\text{ both } G[S] \text{ and } G[S'] \text{ are } k \text{ cliques}\Big],
\end{aligned}$$

where (a) follows from the fact that, for any two vertex sets S and S'. If $|S \cap S'| \leq 1$ (at most one node shared between S and S'), then the set of edges formed among nodes in S are *disjoint* from the set of edges formed among nodes in S'. Thus, by independence, the covariance is zero.

Now, for any given pair of sets S,S', let $\ell = |S \cap S'|$. In order for S and S' to *both* be k-cliques, there are a total of $2\binom{k}{2} - \binom{l}{2}$ possible edges that must be formed (think: inclusion-exclusion principle), so we have

$$\begin{aligned}
\operatorname{Var}[T_k] &\leq \sum_{\ell=2}^{k} \Big|\Big\{(S,S') : |S| = |S'| = k, |S \cap S'| = \ell\Big\}\Big| \cdot 2^{-2\binom{k}{2}+\binom{\ell}{2}} &(1.6)\\
&= \sum_{\ell=2}^{k} \binom{n}{k}\binom{k}{\ell}\binom{n-k}{k-\ell} \cdot 2^{-2\binom{k}{2}+\binom{\ell}{2}}, &(1.7)
\end{aligned}$$

where the last step follows from the following reasoning: there are $\binom{n}{k}$ ways of picking a set S of k vertices from a graph on n vertices. And, for each such set S, there are exactly $\binom{k}{\ell}$ ways to pick ℓ nodes from S that will also be part of another set S'. Once S and the nodes of S that will be shared with S' have been determined, it remains to pick from S^c the remaining $k - \ell$ nodes of S', and there are exactly $\binom{n-k}{k-\ell}$ ways of doing that.

At this point, one can analyze the above sum by brute force, focusing on the exponent of each term. Next we present a more "statistician's approach". Note that the counting step in (1.6) is precisely how hypergeometric distribution (sampling without replacement) arises. Indeed, if we have an urn of n balls among which k balls are red, let H denote the number of red balls if we draw k balls from the urn uniformly at random without replacements. Then $H \sim \text{Hypergeometric}(n, k, k)$. Thus, we can express the same quantity in terms of H as follows:

$$\begin{aligned}
\frac{\operatorname{Var}[T_k]}{\mathbb{E}^2[T_k]} \leq \sum_{\ell=2}^{k} \frac{\binom{k}{\ell}\binom{n-k}{k-\ell}}{\binom{n}{k}} \cdot 2^{\binom{\ell}{2}} \leq \sum_{\ell=2}^{k} \frac{\binom{k}{\ell}\binom{n-k}{k-\ell}}{\binom{n}{k}} \cdot 2^{\ell k/2} \qquad (1.8)\\
= \mathbb{E}[2^{kH/2}\mathbb{1}\left\{H \geq 2\right\}] \leq \mathbb{E}[2^{kH/2}] - \mathbb{P}[H = 0].
\end{aligned}$$

Next we will show that both $\mathbb{P}[H = 0] \to 1$ and $\mathbb{E}[2^{kH/2}] \to 1$. Indeed,

$$\mathbb{P}[H = 0] = \frac{\binom{n-k}{k}}{\binom{n}{k}} = \left(1 - \frac{k}{n}\right)\left(1 - \frac{k}{n-1}\right)\cdots\left(1 - \frac{k}{n-k+1}\right) \to 1,$$

since $k = (2-\epsilon)\log_2 n = o(\sqrt{n})$.

To bound the generating function, we use the comparison between sampling with replacements (binomial) and sampling without replacements (hypergeometric). The following result of Hoeffding (proved in the homework) will be useful in several places in this course:

Lemma 1.2 (Hoeffding's lemma)**.** Binom$(k, \frac{k}{n})$ *dominates* Hypergeometric(n, k, k) *in the order of convex functions. In other words, if* $B \sim$ *Binomial*$(k, \frac{k}{n})$*, then* $\mathbb{E}[f(H)] \le \mathbb{E}[f(B)]$ *for all convex functions* f*.*

Using this lemma, we have

$$\mathbb{E}[2^{kH/2}] \le \mathbb{E}[2^{kB/2}] = \left(1 + \frac{k}{n}\left(2^{\frac{k}{2}} - 1\right)\right)^k \le \exp\left(\frac{k^2}{n}\left(2^{\frac{k}{2}} - 1\right)\right) \to 1.$$

since $k^2 2^{\frac{k}{2}} \ll n$ by the assumption that $k = (2-\epsilon)\log_2 n$.

To summarize, we have shown that $\dfrac{\mathrm{Var}[T_k]}{\mathbb{E}^2[T_k]} \to 0$. By Paley-Zygmund (Lemma 1.1), it follows that $\mathbb{P}[T_k > 0] \to 1$, i.e., $\mathbb{P}[\omega(G) \ge (2-\epsilon)\log_2 n] \to 1$, so we've proven the desiderata.

Remark 1.3. Note that in computing the second moment, (1.8) can be equivalently written as

$$\frac{\mathrm{Var}[T_k]}{\mathbb{E}^2[T_k]} \le \mathbb{E}[2^{k|S\cap S'|/2}\mathbb{1}\left\{|S \cap S'| \ge 2\right\}],$$

where S and S' are independent random k-sets drawn uniformly. This is something we will frequently encounter in computing the second moment, which typically involves *two independent copies* of the same randomness and their *overlap* $|S \cap S'|$.

Remark 1.4. As a small aside, we can further show that not only there exists a clique of size $k = (2-\epsilon)\log_2 n$, there are an *abundance* of them. Indeed, by (1.4) and using $\binom{n}{k} \ge (\frac{n}{k})^k$, we have

$$\mathbb{E}[T_k] = \binom{n}{k}2^{-\binom{k}{2}} \ge \left(\frac{n}{k}\right)^k 2^{-\binom{k}{2}} = n^{\Omega(\log n)} \to \infty.$$

By Lemma 1.1, we have $T_k > o(\mathbb{E}[T_k])$ with probability $1 - o(1)$ (in fact, using the Chebyshev inequality, we can conclude that $T_k \ge (1 - o(1))\mathbb{E}[T_k]$ with high probability). This shows that there exists superpolynomially many cliques of size $(2-\epsilon)\log_2 n$. Unfortunately, the best polynomial-time algorithm can only guarantee to find a clique of size $(1-\epsilon)\log_2 n$ with high probability. We will discuss this next time.

1.3 Grimmett-McDiarmid's greedy algorithm to find cliques of size $(1-\epsilon)\log_2 n$

So far we have shown the following. Let $\omega(G(n, \frac{1}{2}))$ denotes the maximum size of cliques in $G(n, \frac{1}{2})$ which is a random variable. Recall that $\omega(G(n, \frac{1}{2}))$ concentrates around $2\log_2 n$ as shown in the previous lecture. In fact, not only there exists a clique of size $(2-\epsilon)\log_2 n$, there exist an abundance of them. The reason is that

$$\mathbb{E}[\#\text{ of cliques of size } k] = \binom{n}{k}2^{\binom{k}{2}} \to \begin{cases} 0 & \text{if } k = (2+\epsilon)\log_2 n \\ +\infty & \text{if } k = (2-\epsilon)\log_2 n \end{cases}.$$

Now that the statistical aspect of the problem has been understood, what about the computational aspect? The complexity of the exhaustive search is

$$\binom{n}{\log_2 n} \approx n^{\log n}$$

and grows superpolynomially in n. What if we limit ourselves to "efficient" algorithms that run in time polynomial in the size of the graph, say, n^C for some constant C. In the following section, we are going to present a greedy algorithm that runs in polynomial time (in fact, sublinear time) and is able to find cliques of size $(1-\epsilon)\log_2 n$ (a factor-of-two approximation of the maximum clique). In contrast, for the *max clique* problem (finding the maximum clique in a given graph or deciding whether a clique of a given size exists) in the worst case is impossible to approximate even within a factor of $n^{1-\epsilon}$, unless P=NP [Hås99]. This shows the drastic difference between worst-case analsys and average-case analysis, due to the atypicality of the hard instances. Nevertheless, for $G(n, \frac{1}{2})$, it remains open whether there exists an efficient algorithm that finds a clique of size bigger than this threshold, say, $1.01\log_2 n$.

Before we present a greedy algorithm that provably works, let us start with another greedy algorithm which is intuitive but might be difficult to analyze.

Algorithm 1: Greedy algorithm I

```
Start from an arbitrary vertex
Given a clique, repeat:
    Add a vertex randomly from the common neighbors of the existing clique
    If there is no common neighbors, stop and return the clique
```

The justification to this algorithm is the following: given that we have found an m-clique, for a given vertex v outside, the probability that v is connected to all m vertices in the clique is, assuming that each edge happens with probability $\frac{1}{2}$ independently,

$$\mathbb{P}(v \text{ is connected to all } m \text{ vertices}) = 2^{-m}.$$

Therefore, the probability that there exists a v that is connected to all m vertices in the existing clique is

$$\mathbb{P}(\exists v \text{ connected to all } m) = 1-(1-2^{-m})^{n-m} \to 1 \quad \text{if } 2^{-m} \ll 1/n, \text{ e.g., } m = (1-\epsilon)\log_2 n.$$

However, the reasoning is flawed because given the information that we have an existing clique of size m, the probabilities of v connected to each of them is no longer $\frac{1}{2}$. Furthermore, the edges are not conditionally independent either. Therefore it is unclear how to analyze Algorithm 1.

Next we present a variant of the greedy algorithm, due to Grimmett-McDiarmid [GM75],[1] which is easily analyzable.

Theorem 1.2. *Fix any $\epsilon > 0$. With probability tending to one as $n \to \infty$, the output of Algorithm 2 is a clique of size at least $(1-\epsilon)\log_2 n$.*

[1] The original paper [GM75] deals with the number of vertex coloring (so that no adjacent vertices are colored the same) and independent set (subset of vertices that induce an empty graph). Note that cliques in the original graphs correspond to independent sets in the complementary graph, and the complement of $G(n, \frac{1}{2})$ is still $G(n, \frac{1}{2})$.

Algorithm 2: Greedy algorithm II

```
Label the vertices $v_1, \dots, v_n$ arbitrarily
for $t = 1$ to $n$ do
    Given the current clique, if $v_t$ is connected to all vertices in the current clique, add $v_t$ to the clique.
end for
```

Proof. It is obvious that the output of the above algorithm, denoted as S_n, is a clique. It remains to show that with high probability, the size of the clique is at least $(1-\epsilon)\log_2 n$.

Let T_i be the time for the size of the clique to grow from $i-1$ to i. By the design of the algorithm we can see that T_i's are independent and geometrically distributed as[2]

$$T_i \overset{ind}{\sim} \text{Geo}(2^{-(i-1)}) \Rightarrow \mathbb{E}T_i = 2^{i-1}.$$

Therefore,

$$\begin{aligned}
\mathbb{P}(|S_n| \geq k) &= \mathbb{P}(T_1 + T_2 + \cdots + T_k \leq n) \\
&\geq \prod_{i=1}^{k} \mathbb{P}\left(T_i \leq \frac{n}{k}\right) \\
&= \prod_{i=1}^{k} \left(1 - (1 - 2^{-(i-1)})^{n/k}\right) \\
&\geq \left(1 - (1 - 2^{-k})^{n/k}\right)^k \\
&\geq 1 - k(1 - 2^{-k})^{n/k} \\
&\to 1 \quad \text{if } k = (1-\epsilon)\log_2 n.
\end{aligned}$$

□

Remark 1.5. The time complexity of Algorithm 2 is, in expectation,

$$\sum_{i=1}^{\log_2 n} 2^{i-1} \times (i-1) = O(n \log^2 n).$$

This is sublinear in the size of the graph ($\Theta(n^2)$ edges).

[2] This is reminiscent of the coupon collector problem, where it becomes increasingly more difficult to collect the last few uncollected coupons, although here the situation is more drastic.

§ 2. The Planted Clique model and (iterated) degree tests

2.1 Planted Clique model and statistical limits

The *planted clique* model is a random graph model which can be described as follows: First choose a subset K of size k uniformly at random from all n vertices and form a clique. The remaining vertex pairs are connected independently with probability $\frac{1}{2}$. In other words,

$$\forall i, j \quad \mathbb{P}\left[i \sim j\right] = \begin{cases} 1 & \text{if both } i, j \in K \\ \frac{1}{2} & \text{otherwise} \end{cases}.$$

Denote the resulting graph $G \sim G(n, \frac{1}{2}, k)$. (Note that $G(n, \frac{1}{2}, 0)$ is the usual Erdős-Rényi graph $G(n, \frac{1}{2})$.) As mentioned in the Introduction, there are two types of questions one can ask in the Planted Clique model:

- Detection: Testing
$$H_0 : G \sim G(n, \frac{1}{2}), \quad \text{versus} \quad H_1 : G \sim G(n, \frac{1}{2}, k). \tag{2.1}$$

- Recovery: Given $G \sim G(n, \frac{1}{2}, k)$, recover the planted clique, say, exactly; namely, find an estimator $\widehat{K} = \widehat{K}(G) \subset [n]$, such that $\widehat{K} = K$ whp.

It turns out that the statistical limits of both questions are easy to resolve, which we will first get out of the way; it is the algorithmic question that will be our focus for the forthcoming lectures.

Theorem 2.1 (Detecting the planted clique: statistical limit)**.** *Let $\epsilon > 0$ be an arbitrary small constant. Consider the hypothesis testing problem (2.1).*

- *Let $k \geq (2 + \epsilon) \log_2 n$. Let $G \sim G(n, \frac{1}{2}, k)$. Then the test $\mathbb{1}\left\{\omega(G) \geq (2 + \epsilon) \log_2 n\right\}$ succeeds whp.*

- *Let $k \leq (2 - \epsilon) \log_2 n$. Then it is impossible to detect the planted clique, in the sense that*
$$\min_{\phi(\cdot) \in \{0,1\}} \mathbb{P}_{H_0}[\phi(G) = 1] + \mathbb{P}_{H_1}[\phi(G) = 0] = 1 - o(1). \tag{2.2}$$

Theorem 2.2 (Recovering the planted clique: statistical limit)**.** *Let $\epsilon > 0$ be an arbitrary small constant. Let $G \sim G(n, \frac{1}{2}, k)$ with K being the planted k-clique.*

- *Let $k \geq (2 + \epsilon) \log_2 n$. Then whp, K is is the unique largest k-clique.*

- *Let $k \leq (2 - \epsilon) \log_2 n$. Then it is impossible to find the planted clique, in the sense that*
$$\max_{\widehat{K}(\cdot)} \mathbb{P}[\widehat{K}(G) = K] = o(1). \tag{2.3}$$

Remark 2.1 (Interpretation of statistical limits)**.**

Proof of Theorem 2.1. For $k \geq (2+\epsilon)\log_2 n$, the positive direction follows from that of Theorem 2.2 (to be shown next) and Theorem 1.1 which shows that $\omega(G) \leq (2+\epsilon)\log_2 n$ whp under the null model.

Next we prove the impossibility side for $k \geq (2-\epsilon)\log_2 n$. For this we need the basics of hypothesis testing from Appendix A. Recall that the minimum probability of error on the left side of (2.2) is given by the total variation $1 - \mathrm{TV}(P,Q)$, where Q and P are the distribution of the observation G under the null and alternative hypothesis, namely,

$$Q(G) = 2^{-\binom{n}{2}}$$

and

$$P(G) = \frac{1}{\binom{n}{k}} 2^{-\binom{n}{2}+\binom{k}{2}} \sum_{|S|=k} \mathbb{1}\left\{G[S] \text{ is a clique}\right\}.$$

In order to show $\mathrm{TV}(P,Q) = o(1)$, by Lemma A.2 it suffices to show $\chi^2(P\|Q) = o(1)$. The likelihood ratio is given by

$$\frac{P(G)}{Q(G)} = \frac{2^{-\binom{k}{2}}}{\binom{n}{k}} T_k = \frac{T_k}{\mathbb{E}_Q[T_k]},$$

where $T_k = T_k(G)$ is the number of k-cliques in G previously introduced in (1.3). There, in (1.8), we have shown that $\mathrm{Var}_Q(T_k) = o((\mathbb{E}_Q[T_k])^2)$. This is exactly what we need since $\chi^2(P\|Q) = \mathrm{Var}_Q(\frac{P}{Q}(G)) = \frac{\mathrm{Var}_Q(T_k)}{(\mathbb{E}_Q[T_k])^2} = o(1)$. □

Proof of Theorem 2.2. The proof of the positive direction is an exercise in first moment calculation. (Homework 1.) Note that indeed it is not implied by Theorem 1.1 as one needs to rule of the possibility of bigger cliques formed by planted edges plus non-planted edges.

The impossibility for recovery is not implied *a priori* by that of detection proved in Theorem 2.1. For this, let us examine the posterior distribution of the planted clique given the observed graph G. Note that

$$P(G|K=S) = \mathbb{1}\left\{G[S] \text{ is a } k\text{-clique}\right\} 2^{-\binom{n}{2}+\binom{k}{2}}.$$

Since K has a uniform prior, the posterior is simply the uniform distribution of on all k-cliques of G, i.e.,

$$P(K=S|G) = \frac{1}{T_k} \mathbb{1}\left\{G[S] \text{ is a } k\text{-clique}\right\},$$

where $T_k = T_k(G)$ is the total number of k-cliques in G; cf. (1.3). Note that by definition, $T_k \geq 1$. Fix any estimator $\widehat{K} = \widehat{K}(G)$. Then its probability of success is at most

$$\mathbb{P}\left[\widehat{K}(G) = K\right] = \mathbb{E}\left[\mathbb{P}\left[\widehat{K}(G) = K \middle| G\right]\right] \leq \mathbb{E}\left[\frac{1}{T_k}\right] \leq \mathbb{P}\left[T_k \leq t\right] + \frac{1}{t},$$

where the first inequality follows from the uniformity of the posterior. Note that $T_k(G) \geq T_k(G[K^c])$, where $G[K^c] \sim G(n-k, \frac{1}{2}$. For all sufficiently large n, $k \leq (2-\epsilon)\log_2 n \leq (2-\epsilon/2)\log_2(n-k)$. So, as shown in Remark 1.4, $T_k(G[K^c]) \geq n^{\Omega(\log n)} \equiv t$ whp, completing the proof. □

2.2 Overview of algorithmic approaches

Ignoring the computation cost, we can use exhaustive search to recover the planted clique with high probability if $k \geq (2+\epsilon)\log_2 n$, because the maximum clique in $G(n, \frac{1}{2})$ is approximately $2\log_2 n$, as shown in Section 1.2. For the same reason, if $k \leq (2-\epsilon)\log_2 n$, recovering the planted clique is information-theoretically impossible, as there is an abundance of cliques of such size in $G(n, \frac{1}{2})$ (cf. Remark 1.4). However, what if we only consider efficient algorithms? It turns out the state of the art can only find planted cliques of size $k = \Omega(\sqrt{n})$.

In the following sections, we will discuss a number of efficient algorithms that is able to recover the planted clique with high probability if $k = \Theta(\sqrt{n})$.

- **Degree test** method, which works for $k = \Omega(\sqrt{n \log n})$. We will discuss an iterative version that works for $k = C\sqrt{n}$ for some $C > 0$, that runs in linear time.[1]

- **Spectral method**, which works for $k = C\sqrt{n}$ for some C. This can be improved to arbitrarily small C at the price of time complexity $n^{O(1/C^2)}$.

- **Semi-definite programming** approach, which also work for $k = C\sqrt{n}$. The added advantage is robustness, which the previous methods lack.

Remark 2.2. We see that the gap between the information-theoretical limit and what computationally efficient algorithms can achieve is significantly larger for the planted clique problem ($\log_2 n$ versus $\sqrt{n}$) than the counterpart for the maximum clique problem in $G(n, \frac{1}{2})$ ($2\log_2 n$ versus $\log_2 n$).

Exercise (Slightly smarter exhaustive search). To find the hidden k-clique in $G(n, \frac{1}{2}, k)$, exhaustive search takes $\binom{n}{k} \sim n^k$ time. Here is an $n^{\Theta(\log n)}$-time algorithm for all $k \geq C\log_2 n$ for a sufficiently large constant C.

1. By exhaustive search we can find a clique T of size $C\log_2 n$. Then it holds that $|T \cap K| \geq (C-2-\epsilon)\log_2 n$ whp. (Why?)

2. Let S denote the set of all vertices that has at least $3|T|/4$ neighbors in T. Then one can show that $K \subset S$ with high probability.

3. Now S might also contain some non-clique vertices, which requires some cleanup. So we report the k highest-degree vertices in the induced subgraph $G[S]$. Hint: show $|S \setminus K|$ is relatively small. Be careful with the union bound as T is random.

2.3 Degree Test

By degree test we meant declaring the k vertices with the largest degrees as the estimated clique. The motivation is that the vertices in the clique tend to have a higher degree than those outside. To analyze this method, let d_i denote the degree of a vertex i. If $i \notin K$, then

$$d_i \sim \text{Binom}\left(n-1, \frac{1}{2}\right), \quad \mathbb{E}d_i \approx \frac{n}{2}.$$

If $i \in K$, then

$$d_i \sim k - 1 + \text{Binom}\left(n-k, \frac{1}{2}\right), \quad \mathbb{E}d_i \approx \frac{n+k}{2}.$$

[1] Since the graph is dense, here linear time means $O(n^2)$.

Therefore, the separation in mean is proporational to k. Usually, we need the separation to be at least as large as the standard deviation. Therefore, $k = \Omega(\sqrt{n})$ is clearly necessary. Furthermore, for the degree test method to work, we need an additional $\sqrt{\log n}$ factor to accommodate for the possibility that some of the $n-k$ non-clique vertices will have an atypically high degree, and some of the k vertices in the clique will have an atypically low degree. This will be carried out by an union bound, as we shall see later. In fact, although the degrees are not mutually independent (d_i and d_j are positively correlated through $\mathbb{1}\{i \sim j\}$), they are almost independent so $\sqrt{\log n}$ factor is necessary in order for the degrees to fully separate.

Formally, the degree-based estimator is just

$$\widehat{K} = \text{ set of } k \text{ vertices with the highest degree }. \tag{2.4}$$

We will show that with high probability, the degree based estimator can recover the true planted clique of size $k = \Omega(\sqrt{n \log n})$. This simple observation is usually attributed to [Kuč95].

Theorem 2.3. *$\mathbb{P}(\widehat{K} = K) \to 1$, provided $k \geq C\sqrt{n \log n}$ for some absolute constant $C > 0$.*

It is obvious from the definition of the degree-based estimator (2.4) that a sufficient condition for $\widehat{K} = K$ is that

$$\min_{i \in K} d_i > \max_{i \notin K} d_i. \tag{2.5}$$

Before we prove Theorem 2.3, we first review some basic facts about Gaussian approximation and concentration inequalities.

Lemma 2.1. *Suppose that $X_1, \ldots, X_n \sim N(0,1)$. Then for any fixed $\epsilon > 0$,*

$$X_{\max} = \max_{i \in 1, \ldots, n} X_i \leq \sqrt{(2+\epsilon)\log n} \quad w.h.p.$$

Proof.

$$\mathbb{P}(X_{\max} > t) \leq n\mathbb{P}(X_1 > t) \leq ne^{-t^2/2} \to 0 \text{ if } t > \sqrt{(2+\epsilon)\log n}.$$

□

Note that Lemma 2.1 is tight if $X_1, \ldots, X_n$ are independent.

Using the heuristic of approximating binomials by Gaussians, we can analyze the degree test as follows: If $i \notin K$, then

$$d_i \sim \text{Binom}\left(n-1, \frac{1}{2}\right) \approx N(n/2, n/4)$$

Using Lemma 2.1, the right hand side of Equation (2.5) is approximately

$$\max_{i \notin K} d_i \leq \sqrt{2\log(n-k)\frac{n}{4}} + \frac{n}{2} \approx \frac{n}{2} + \frac{1}{2}\sqrt{2n \log n}.$$

Similarly, since for $i \in K$ we have $d_i \sim k + \text{Binom}(n-k, \frac{1}{2})$,

$$\min_{i \in K} d_i \geq k + \frac{n-k}{2} - \sqrt{2 \log k \frac{n-k}{4}} \approx \frac{n+k}{2} - \frac{1}{2}\sqrt{2n \log k}.$$

Therefore (2.5) holds with high probability if $k \geq \sqrt{Cn \log n}$ for some large constant C.

To justify the above Gaussian intuition, we use Hoeffding's inequality, one of the basic concentration inequalities. We will prove it in Lecture 4 (see Lemma 4.4).

Lemma 2.2 (Hoeffding's inequality)**.** *Let $S = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are independent and $a \le X_i \le b, \forall i$. Then*

$$\mathbb{P}(|S - ES| \ge t) \le 2\exp\left(-\frac{2t^2}{n(b-a)^2}\right).$$

If we apply the Hoeffding inequality to binomial distribution, we get the following corollary:

Corollary 2.1. *If $S \sim \text{Binom}(n, p)$,*

$$\mathbb{P}(|S - np| \ge t) \le 2\exp\left(-\frac{2t^2}{n}\right).$$

Two remarks are in order.

Remark 2.3. In the large-deviation regime where $t = \Theta(n)$, we have

$$\mathbb{P}(|S - np| \ge t) = \exp\left(-\Theta(n)\right).$$

In the moderate-deviation regime where $O(\sqrt{n}) \le t = o(n)$, e.g., if $t = n^{\frac{1}{2}+\epsilon}$, then we have

$$\mathbb{P}(|S - np| \ge t) = \exp\left(-\Theta(n^{2\epsilon})\right).$$

When $t = o(\sqrt{n})$, the inequality becomes meaningless.

Remark 2.4. The bound is good if p is a constant like $\frac{1}{2}$. If $p = o(1)$, i.e., for *sparse* graphs, then the variance is $np \ll n$ and we should aim for tail bound like $\exp(-\frac{2t^2}{np})$ instead if $p \gg \frac{1}{n}$. However, Hoeffding's inequality does not capture this.

Using Hoeffling's inequality, we obtain the following lemma about the degree test:

Lemma 2.3. *Suppose $G \sim G(n, \frac{1}{2}, k)$ with the planted clique K. Let $\widehat{K}$ be the set of the k highest degree vertices in G. Then,*

$$\mathbb{P}\left[\widehat{K} = K\right] \ge 1 - 2n\exp\left(-\frac{k^2}{8n}\right).$$

Therefore, if $k = \Omega(\sqrt{n \log n})$, then $\widehat{K} = K$ with high probability.

For the next algorithm, we need the following lemma that quantifies the convergence rate in the central limit theorem, in terms of the Kolmogorov distance, that is, the sup norm between CDFs; this result is due to Berry-Esseen. We state without proof the version for the normal approximation of binomials (see, e.g., [Dur10, Theorem 3.4.9]).

Lemma 2.4 (Berry-Esseen theorem)**.** *For all p, n,*

$$\sup_{x \in \mathbb{R}} \left|\mathbb{P}\left(\text{Binom}(n,p) \le x\right) - \mathbb{P}\left(N(np, np(1-p)) \le x\right)\right| \le \frac{2}{\sqrt{np(1-p)}}.$$

2.4 Iterating the degree test

The degree test works for $k = \Omega(\sqrt{n \log n})$. Now we will present an iterative algorithm by Dekel–Gurel-Gurevich–Peres [DGGP11] that is able to find K in $O(n^2)$ times when $k = |K| = C\sqrt{n}$ for some sufficiently large $C > 0$.

There are essentially two ideas:

1. The first idea is this: if we define the "relative size" as $\frac{k^2}{n}$, then we know from Lemma 2.3 that it needs to exceed $\log n$ in order for the degree test to succeed. If we can subsample the graph cleverly, then we might be able to gradually increase the relative size until it reaches this threshold. However, blindly subsample each vertex independently with probability τ clearly does not work, as $n \to n\tau$ and $k \to k\tau$ and this will only decrease the relative size. Instead, we are going to subsample in such a way that prefers clique vertices (e.g., based on degrees!), so that $n \to \tau n$ and $k \to \rho k$ and, upon choosing the parameters appropriately, $\rho > \sqrt{\tau}$. This way, the relative size will grow by a constant factor in each iteration, and it takes $\log \log n$ rounds to pass the $\log n$ threshold.

   Specifically, we will generate a sequence of graphs $G = G_0 \supset G_1 \supset \cdots \supset G_T$, so that each G_t is an instance of the planted clique model $G(n_t, \frac{1}{2}, k_t)$, with
   $$n_t \approx n'_t \triangleq \tau^t n, \quad k_t \approx k'_t \triangleq \rho^t k \tag{2.6}$$
   where
   $$\tau = (1-\alpha)Q(\beta), \quad \rho = (1-\alpha)Q(\beta - C\sqrt{\alpha}), \tag{2.7}$$
   and $Q(t) \triangleq \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ denotes the complementary CDF of the standard normal distribution.

2. The catch with the subsampling idea is that we ended up with recovering just a subset of the hidden clique. Nevertheless, provided this *seed* set is not too small, it is not hard to blow it up to the entire clique. Indeed, suppose that we have successfully found a seed set $S \subseteq K$, we can recover the whole planted clique by first taking the union of the seed set and its common neighbors and then finding the k highest-degree vertices therein.

The details of the iterated degree test is summarized in Algorithm 3. Here, for each vertex v and subset S of vertices,
$$d_S(v) = \sum_{j \in S} \mathbb{1}\{j \sim i\}$$
denote the number of neighbors of v in S (the degree of v restricted on S).

Theorem 2.4. *If $k = C\sqrt{n}$ for sufficiently large $C > 0$, then $\widehat{K} = K$ with high probability.*

Remark 2.5. Before proving the theorem, let us explain the first two steps in Algorithm 3:

- In Step 1, the vertex V_t is chosen based on its degree, since vertices in the clique tend to have a higher degree. Thus, intuitively, we could have chosen V_t as the vertices in G_{t-1} whose degree exceeds a given threshold. However, in this way we created a lot of dependency and we cannot ensure each G_t is still an instance of the planted clique model (because we want to apply the degree test and invoke Lemma 2.3). Instead, what we did is to choose a "test set" S_t and compute the degree by withholding this test set. This is a commonly used trick for type of problems, which we will revist later in stochastic block models.

Algorithm 3: Iterative degree testing algorithm [DGGP11]

```
Step 1: Given parameters $\alpha, \beta, T$, we will generate $G = G_0 \supset G_1 \supset \cdots \supset G_T$ as follows
for $t = 0$ to $T - 1$ do
    Given the current $G_t = (V_t, E_t)$, pick a test subset $S_t \subset V_t$ by including each vertex with probability $\alpha$.
    Let $V_{t+1} \triangleq \{v \in V_t \setminus S_t \colon \; d_{S_t}(v) \geq \frac{1}{2}|S_t| + \frac{\beta}{2}\sqrt{|S_t|}\}$, namely, those vertices whose number of neighbors in the test set is statistically siginificant.
    Denote the induced subgraph $G_{t+1} = G[V_{t+1}]$
end for
Step 2: Let $\widetilde{K}$ = set of $\frac{k'_T}{2}$ highest-degree vertices on $G_T$, where $k'_T$ is defined in (2.6).
Step 3: Let $K'$ be the union of $\widetilde{K}$ and its common neighbors. Report $\widehat{K}$, the $k$ highest-degree vertices in $G' = G[K']$.
```

- In Step 2, since $G_T \sim G(n_T, \frac{1}{2}, k_T)$, according to the degree test we should choose $\widetilde{K}$ as the k_T highest-degree vertices. However, since k_T is not observed, we use $\frac{k'_T}{2}$ as a conservative proxy, which is a high-probability lower bound for k_T.

To prove the theorem, we will prove a series of claims, upon which the proof of the theorem becomes straightforward. In order to focus on the main ideas, we will be sloppy with notations like " $\approx$" and "whp".

Claim 2.1. *Define $n_t = |V_t|$ and $k_t = |K \cap V_t|$. For all t, G_t is an instance of $G(n_t, \frac{1}{2}, k_t)$. In other words, conditioned on (n_t, k_t), $G_t \sim G(n_t, \frac{1}{2}, k_t)$.*

Proof. This claim is true because the vertex set V_t is chosen without exposing any edges inside S_t^c (and hence V_t). Indeed, by induction, it suffices to consider $t = 0$, i.e., $G_0 \to G_1$. Note that each vertex $v \in S_0^c$ is included in V_1 if its degree in the test set S_0, namely, $d_{S_0}(v)$, exceeds a given threshold. Therefore $G_1 = (V_1, E_1)$ is distributed as $G(n_1, \frac{1}{2}, k_1)$, where $n_1 = |V_1|$ and $k_1 = |V_1 \cap K|$. This can also been seen from the perspective of the adjacency matrix: V_1 is determined based on the submatrix $A_{S_0^c, S_0}$ and hence independent of $A_{S_0^c, S_0^c}$ (see Fig. 2.1). □

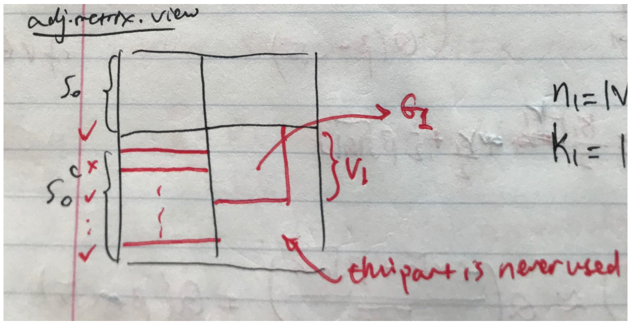


Figure 2.1: The vertex set V_1 is chosen by withholding the test set S_0.

Claim 2.2. *Let $G_t = (V_t, E_t)$, then*

$$n_t \approx n'_t \triangleq \tau^t n \text{ and } k_t \approx k'_t \triangleq \rho^t k \quad w.h.p.,$$

where τ and ρ are defined in (2.7).

Proof. Again, by induction, it suffices to consider $t = 0$, i.e., $G_0 \to G_1$. We will use the Berry-Esseen inequality (Lemma 2.4) to approximate binomials by Gaussians. Since $|S_0| \sim \text{Binom}(n, \alpha) \overset{w.h.p}{\approx} \alpha n$, it follows that $\forall v \in S_0^c$,

$$d_{S_0}(v) \sim \begin{cases} \text{Binom}(|S_0|, \frac{1}{2}) & \text{if } v \notin K \\ |S_0 \cap K| + \text{Binom}(|S_0 \backslash K|, \frac{1}{2}) & \text{if } v \in K \end{cases}$$

where $|S_0 \cap K| \sim \text{Binom}(k, \alpha) \approx k\alpha = C\sqrt{n}\alpha$ whp. Thus,

$$\mathbb{P}\left(d_{S_0}(v) \geq \frac{1}{2}|S_0| + \frac{\beta}{2}\sqrt{|S_0|}\right) \overset{\text{B-E}}{\approx} \begin{cases} Q(\beta) & \text{if } v \notin K \\ Q(\beta - C\sqrt{\alpha}) & \text{if } v \in K \end{cases} .$$

Finally, in summary,

$$|V_1| = \sum_{v \in V_0} \mathbb{1}\{v \in S_0^c\}\mathbb{1}\{d_{S_0}(v) \geq \frac{1}{2}|S_0| + \frac{\beta}{2}\sqrt{|S_0|}\} \approx |V_0| \underbrace{(1-\alpha)Q(\beta)}_{\triangleq \tau} .$$

Similarly,

$$k_1 = |K \cap V_1| = \sum_{v \in K} \mathbb{1}\{v \in S_0^c\}\mathbb{1}\{d_{S_0}(v) \geq \frac{1}{2}|S_0| + \frac{\beta}{2}\sqrt{|S_0|}\} \approx k \underbrace{(1-\alpha)Q(\beta - C\sqrt{\alpha})}_{\triangleq \rho} .$$

□

Claim 2.3. *Let $\widetilde{K}$ be the set of the $\frac{k'_T}{2}$ highest-degree vertices in G_T. Choose $T = C_0 \log\log n$ (so that whp $\frac{k_T^2}{n_T} \geq \log^2 n$ say, and $n_T \approx n'_T = \rho^T n \geq \frac{n}{\mathsf{polylog}(n)}$ and $k_T \approx k'_T = \tau^T k \geq \frac{k}{\mathsf{polylog}(n)}$.) Then $\widetilde{K} \subset K$ with high probability.*

Proof. By Lemma 2.3, with probability $\geq 1 - n_T e^{-k_T^2/(8n_T)} \geq 1 - e^{-\mathsf{polylog}(n)}$, the nodes in the hidden clique have the highest degrees in $G_T \sim G(n_T, \frac{1}{2}, k_T)$. On the high probability event that $k_T \geq \frac{k'_T}{2}$, we have $\widetilde{K} \subset K$. □

Now that we have shown $\widetilde{K}$ is a subset of the true clique, we still need to expand it to the entire clique. Think of $\widetilde{K}$ as a "seed set" and the main point is in this case $s = |\widetilde{K}| \geq (1+\epsilon)\log n$ seeds suffice. However, the caveat is that $\widetilde{K}$ obtained from steps 1 and 2 is random and may depend on the entire graph. Fortunately, at this point, this can be addressed by taking a union bound over all s-subsets of K.

Claim 2.4 (Clean-up)**.** *With high probability, the following holds: Let $\widetilde{K}$ is an s-subset of K (which can be adversarilly chosen). Let K' be the union of $\widetilde{K}$ and its common neighbors. Let $\widehat{K}$ be the k highest-degree vertices on $G' = G[K']$. If $k = |K| \geq C \log n$ for a suffciently large constant C and $s \geq (1+\epsilon)\log_2 n$ for any constant $\epsilon \in (0,1)$, then $\widehat{K} = K$ with high probability.*

Proof. Let

$$K' = \widetilde{K} \cup \text{ common neighbors } = K \cup F,$$

where F denotes the non-clique common neighbors. We first show that $|F|$ is small. Fix set $\widetilde{K}$. For any node $u \in [n]\backslash K$,

$$\mathbb{P}[u \in F] = \mathbb{P}\left[u \sim v, \forall v \in \widetilde{K}\right] = \prod_{v \in \widetilde{K}} \mathbb{P}[u \sim v] = 2^{-s}.$$

Moreover, the events $\{u \in F\}$ are mutually independent across all $u \in [n]\backslash K$. Thus, for a given set $\widetilde{K}$,

$$\mathbb{P}\left[|F| \geq \ell\right] \leq \binom{n}{\ell} 2^{-s\ell}.$$

Taking a union bound over all possible $\widetilde{K}$ yields that

$$\begin{aligned}\mathbb{P}\left[\exists \widetilde{K} : |F| \geq \ell\right] &\leq \binom{k}{s}\binom{n}{\ell} 2^{-s\ell} \\ &\leq k^s n^\ell 2^{-s\ell} \\ &\leq 2^{s \log_2 k + \ell \log_2 n - s\ell} \\ &= 2^{(1+\epsilon)\log_2 n \log_2 k - \epsilon \ell \log_2 n} \to 0\end{aligned}$$

if $\ell = \frac{2}{\epsilon} \log k$.

Now in G', for any $v \in K$, we have by definition $d(v) \geq k-1$; for any $v \notin K$, we have

$$d(v) \leq |F| + d_K(v) \leq \underbrace{|F|}_{\leq O_P(\log k)} + \underbrace{\max_{v \notin K} d_K(v)}_{\leq k/2 + O_P(\sqrt{k \log n})} < k-1,$$

where the last inequality holds under the assumption that $k \geq C \log n$ for a sufficiently large constant C. This shows that the k highest-degree vertices in G' are precisely the true clique K. □

Our goal is to use the spectral method for statistical inference. As we will see, in planted clique and many related planted problems, the first few eigenvectors of the population matrix $\mathbb{E}X$ contain the information about the planted structures that we are interested in. Since we only have observations X at hand and do not know $\mathbb{E}X$, we compute the first few eigenvectors of X instead. Writing $X = \mathbb{E}X + (X - \mathbb{E}X)$, we expect that the error of estimating the first few eigenvectors of $\mathbb{E}X$ can be bounded by the size of the pertubation $X - \mathbb{E}X$.

3.1 Review of linear algebra

3.1.1 Eigendecomposition

Suppose that X is a symmetric real valued matrix in $\mathbb{R}^{n\times n}$.

Definition 3.1. The pair (λ, v) with $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^n$ is an eigenpair of X, consisting of an eigenvalue λ and an eigenvector v, if

$$Xv = \lambda v.$$

We order the eigenvalues of X by their sizes such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The corresponding eigenvectors $[v_1, \ldots, v_n]$ form an orthonormal basis (ONB) of $\mathbb{R}^n$. Denote $V = [v_1, \ldots, v_n]$, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. We can write the eigendecomposition of X as

$$X = V\Lambda V^\top = \sum_{i=1}^{n} \lambda_i v_i v_i^\top.$$

Also note that $\operatorname{rank}(X) = r \Leftrightarrow$ there exist exactly r nonzero λ_i's.

3.1.2 Singular value decomposition (SVD)

Now suppose that $X \in \mathbb{R}^{m\times n}$ is a real valued rectangular matrix. The singular value decomposition (SVD) of X is

$$X = U\Sigma V^\top = \sum \sigma_i u_i v_i^\top,$$

where $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$, $\sigma_i \geq 0$, $U = [u_1, \ldots, u_r] \in \mathbb{R}^{m\times r}$ and $V = [v_1, \ldots, v_r] \in \mathbb{R}^{n\times r}$. The columns of U are orthonormal and we call them left singular vectors and likewise the columns of V are orthonormal too and we call them right singular vectors.
We can calculate Σ, U and V by taking eigendecompositions of $XX^\top$ and $X^\top X$. Indeed,

$$XX^\top = U\Sigma^2 U^\top \in \mathbb{R}^{m\times m} \qquad \text{and} \qquad X^\top X = V\Sigma^2 V^\top \in \mathbb{R}^{n\times n},$$

and

$$\sigma_i = \sqrt{\lambda_i(XX^\top)} = \sqrt{\lambda_i(X^\top X)}.$$

3.1.3 Matrix norms

Suppose again that $X \in \mathbb{R}^{m\times n}$. There are multiple ways to define a norm on X.

- We view X as a mn-dimensional vector with euclidean norm and define the Frobenius norm
$$\|X\|_F = \|\mathrm{vec}(X)\|_2 = \sqrt{\sum_{i,j} X_{ij}^2}.$$

- We view X as a linear operator from $(\mathbb{R}^n, \|\cdot\|_p) \to (\mathbb{R}^m, \|\cdot\|_q)$ with operator norm
$$\|X\|_{p\to q} = \sup_{\|v\|_p=1} \|Xv\|_q.$$
For this course the most relevant matrix is the case of $p = q = 2$, where we equip $\mathbb{R}^n$ with the euclidean inner product. We denote
$$\|X\|_{2\to 2} =: \|X\|_{op},$$
also known as the spectral norm.

We now prove that
$$\|X\|_{op} = \sigma_{\max}(X).$$

Using the SVD of X:
$$\|X\|_{op}^2 = \sup_{\|v\|_2=1} \|Xv\|_2^2 = \sup_{\|v\|_2=1} \left\|\sum \sigma_i u_i v_i^\top v\right\|_2^2 = \sup_{\|v\|_2=1} \sum \sigma_i^2 \langle v_i, v\rangle^2 = \sigma_{\max}(X)^2.$$

Remark 3.1.
- $\|\cdot\|_{op}$ is a norm and $\|X\|_{op} = \|X^\top\|_{op}$.
- $\|XY\|_{op} \le \|X\|_{op}\|Y\|_{op}$.
- If $X = x$ is a vector then $\|X\|_{op} = \|x\|_2$.
- $\|\cdot\|_{op}$ is orthogonal invariant, i.e. for any $R \in \mathbb{O}(n)$, $R' \in \mathbb{O}(m)$ we have $\|R'XR\|_{op} = \|X\|_{op}$.
- If $X = [X_1, \dots, X_n]$ has orthonormal rows (columns), then $\|X\|_{op} = 1$.

Remark 3.2. Recall the matrix inner product: $\langle X, Y\rangle = \mathrm{trace}(Y^\top X) = \sum_{i,j} X_{ij}Y_{ij}$. Using this we can write
$$\|X\|_{op} = \sigma_{\max}(X) = \sup_{\|u\|_2=\|v\|_2=1} \langle X, uv^\top\rangle = \sup_{\|A\|_F=1,\ \mathrm{rank}(A)=1} \langle X, A\rangle.$$

Likewise, if X is real and symmetric we have that
$$\lambda_{\max}(X) = \sup_{\|v\|_2=1} \langle X, vv^\top\rangle, \quad \|X\|_{op} = \sigma_{\max}(X) = \sup_{\|v\|_2=1} |\langle X, vv^\top\rangle|.$$

Similar relations hold for $\sigma_{\min}$ and $\lambda_{\min}$ if one substitutes the sup's above for inf's.

3.2 Pertubation of eigenstructures

In this section we assume that we are given two matrices, X and $Y = X + Z$ where Z is a 'pertubation' of X. We are interested if eigenvectors and eigenvalues of X and Y are close when Z is 'small'. Unfortunately, in general this is not the case.

3.2.1 Negative results

Eigenvalues The eigenvalues λ_i are the roots of the polynomial $\det(\lambda I - X) = 0$, which is a polynomial in λ of degree n. Although the roots are continuous in the coefficients of the polynomial, in general the modulus of continuity is not Lipschitz and only $\frac{1}{\text{degree}}$-Hölder continuous, and this is tight. Indeed, consider the two matrices

$$X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad X_\varepsilon = \begin{bmatrix} 0 & 1 \\ \varepsilon & 0 \end{bmatrix}$$

Then $\lambda_1(X) = \lambda_2(X) = 0$, but $\lambda_1(X_\varepsilon) = \sqrt{\varepsilon}$ and $\lambda_2 = -\sqrt{\varepsilon}$. More generally consider

$$X = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & 1 \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix} \quad \text{and} \quad X_\varepsilon = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & 1 \\ \varepsilon & \dots & \dots & \dots & 0 \end{bmatrix}.$$

One can show that $\lambda_i(X) = 0$ but that $\lambda_i(X_\varepsilon) = (-1)\varepsilon^{1/n} e^{2\pi i \mathbf{j}/n}$ for $i = 1, \dots, n$, where $\mathbf{j}$ denotes the imaginary part.

Therefore we need more assumptions on X to be able to obtain Lipschitz bounds, e.g. that X is a real and symmetric matrix.

Eigenvectors But even in the symmetric case eigenvector pertubations may fail dramatically. For $\varepsilon > 0$ consider

$$X = \begin{bmatrix} 1+\varepsilon & 0 \\ 0 & 1-\varepsilon \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{bmatrix}.$$

The eigenvalues of these two matrices are the same

$$\lambda_1(X) = \lambda_1(Y) = 1 + \varepsilon, \ \lambda_2(X) = \lambda_2(Y) = 1 - \varepsilon.$$

However, the eigenvectors are far apart:

$$v_1(X) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2(X) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \textbf{but} \quad v_1(Y) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v_2(Y) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The lesson from this is that we need separation between the eigenvalues, a spectral (eigen) gap.

3.2.2 Pertubation bound for eigenvalues

Let X, Y, Z be real symmetric matrices in $\mathbb{R}^{n \times n}$ and suppose $Y = X + Z$. We have that

$$\begin{aligned} \lambda_1(X) + \lambda_n(Z) = \lambda_1(X) + \inf_{\|v\|_2 = 1} \langle Z, vv^\top \rangle \\ \leq \sup_{\|v\|_2 = 1} \langle X + Z, vv^\top \rangle = \lambda_1(Y) \\ \leq \lambda_1(X) + \sup_{\|v\|_2 = 1} \langle Z, vv^\top \rangle = \lambda_1(X) + \lambda_1(Z) \end{aligned}$$

and therefore

$$|\lambda_1(X) - \lambda_1(Y)| \leq \max(|\lambda_1(Z)|, |\lambda_n(Z)|) = \|Z\|_{op}.$$

More generally we have the following theorem (homework):

Theorem 3.1 (Weyl's inequality / Lidskii's inequality)**.**

$$|\lambda_i(X) - \lambda_i(Y)| \leq \|Z\|_{op}.$$

3.2.3 Pertubation bounds for eigenspaces

Let X, Y, Z again be real symmetric matrices in $\mathbb{R}^{n\times n}$ and suppose $Y = X + Z$. Suppose that $X = \sum_i \lambda_i u_i u_i^\top$ and $Y = \sum \rho_i v_i v_i^\top$. We want to prove a pertubation bound for $u \triangleq u_1$ and $v \triangleq v_1$ and more generally for $U = [u_1, \dots, u_r]$ and $V = [v_1, \dots, v_r]$. However, considering $\|u - v\|_2$ makes no sense as u and v are only determined up to their sign, and similarly U and V are only defined up to orthogonal transformation. There are two possible workarounds:

- Consider the distance

$$\min_{s\in\{\pm 1\}} \|u + sv\|_2 = \sqrt{2 - 2|\langle u, v\rangle|} = \sqrt{2 - 2\cos\theta} = 2\sin\frac{\theta}{2} \leq \sqrt{2}\sin\theta, \tag{3.1}$$

  where $\cos(\theta) \triangleq |\langle u, v\rangle|$, and more generally, $\inf_{R\in O(r)} \|U - VR\|$.

- Consider the distance between the linear subspaces spanned by u and v, defined through their respective projection matrices:

$$\left\| uu^\top - vv^\top \right\|_F^2 = 2(1 - \langle u, v\rangle^2) = 2\sin^2(\theta),$$

  and in the general case $\|UU^\top - VV^\top\|_F$ or $\|UU^\top - VV^\top\|_{op}$.

Theorem 3.2 (Davis-Kahan)**.** *Let* $\cos\theta = |\langle u_1, v_1\rangle|$. *Suppose* $\max(\rho_1 - \lambda_2, \lambda_1 - \rho_2) > 0$. *Then*

$$\sin\theta \leq \frac{\|Z\|_{op}}{\max(\rho_1 - \lambda_2, \lambda_1 - \rho_2)}.$$

Proof. Assume that $\rho_1 \geq \lambda_2$. Let us start from the eigenvalue equations:

$$Xu = \lambda_1 u \quad \text{and} \quad Yv = \rho_1 v.$$

Denote $U_\perp = [u_2, \dots, u_n] \in \mathbb{R}^{n\times n-1}$. Then

$$U_\perp^\top X = \begin{bmatrix} u_2^\top \\ \vdots \\ u_n^\top \end{bmatrix} X = \begin{bmatrix} \lambda_2 u_2^\top \\ \vdots \\ \lambda_n u_n^\top \end{bmatrix} = \begin{bmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} u_2^\top \\ \vdots \\ u_n^\top \end{bmatrix}.$$

Hence

$$\begin{aligned} U_\perp^\top (X + Z)v = \rho_1 U_\perp^\top v \Leftrightarrow & \begin{bmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} U_\perp^\top v + U_\perp^\top ZV = \rho_1 U_\perp^\top v \\ \Leftrightarrow & \begin{bmatrix} \rho_1 - \lambda_2 & & \\ & \ddots & \\ & & \rho_1 - \lambda_n \end{bmatrix} U_\perp^\top v = U_\perp^\top ZV \\ \Leftrightarrow & \; U_\perp^\top v = \begin{bmatrix} \frac{1}{\rho_1 - \lambda_2} & & \\ & \ddots & \\ & & \frac{1}{\rho_1 - \lambda_n} \end{bmatrix} U_\perp^\top ZV. \end{aligned}$$

Taking the $\|\cdot\|_{op}$-norm on both sides gives

$$\|U_\perp^\top v\|_2 \le \left\| \begin{bmatrix} \frac{1}{\rho_1-\lambda_2} & & \\ & \ddots & \\ & & \frac{1}{\rho_1-\lambda_n} \end{bmatrix} \right\|_{op} \left\|U_\perp^\top\right\|_{op} \|Z\|_{op} = \frac{\|Z\|_{op}}{\rho_1-\lambda_2}.$$

Finally, note that

$$\|U_\perp^\top v\|_2^2 = v^\top U_\perp U_\perp^\top v = v^\top (I - uu^\top) v = 1 - \langle u, v\rangle^2 = \sin^2(\theta).$$

If $\rho_1 < \lambda_2$, then $\lambda_1 > \rho_2$. Exchanging the roles of X and Y we obtain the other statement. □

More generally, considering the first r eigenvectors we have for $U = [U_1, \ldots, U_r]$ and $V = [V_1, \ldots V_r]$ that for any unitarily invariant norm $\|\cdot\|$,

$$\|U_\perp^\top V\| \le \frac{\|Z\|}{\max(\rho_r - \lambda_{r+1}, \lambda_r - \rho_{r+1})}.$$

One can generalize this to singular vectors by a technique sometimes called self-adjoint dilation:[1] For $X = U\Sigma V^\top \in \mathbb{R}^{m\times n}, Y = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$ consider the matrix

$$\begin{bmatrix} 0 & X \\ X^\top & 0 \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)},$$

and likewise for Y. Observe that

$$\begin{bmatrix} 0 & X \\ X^\top & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \sigma_1 \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 0 & X \\ X^\top & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ -v_1 \end{bmatrix} = -\sigma_1 \begin{bmatrix} u_1 \\ -v_1 \end{bmatrix}.$$

Now we can apply the Davis-Kahan Theorem (and $\sin\frac{\theta}{2} \le \sin\theta$) to obtain

$$\min_{s\in\{\pm 1\}} \left\| \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} + s \begin{bmatrix} \widetilde{u}_1 \\ \widetilde{v}_1 \end{bmatrix} \right\|_2 \le \frac{2\left\| \begin{bmatrix} 0 & Z \\ Z^\top & 0 \end{bmatrix} \right\|_{op}}{|\sigma_1(X) - \sigma_2(Y)|} = \frac{2\|Z\|_{op}}{|\sigma_1(X) - \sigma_2(Y)|}.$$

[1] Thanks for Cheng Mao for pointing this out.

§ 4. Basic Random Matrix Theory

Let $Z = (Z_{ij}) \in \mathbb{R}^{n \times n}$ be an *i.i.d* or a symmetric with upper part *i.i.d* (i.e. $Z_{ij} = Z_{ji} \overset{i.i.d.}{\sim} P$ for $1 \le i \le j \le n$) matrix with $\mathbb{E}Z_{ij} = 0$ and bounded variance and sub-Gaussian tails. We will prove next that w.h.p.

$$\|Z\|_{\text{op}} \le C\sqrt{n}.$$

In comparison we have w.h.p.

$$\|Z\|_F \asymp n.$$

For now we will discuss some intuition why $\sqrt{n}$ is the right order for the operator norm by focusing on $P = \mathcal{N}(0,1)$. First observe that

$$Z = \begin{bmatrix} Z_1^\top \\ \vdots \\ Z_n^\top \end{bmatrix}$$

and $\|Z_i\|_2 \asymp \sqrt{n}$ by the CLT.

<u>Intuition 1</u>: Observe that $\theta_i \triangleq \frac{Z_i}{\|Z_i\|_2} \sim \text{Unif}(S^{n-1})$, where $\text{Unif}(S^{n-1})$ stands for the unit sphere in $\mathbb{R}^n$. Thus $\langle \theta_i, \theta_j \rangle \asymp \frac{1}{n}\langle Z_i, Z_j \rangle = \frac{\sum_k Z_{ik} Z_{jk}}{n} \asymp \frac{1}{\sqrt{n}}$, where the last step holds by CLT as $Z_{ik}Z_{jk}$ have zero mean, variance 1, and mutually independent across all k. Therefore the rows of Z are almost orthogonal and hence the operator norm should roughly equal the $\|\cdot\|_2$-norm of the largest row.

<u>Intuition 2</u>: Since $\|Z\|_{\text{op}} = \sup_v \frac{\|Zv\|_2}{\|v\|_2}$, fix a particular $v \in S^{n-1}$. Then

$$Zv = \begin{bmatrix} \langle Z_1, v \rangle \\ \vdots \\ \langle Z_n, v \rangle \end{bmatrix} \sim \mathcal{N}(0, I_n).$$

This shows that $\|Z\|_{\text{op}} \ge \sqrt{n}$. A better choice is $v = Z_1$. Indeed, in that case

$$Zv = \begin{bmatrix} \|Z_1\|_2^2 \\ \langle Z_2, Z_1 \rangle \\ \vdots \\ \langle Z_n, Z_1 \rangle \end{bmatrix}.$$

As before $\|Z_1\|_2^2 \asymp n$ and $\sum_{i>1} \langle Z_i, Z_1 \rangle^2 \asymp n^2$ which yields that $\|Zv\|_2 \asymp \sqrt{2}n$, which shows that $\|Z\|_{\text{op}} \ge \sqrt{2n}$. In fact one can do even better and prove that w.h.p.

$$\|Z\|_{\text{op}} = (2 + o(1))\sqrt{n}.$$

4.1 Gaussian Random Matrix

For simplicity, let's consider a symmetric $n \times n$ Z having independent $N(0,1)$ above-diagonals and $N(0,2)$ diagonals. It will become transparent that the variance of the diagonal is immaterial, provided it is small, say, a constant. This model is referred to as Gaussian Orthogonal Ensemble (GOE). We are interested in

$$\|Z\|_{\text{op}} = \sigma_{\max} = \max_{u,v\in S^{n-1}} \langle Z, uv^\top \rangle. \tag{4.1}$$

Note that we can write $Z = \frac{1}{\sqrt{2}}(W + W^\top)$, where W is an $n \times n$ matrix whose n^2 entries are all iid $N(0,1)$.

Remark 4.1. The distributions of the diagonals are not important for the operator norm. To see this, note

$$\|Z\|_{\text{op}} \le \|Z_o\|_{\text{op}} + \|\text{diag}(Z)\|_{\text{op}},$$

where Z_o is the same as Z except that the diagonals are set to zero, and $\text{diag}(Z) = \text{diag}(Z_{ii})$. By union bound, $\|\text{diag}(Z)\|_{\text{op}} = \max_{1\le i\le n} |Z_{ii}| = O_p(\sqrt{\log n}) \ll O_p(\sqrt{n})$, and thus negligible.

Starting from the variational formula (4.1), note that for fixed $u, v \in \mathbb{S}^{n-1}$, we have

$$\langle Z, uv^\top \rangle = \frac{1}{\sqrt{2}} \langle W + W^\top, uv^\top \rangle = \frac{1}{\sqrt{2}} \langle W, uv^\top + uv^\top \rangle \sim N\left(0, \frac{1}{2}\|uv^\top + vu^\top\|_{\text{F}}^2\right) = N(0,2),$$

because $\|uv^\top + vu^\top\|_{\text{F}}^2 = 2 + \langle uv^\top, vu^\top \rangle = 4$. However, we need to deal with all u, v simultaneously,
To bound $\mathbb{P}(\max_{v\in S^{n-1}} \langle Z, vv^\top \rangle > t)$, we would like to apply the union bound. However, the sphere here is not a finite set. In order to handle this, we can use the discretization technique — the ϵ-net argument – to approximate a continuous max by a discrete max.

Definition 4.1. $V \subset S^{n-1}$ is called an *ϵ-net (covering)*, if $\forall u \in S^{n-1}$, $\exists v \in V$ s.t. $\|u - v\|_2 \le \epsilon$.

Lemma 4.1. *Let $\epsilon < \frac{1}{2}$. For any ϵ-net V,*

$$\max_{u,v\in V} \langle Z, uv^\top \rangle \le \|Z\|_{\text{op}} \le \frac{1}{1-2\epsilon} \max_{u,v\in V} \langle Z, uv^\top \rangle.$$

Proof. We only need to show the right inequality. Choose $u \in S^{n-1}$ such that $\langle Z, uv^\top \rangle = \|Z\|_{\text{op}}$. Then $\exists \widetilde{u}, \widetilde{v} \in V$ such that $\|u - \widetilde{u}\|_2 \le \epsilon$ and $\|v - \widetilde{v}\|_2 \le \epsilon$. It follows that

$$\begin{aligned}
\|Z\|_{\text{op}} = \langle Z, uv^\top \rangle &= \langle Z, \widetilde{u}\widetilde{v}^\top \rangle + \langle Z, uv^\top - \widetilde{u}\widetilde{v}^\top \rangle \\
&= \langle Z, \widetilde{u}\widetilde{v}^\top \rangle + \langle Z, (u - \widetilde{u})v^\top \rangle + \langle Z, \widetilde{u}(v - \widetilde{v})^\top \rangle \\
&\le \langle Z, \widetilde{u}\widetilde{v}^\top \rangle + \|Z(u - \widetilde{u})\|_2 + \|Z(v - \widetilde{v})\|_2 \\
&\le \max_{u,v\in V} \langle Z, uv^\top \rangle + 2\epsilon \|Z\|_{\text{op}}.
\end{aligned}$$

□

Next we give a simple bound on the cardinality of the ϵ-net.

Definition 4.2. For $A \subset \mathbb{R}^d$, $V = \{v_1, \dots, v_m\} \subset A$ is called an *ϵ-packing*, if $\forall i \ne j$, $\|v_i - v_j\|_2 \ge \epsilon$. An ϵ-packing V is *maximal* if it cannot be made bigger, i.e., $\forall u \in A \backslash V$, $V \cup \{u\}$ is not an ϵ-packing.

We make two key observations for these concepts:

- Any maximal ϵ-packing is an ϵ-net.
- For any ϵ-packing V of A, $|V| \leq \mathrm{vol}(A + \frac{\epsilon}{2}B)/\mathrm{vol}(\frac{\epsilon}{2}B)$, where B is the unit norm ball. Here $A + B \triangleq \{x + y : x \in A, y \in B\}$ is the Minkowski sum of two sets.

The first observation is just by definition. We can construct a maximal ϵ-packing through greedy search. The second one is because we can put $|V|$ balls of radius $\frac{\epsilon}{2}$ into $A + \frac{\epsilon}{2}B$ and keep them disjoint. So the total volume of balls should not exceed that of the $A + \frac{\epsilon}{2}B$. Among many measures, we choose volume because it's location invariant. We can summarize the observations as

$$\text{size of the smallest covering } \leq \text{ size of any maximal packing } \leq \text{ volume ratio.}$$

Now set $A = S^{n-1}$. Then $A + \frac{\epsilon}{2}B \subset B + \frac{\epsilon}{2}B = (1 + \frac{\epsilon}{2})B$.[1] The volume ratio

$$\frac{\mathrm{vol}(A + \frac{\epsilon}{2}B)}{\mathrm{vol}(\frac{\epsilon}{2}B)} \leq \frac{\mathrm{vol}((1 + \frac{\epsilon}{2})B)}{\mathrm{vol}(\frac{\epsilon}{2}B)} = \frac{(1 + \frac{\epsilon}{2})^n \mathrm{vol}(B)}{(\frac{\epsilon}{2})^n \mathrm{vol}(B)} = \left(1 + \frac{2}{\epsilon}\right)^n .$$

What we discussed above concludes the following lemma.

Lemma 4.2 (Size of ϵ-net)**.** *There exists an ϵ-net V for S^{n-1}, of size $|V| \leq \left(1 + \frac{2}{\epsilon}\right)^n$.*

Remark 4.2. The above upper bound is essentially tight. To see this, for any ϵ-net V, we have $S^{n-1} \subset \cup_{v \in V} (v + \epsilon B)$ so $S^{n-1} + \epsilon B \subset \cup_{v \in V}(v + 2\epsilon B)$. Thus by the union bound,

$$\mathrm{vol}(S^{n-1} + \epsilon B) \leq \mathrm{vol}\left(\cup_{v \in V} (v + 2\epsilon B)\right) \leq \sum_{v \in V} \mathrm{vol}(2\epsilon B) = |V|(2\epsilon)^n \mathrm{vol}(B).$$

For small ϵ, $S^{n-1} + \epsilon B = (1 + \epsilon)B \backslash (1 - \epsilon)B$ is a spherical shell. So $\mathrm{vol}(S^{n-1} + \epsilon B) = ((1 + \epsilon)^n - (1 - \epsilon)^n)\mathrm{vol}(B)$, and we get $2^n|V| \geq (\frac{1}{\epsilon} + 1)^n - (\frac{1}{\epsilon} - 1)^n \asymp n(\frac{1}{\epsilon})^{n-1}$

Theorem 4.1. *$\|Z\|_{op} \leq C\sqrt{n}$ whp for some universal constant C.*

Proof. Set $\epsilon = \frac{1}{4}$ and choose V as in Lemma 4.2 with $|V| \leq 9^n$. By Lemma 4.1, $\|Z\|_{\mathrm{op}} \leq 2 \max_{u,v \in V} \langle Z, uv^\top \rangle$. For $\forall t > 0$,

$$\begin{aligned} \mathbb{P}\left(\max_{u,v \in V} \langle Z, uv^\top \rangle > t\right) &\leq \sum_{u,v \in V} \mathbb{P}(\langle Z, uv^\top \rangle > t) \\ &\leq |V|^2 \cdot 2e^{-\frac{t^2}{4}} = 2e^{n \log 9 - \frac{t^2}{4}}. \end{aligned}$$

Choose $t = \frac{C}{2}\sqrt{n}$ with a universal constant $C > 4\sqrt{\log 9}$. Then we know $\|Z\|_{\mathrm{op}} \leq C\sqrt{n}$ with probability at least $1 - 2e^{-C'n}$, where $C' = C^2/16 - \log 9 > 0$.

□

[1] The first inclusion does not lose much volume, because the volume of a ball in high dimension is concentrated near the shell anyway.

4.2 Sub-Gaussian Random Matrix

Reviewing the whole proof of Theorem 4.1, we can see there is only one part that the Gaussian assumption is used: the tail bound $\mathbb{P}(|\langle Z, uv^\top\rangle| > t) \leq 2e^{-\frac{t^2}{4}}$. Thus the result of Theorem 4.1 can be naturally extended to other random variables with such tail bound.

Definition 4.3. A random variable X is said to be sub-Gaussian with parameter σ^2, or σ^2-SG in short, if $\forall\lambda$, $\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \leq e^{\sigma^2\lambda^2/2}$.

For a σ^2-SG random variable X and $t > 0$, Chernoff bound yields that $\mathbb{P}(X - \mathbb{E}X > t) \leq e^{\sigma^2\lambda^2/2-\lambda t}$ for all $\lambda \geq 0$. Choosing $\lambda = t/\sigma^2$, we have $\mathbb{P}(X - \mathbb{E}X > t) \leq e^{-\frac{t^2}{2\sigma^2}}$, and similarly for the other tail. Overall, a σ^2-SG random variable satisfies the same tail bound as $N(0, \sigma^2)$, namely

$$\mathbb{P}(|X - \mathbb{E}X| > t) \leq 2e^{-\frac{t^2}{2\sigma^2}}. \tag{4.2}$$

We can also view the tail bound as the definition of σ^2-SG. Note that σ^2-SG random variables have variance at most σ^2, which can be shown easily through Taylor expansion of MGF. Sometimes σ^2 is called the "variance proxy".

We also need some basic observations on subgaussianity. The proof is omitted.

Lemma 4.3. *1. If X is σ^2-SG and $\tau^2 > \sigma^2$, then X is τ^2-SG.*

2. If X is σ^2-SG, then αX is $\alpha^2\sigma^2$-SG.

3. If X is σ^2-SG, then $X + \mu$ is σ^2-SG for any constant μ.

4. If $X_1, \ldots, X_n$ are independent and each X_i is σ_i^2-SG, then $\sum_{i=1}^n X_i$ is $(\sum_{i=1}^n \sigma_i^2)$-SG.

We are now ready to give an extension of Theorem 4.1:

Theorem 4.2. *Let $Z = (Z_{ij})_{n\times n}$ a real symmetric matrix with $\mathbb{E}Z = 0$. Assume that, for $1 \leq i \leq j \leq n$, Z_{ij} are independent and σ^2-SG. Then $\|Z\|_{op} \leq C\sqrt{n\sigma^2}$ with probability at least $1 - 2e^{-C'n}$ for some universal constant C, C'.*

Proof. For any $u, v \in S^{n-1}$, Lemma 4.3 shows that $\langle Z, uv^\top\rangle = \sum_i Z_{ii}u_iv_i + 2\sum_{i<j} Z_{ij}u_iv_j$ is SG with parameter

$$\sigma^2\left(\sum_i u_i^2v_i^2 + 4\sum_{i<j} u_i^2v_j^2\right) \leq 2\sigma^2\left(\sum_i u_i^2\right)\left(\sum_i v_i^2\right) = 2\sigma^2.$$

Thus we get from (4.2) that $\mathbb{P}(|X - \mathbb{E}X| > t) \leq 2e^{-\frac{t^2}{4\sigma^2}}$ for all $t > 0$. The rest is identical to the proof of Theorem 4.1. □

In order to analyze spectral method for the Planted Clique model, we need to deal with Bernoulli random matrices. So let's find the SG parameter of Bernoulli random variables. The following result of Hoeffding, previously stated as Lemma 2.2, shows that bounded random variables are SG.

Lemma 4.4 (Hoeffding)**.** *If $X \in [-a, a]$ a.s. for some $a > 0$, then it is $4a^2$-SG.*

Proof. First, we prove when X is a Rademacher random variable:

$$\mathbb{E}e^{\lambda X} = \frac{1}{2}(e^{\lambda} + e^{-\lambda}) = \sum_{k\geq 0, k \text{ even}}^{\infty} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2^k k!} = e^{\lambda^2/2}. \tag{4.3}$$

Second, when $|X| \leq a$ a.s., we apply a *symmetrization* argument. Let X' be an independent copy of X and ϵ a Rademacher random variable independent of everything else. Then $X - X'$ has a symmetric distribution and $X - X' \overset{\text{law}}{=} \epsilon(X - X')$. Then

$$\begin{aligned}
\mathbb{E}e^{\lambda(X-\mathbb{E}X)} &= \mathbb{E}e^{\lambda(X-\mathbb{E}X')} = \mathbb{E}e^{\lambda X}e^{-\lambda\mathbb{E}X'} \\
&\overset{(a)}{\leq} \mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda X'} = \mathbb{E}e^{\lambda(X-X')} \\
&= \mathbb{E}e^{\lambda\epsilon(X-X')} \\
&= \mathbb{E}\left(\mathbb{E}(e^{\lambda\epsilon(X-X')}|X, X')\right) \\
&\overset{(b)}{\leq} \mathbb{E}\left(e^{\lambda^2(X-X')^2/2}\right) \overset{(c)}{\leq} e^{2\lambda^2 a^2}
\end{aligned}$$

where (a) applies Jensen's inequality; (b) applies (4.3); and (c) is because $|X - X'| \leq 2a$ a.s.. □

Since Bernoulli random variable is 1-SG, an immediate consequence is the following tail bound for binomial random variables:

$$\mathbb{P}\left[\text{Binom}(n, p) \geq np + t\right] \leq \exp\left(-\frac{t^2}{2n}\right) \tag{4.4}$$

Combining Lemma 4.4 with Theorem 4.2, we get the following corollary:

Corollary 4.1. *Let A be the adjacency matrix of an (inhomogeneous) Erdős-Rényi graph G with n vertices, where $i \sim j$ with probability p_{ij} independently. Then $\|A - \mathbb{E}A\|_{op} \leq C\sqrt{n}$ with probability at least $1 - 2e^{-C'n}$ for some universal constant C, C'.*

For $G \sim G(n, p)$, the above result is does not depend on p, because Lemma 4.4 simply states that Bern(p) is 1-SG regardless of the value of p. As such, the resulting bound is tight for dense graphs with $p = O(1)$, but loose for sparse graphs with $p = o(1)$. For the latter case, it is natural to expect the spectral fluctuation to be $O(\sqrt{np}$ because the variance of Bern(p) is on the order p. Indeed, we can improve the sub-Gaussian constant to $\sigma^2(p) \lesssim \frac{1}{\log\frac{1}{p}}$, which is the best possible (check Def. 4.3 by setting $\lambda = \log(1/p)$), leading to the improved result that whp,

$$\|A - \mathbb{E}A\|_{\text{op}} \lesssim \sqrt{\frac{n}{\log\frac{1}{p}}}$$

However, this is still loose as the correct behavior is indeed $O(\sqrt{np})$ when p is not too small, namely, $p = \Omega(\frac{\log n}{n})$. (We will come back to this in Lecture 10.) Overall this means dealing with sparse graphs using sub-Gaussian technology leads to highly suboptimal result.

We now apply the perturbation result from Lecture 3 and the basic random matrix result from Lecture 4 to the hidden clique problem, following [AKS98].

5.1 Spectral methods for Planted Clique Model

Let $G \sim G(\frac{1}{2}, n, k)$ where the hidden k-clique is denoted by $K \subset [n]$. Denote the adjacency matrix of G by A, with

$$A_{ij} = \begin{cases} 1 & i,j \in K \\ \text{Bern}(\frac{1}{2}) & o.w. \end{cases}.$$

It will be more convenient to work with the signed adjacency matrix W (± 1 as opposed to $0/1$-valued), where

$$W_{ij} = \begin{cases} 2A_{ij} - 1 & i \neq j \\ 0 & i = j \end{cases}.$$

The following is a spectral method to find the clique:

1. Find the top eigenvector u of W.
2. Let $\widetilde{K}$ consist of those vertices $i \in [n]$ with the k largest $|u_i|$.
3. (Clean up) Denote by $\widehat{K}$ the set of vertices having $\geq \frac{3k}{4}$ neighbors in $\widetilde{K}$. In other words, $\widehat{K} = \{i \in [n] : d_{\widetilde{K}}(i) \geq \frac{3k}{4}\}$.

Theorem 5.1 ([AKS98])**.** *If $k \geq C\sqrt{n}$ for some large constant C, then $\mathbb{P}(\widehat{K} = K) \to 1$.*

Proof. First, we show $\widetilde{K}$ is that approximately correct: $|\widetilde{K} \cap K| \geq (1-\epsilon)k$ whp for some $\epsilon = \epsilon(C)$. Let $W^* = \xi\xi^\top$, where $\xi = \mathbb{1}_K = (\mathbb{1}\{i \in K\})_{1 \leq i \leq n}$ is the indicator vector for the planted clique. Since W^* is rank-one, so $\lambda_1(W^*) = \|\xi\|_2^2 = k$ with top eigenvector $v = \frac{1}{\sqrt{k}}\xi$, and $\lambda_2(W^*) = 0$. By Davis-Kahan's $\sin\Theta$-theorem (Theorem 3.2 together with (3.1)), provided that $\lambda_1(W^*) - \lambda_2(W) > 0$,

$$\min_{s \in \{\pm\}} \|u + sv\|_2 \leq \frac{2\|W - W^*\|_{\text{op}}}{\lambda_1(W^*) - \lambda_2(W)}. \tag{5.1}$$

WLOG, assume the LHS is $\|u - v\|_2$. Note that

$$\|W - W^*\|_{\text{op}} \leq \|\mathbb{E}W - W\|_{\text{op}} + \|\mathbb{E}W - W^*\|_{\text{op}} \leq \|\mathbb{E}W - W\|_{\text{op}} + 1 \leq C_0\sqrt{n} \tag{5.2}$$

whp for some universal $C_0 > 1$ and all sufficiently large n by Theorem 4.2. On the same event, by Weyl's inequality (Theorem 3.1), $\lambda_2(W) = \lambda_2(W) - \lambda_2(W^*) \leq \|W - W^*\|_{\text{op}} \leq C_0\sqrt{n}$. Plugging this back into (5.1), we get that

$$\|u - v\|_2 \leq \frac{2C_0}{C - C_0} \leq \epsilon \tag{5.3}$$

holds whp for C big enough.

Second, we claim that $\|u - v\|_2 \leq \epsilon$ actually implies

$$|\widetilde{K} \cap K| \geq (1 - \epsilon')k \tag{5.4}$$

for some small ϵ' depending on ϵ. To see this, note that $|K| = |\widetilde{K}| = k$ and thus $|K\backslash\widetilde{K}| = |\widetilde{K}\backslash K|$ and it is equivalent to show this is at most $\epsilon' k$. Moreover,

$$\epsilon^2 \geq \|u - v\|_2^2 = \sum_{i \in K} \left(u_i - \frac{1}{\sqrt{k}} \right)^2 + \sum_{i \notin K} u_i^2.$$

Now, we separately consider two cases.

- If $|u_i| \leq \frac{1}{2\sqrt{k}}$ for all $i \notin \widetilde{K}$, then

$$\epsilon^2 \geq \sum_{i \in K\backslash\widetilde{K}} \left(\frac{1}{\sqrt{k}} - u_i \right)^2 \geq \frac{1}{4k} |K\backslash\widetilde{K}|.$$

- If $|u_j| > \frac{1}{2\sqrt{k}}$ for some $j \notin \widetilde{K}$, then by the definition of $\widetilde{K}$, we have $|u_i| > \frac{1}{2\sqrt{k}}$ for all $i \in \widetilde{K}$. It follows that

$$\epsilon^2 \geq \sum_{i \in \widetilde{K}\backslash K} u_i^2 \geq \frac{1}{4k} |\widetilde{K}\backslash K|.$$

In sum, in either case, (5.4) holds with $\epsilon' = 4\epsilon^2$.

Third, we claim that $\widehat{K} = K$ with high probability. We proceed on the event of $\|u - v\|_2 \leq \epsilon$.

- For each vertex $i \in K$, $d_{\widetilde{K}}(i) \geq d_{\widetilde{K} \cap K}(i) = |\widetilde{K} \cap K| - 1 \geq (1 - \epsilon')k - 1$. So we have $i \in \widehat{K}$ when $\epsilon' < \frac{1}{4}$.

- For each vertex $i \notin K$, $d_{\widetilde{K}}(i) \leq d_K(i) + |\widetilde{K}\backslash K|$. From above, we know $|\widetilde{K}\backslash K| \leq \epsilon' k$. And $d_K(i) \sim \text{Bin}(k, \frac{1}{2})$. By Hoeffding's inequality, for all $\epsilon' \leq 1/8$,

$$\mathbb{P}\left(d_K(i) \geq (\frac{3}{4} - \epsilon')k \right) \leq \mathbb{P}\left(d_K(i) \geq \frac{5}{8}k \right) \leq e^{-\frac{k}{32}}.$$

  Applying a union bound over all vertices $i \notin K$ gives that

$$\mathbb{P}\left(\exists i \notin K : d_K(i) \geq (\frac{3}{4} - \epsilon')k \right) \leq n e^{-\frac{k}{32}}.$$

In all, on the events $\|u - v\| \leq \epsilon$ and $d_K(i) < (\frac{3}{4} - \epsilon')k$ for all $i \notin K$, we have $\widehat{K} = K$. To wrap up the whole proof, we choose $\epsilon = \frac{1}{8}$. Then $\epsilon' = 4\epsilon^2 = \frac{1}{16}$. Choose $C \geq 18C_0$ so the second inequality in (5.3) is guaranteed. Therefore,

$$\begin{aligned}
\mathbb{P}(\widehat{K} \neq K) &\leq \mathbb{P}\left(\|u - v\| > \epsilon \right) + \mathbb{P}\left(\exists i \notin K : d_K(i) > (\frac{3}{4} - \epsilon')k \right) \\
&\leq \mathbb{P}(\|W - \mathbb{E}W\|_{\text{op}} > C_0\sqrt{n}) + n e^{-\frac{k}{32}} \\
&\leq 2e^{-C_0' n} + n e^{-\frac{C}{32}\sqrt{n}} \to 0.
\end{aligned}$$

□

Remark 5.1. 1. Alternatively, one can take u as the second leading eigenvector of A. The top eigenvector of A is almost deterministic and not informative, since it is almost proportional to the all-ones vector.

2. Thresholding technique is widely used in non-parametric estimation. Here, the step 3 (clean up) can be viewed as a version of thresholding.

5.2 Improving the constant

Next we show that the constant C in Theorem 5.1 can be made arbitrarily small, at the price of increasing the time complexity (still $\mathsf{poly}(n)$ but with a bigger exponent). This part is generic and applies to any algorithm. The idea is as following. Fix a subset of vertices $S \subset V$, $|S| = s$. Define $N_*(S)$ as the set of common neighbors of S, i.e., $N_*(S) = \{v \in V \backslash S : \forall u \in S, v \sim u\} = \left(\bigcap_{u \in S} N(u)\right) \backslash S$. Let's say $s = 2$. Next, consider the induced subgraph $G' = G[N_*(S)]$. If $S \subset K$, then $G' = G(|N_*(S)|, \frac{1}{2}, k-2)$ and $|N_*(S)| \sim \text{Bin}(n-k, \frac{1}{4}) + k - 2 = (1 + o_P(1))\frac{n}{4}$ when $k = o(n)$. So as we can see, by working on this subgraph, n decreases *exponentially* while k decreases *linearly*.

The upgraded algorithm is thus summarized below:

> For any s-subset $S \subset V$, run the existing algorithm on $G' = G[N_*(S)]$ and output Q. Repeat until $S \cup Q$ is a k-clique. And the final output is $S \cup Q$.

When the search over S finds $S \subset K$, the requirement in Theorem 5.1 asks for $k - s \geq C\sqrt{n \cdot 2^{-s}}$ to guarantee the success of the spectral method, namely, $Q = K \backslash S$. So suppose we aim to find planted clique of size $k \geq \delta\sqrt{n}$, where δ is an arbitrary constant. Picking $s = 2\log_2 \frac{C}{\delta}$, the algorithm above is guaranteed to be find K whp. The extra search time is at most $\binom{n}{s} = n^{O(\log \frac{1}{\delta})}$ that is polynomial in n.

§ 6. Semidefinite programming I: Planted clique

In this lecture we discuss semidefinite programming (SDP) relaxation in the context of the planted clique problem. We discuss two SDPs:

- A standard form of SDP relaxation for the planted clique problem (after [FK00, HWX16])
- A convexified maximum likelihood estimator with nuclear norm constraint, which can also be written as SDP (after [CX16]).

We show that both methods succeed in finding the hidden clique of size $\Omega(\sqrt{n})$ with high probability, using the following two types of proof techniques respectively:

- Dual proof: we construct the needed Lagrangian multipliers (also called dual certificates or dual witnesses) that together with the desired solution X^* fulfill the KKT condition, thereby certifying the optimality of X^*.
- Primal proof: we show that no feasible solution other than the desired X^* achieve a higher objective function.

Although the two methods are of distinct nature (one is constructive and one is non-construcive), for analyzing convex programs both methods are ultimately equivalent; nevertheless, the specific execution (e.g. explicit construction of dual certificates) need not be the same.

In addition, we show that the standard SDP relaxation of the hidden clique problem, unlike previously discussed methods like the degree test or the spectral method, is robust with respect to certain adversarial perturbation. Aside from robustness, SDP also possesses the advantages of not requiring any cleanup step unlike spectral methods (Section 5.1) or the iterative degree tests in (Section 2.4), and that it can be solved in polynomial time. (See Section 6.2.4 on rounding an approximate solution.) However, in practice solving a large SDP can be quite slow.

6.1 The planted clique problem and spectral method revisited

Recall the in the planted clique model, a clique K of size k is planted in the Erdös-Rényi graph with success probability $1/2$. Denote the corresponding adjacency matrix as A, and $G(n, 1/2, k)$ the distribution A is generated from. Define $W \in \mathbb{R}^{n\times n}$ to be the following transformation of A:

$$W_{ij} = \begin{cases} 2A_{ij} - 1, & i \neq j; \\ 0, & i = j. \end{cases}$$

Note that W takes value 1 on edges connecting two members within the clique, and is *i.i.d.* $\mathrm{Rad}(1/2)$ for all other above-diagonal entries. It is easy to see that the MLE for the planted clique problem can be written as

$$\begin{aligned} \widehat{u}_{\text{MLE}} = \arg\max_u & \sum_{i,j} u_i u_j W_{ij} \\ \text{s.t.} \quad & u \in \{0,1\}^n, \\ & \sum u_i = k, \end{aligned}$$

or equivalently in matrix form,

$$\begin{aligned} \widehat{u}_{\text{MLE}} = \arg\max_u \ & \langle W, uu^\top \rangle \\ \text{s.t.} \quad & u \in \{0,1\}^n, \\ & \|u\|^2 = k. \end{aligned} \tag{6.1}$$

Notice that the spectral method in Section 5.1 that uses the top eigenvector of W is a relaxation of (6.1). By relaxation we mean enlarging the constraint set in order to speed up the computation. Indeed, the spectral method takes the top eigenvector of W by solving

$$\begin{aligned} \widehat{u}_{\text{spectral}} = \arg\max_u \ & \langle W, uu^\top \rangle \\ \text{s.t.} \quad & u \in \mathbb{R}^n, \\ & \|u\|^2 = k. \end{aligned} \tag{6.2}$$

One would hope that after relaxation, the optimality of the optimizer in the original problem is not lost. Unfortunately that is not the case for the relaxation done in (6.2). We proved previously in Lecture 5 that the spectral method typically only recovers most of the members in the clique, thus requires a cleanup step to recover the entire clique. In other words, we relaxed the constraint set too much. We will develop two tighter relaxations to (6.1), namely our semidefinite programs. The standard form of SDP is derived using the idea of lifting. Let us first illustrate the lifting idea on spectral program (6.2).

Definition 6.1. An optimization problem is said to be convex if the objective function is a convex function and the constraint set is a convex set.

Although easy to solve, the optimization (6.2) is not convex. Nevertheless, it can be written as a convex optimization via the lifting idea.

Definition 6.2. A symmetric matrix $X \in \mathbb{R}^{n\times n}$ is called positive semidefinite (PSD), denote by $X \succeq 0$, if $y^\top X y \geq 0$ for all $y \in \mathbb{R}^n$.

Consider the following program:

$$\begin{aligned} \widehat{X}_{\text{spectral}} = \arg\max_X \ & \langle W, X \rangle \\ \text{s.t.} \quad & X \succeq 0, \\ & \mathsf{Tr}(X) = k. \end{aligned} \tag{6.3}$$

Proposition 6.1. *The optimization* (6.2) *is equivalent to* (6.3).

To prove Proposition 6.1 we need to introduce the notion of extremal points in a convex set. Let S be a convex set. We say a point $s \in S$ is an *extreme point* of S if it cannot be written as convex combinations of other points in S. The importance of extremal points is that points in a convex set S can be written as convex combinations of extreme points. There are various results of this flavor in convex analysis, the most general being the Krein-Milman theorem. We recall the following result for Euclidean spaces:

Theorem 6.1 (Carathéodory theorem). *Suppose that S is a convex set in $\mathbb{R}^d$. Then each $s \in S$ can be written as a convex combination of at most $d+1$ extreme points of S.*

Proof of Proposition 6.1. Write $X = uu^\top$ to reformulate (6.2) as

$$\begin{aligned} &\max_X \ \langle W, X \rangle \\ &\text{s.t.} \quad X \succeq 0, \\ &\qquad\ \ \mathsf{Tr}(X) = k, \\ &\qquad\ \ \operatorname{rank}(X) = 1. \end{aligned}$$

The lifting step only drops the rank one constraint on X. The proposition can be proved by arguing that dropping the rank constraint does not incur any sub-optimality. Notice that in (6.3), the objective function is linear in X and the constraint set is convex. Therefore optimality occurs at one of the extreme points. To see that simply notice that for each feasible X, by Carathéodory's theorem it can be written as a convex combination of a finite number of extreme points in the feasible set. Write $X = \sum_i X_i \alpha_i$, which gives

$$\langle W, X \rangle = \sum_i \langle W, X_i \rangle \alpha_i.$$

Therefore the objective function evaluated at X has to be beaten (or at least match) its value at one of the extreme points. Given that all extreme points of the feasible set $\{X : X \succeq 0, \mathsf{Tr}(X) = k\}$ are of rank one, the rank constraint is automatically enforced by the optimization (6.3). □

6.2 Standard SDP relaxation

6.2.1 Formulation

Start by rewriting the MLE program (6.1) in a lifted form.

$$\begin{aligned} \widehat{X}_{\mathrm{MLE}} = \arg&\max_X \ \langle W, X \rangle \\ &\text{s.t.} \quad X \succeq 0, \\ &\qquad\ \ 0 \le X \le \mathbf{J}, \text{(entrywise)} \\ &\qquad\ \ \mathsf{Tr}(X) = k, \\ &\qquad\ \ \langle X, \mathbf{J} \rangle = k^2, \\ &\qquad\ \ \operatorname{rank}(X) = 1. \end{aligned} \tag{6.4}$$

Proposition 6.2. *The MLE optimization (6.1) is equivalent to (6.4).*

Proof. It is easy to check that for all u in the feasible set of (6.1), the matrix $uu^\top$ is in the feasible set of (6.4). We only need to check the other direction. In other words, we need to show that every X in the feasible set of (6.4) can be written as $uu^\top$ for some u in the feasible set of (6.1).

The positive semidefinite constraint combined with the rank one constraint imply that $X = uu^\top$ for some $u \in \mathbb{R}^n$. The trace constraint gets translated to $\sum_i u_i^2 = k$; the constraint $\langle X, \mathbf{J} \rangle = k^2$ is equivalent to $|\sum_i u_i| = k$. What's more, by looking at the diagonal entries $X_{ii} = u_i^2 \leq 1$ we have $u_i \in [-1, 1]$. Also, $X_{ij} \geq 0$, so all the non-zeros of u are either all positive or all negative. Assume the former. Then $\sum_i u_i^2 = k$ and $\sum_i u_i = k$ force the integrality $u_i \in \{0, 1\}$. Thus u lies in the feasible set of (6.1). □

We consider the following SDP relaxation of the MLE by dropping the rank-one constraint and the constraint that $X \leq \mathbf{J}$ in (6.4). (The first relaxation is the most important one as it makes the program convex. We could have kept the box constraint $X \leq \mathbf{J}$ but it turns out we don not need it for the proof of success.) Define

$$\begin{aligned} \widehat{X}_{\text{SDP}} = \arg\max_X \ & \langle W, X \rangle \\ \text{s.t.} \ & X \succeq 0, \\ & X \geq 0, \\ & \mathsf{Tr}(X) = k, \\ & \langle X, \mathbf{J} \rangle = k^2. \end{aligned} \tag{6.5}$$

Remark 6.1. Recall that the standard form for a linear program (LP) is

$$\begin{aligned} \max_x \ & \langle a, x \rangle \\ \text{s.t.} \ & \langle b_i, x \rangle \leq 0 \quad \text{for } i = 1, ..., m. \end{aligned}$$

Compare with the standard form for an SDP:

$$\begin{aligned} \max_{X \text{ symmetric}} \ & \langle W, X \rangle \\ \text{s.t.} \ & \langle B_i, X \rangle \leq 0 \quad \text{for } i = 1, ..., m, \\ & X \succeq 0. \end{aligned}$$

The set of all positive semidefinite matrices form a cone ($X \succeq 0$ and $\alpha \geq 0$ then $\alpha X \succeq \geq 0$). The PSD constraint is a conic condition. Notice that a matrix $X \in \mathbb{R}^{n \times n}$ is call PSD if $\langle X, uu^\top \rangle \geq 0$ for all $u \in \mathbb{R}^n$. Hence the PSD constraint can be viewed as a continuum of linear constraints.

6.2.2 Statistical guarantee: dual proof

Denote by $\xi = (\mathbb{1}\{i \in K\})_{1 \leq i \leq n}$ the indicator vector of the hidden clique K. Let $X^* = \xi\xi^\top$. We show that if the size of the clique is of order $\sqrt{n}$, then the SDP relaxation (6.5) is unique solved by X^*. Thus we can recover the hidden clique from the solution of the SDP.

Theorem 6.2. *There exists a constant $C > 0$ such that if $k \geq C\sqrt{n}$, then with high probability X^* is the unique maximizer of the SDP relaxation* (6.5).

Remark 6.2. With some work the constant C can be reduced to 1. We will not optimize the choice of C in the proof of the theorem.

Remark 6.3. Clearly X^* is *an* optimizer of (6.5). To prove Theorem 6.2 we will mostly need to establish the uniqueness of X^*. To see the optimality of X^*, note that for each X in the feasible set of (6.5),

$$\langle W, X \rangle = \langle W + \mathbf{I}, X \rangle - \mathsf{Tr}(X) \leq \langle \mathbf{J}, X \rangle - \mathsf{Tr}(X) = k^2 - k.$$

where the inequality is from $W + \mathbf{I} \le \mathbf{J}$ and the entrywise non-negativity of X. This is achieved by $\langle W, X^* \rangle = k^2 - k$, because W takes value 1 for all off-diagonal entries in $K \times K$. Therefore $\langle W, X \rangle \le \langle W, X^* \rangle$ for all feasible X.

The proof is via the standard dual approach to optimality: we will construct a set of Lagrangian multipliers for (6.5) (also called "dual certificates" or "dual witnesses") that certifies the optimality of X^*.

Attach to each constraint in (6.5) a Lagrangian multiplier, i.e., $S \succeq 0$, $B \ge 0$, $\eta, \lambda \in \mathbb{R}$. Write down the Lagrangian for the SDP (6.5):

$$L(X, S, B, \eta, \lambda) = \langle W, X \rangle + \langle S, X \rangle + \langle B, X \rangle + \eta(k - \mathsf{Tr}(X)) + \lambda \left(k^2 - \langle X, \mathbf{J} \rangle\right).$$

Since the inner products of two PSD matrices is always nonnegative, one important observation is that

$$\max_X \langle W, X \rangle \le \min_{S \succeq 0, B \ge 0, \eta, \lambda} \max_X L(X, S, B, \eta, \lambda). \tag{6.6}$$

Lemma 6.1. *Suppose there exists $S \succeq 0$, $B \ge 0$, $\eta, \lambda \in \mathbb{R}$ such that*

$$W + S + B - \eta \mathbf{I} - \lambda \mathbf{J} = 0, \quad \textit{(first-order condition)}$$

$$\langle S, X^* \rangle = 0, \quad \langle B, X^* \rangle = 0, \quad \textit{(complementary slackness)}$$

$$\lambda_{n-1}(S) > 0, \quad \textit{(uniqueness)}$$

then $X^* = \xi \xi^\top$ is the unique global maximizer for (6.5).

Proof. At the high level, the reasoning for any duality result is always of the following type:

1. First-order condition ensures $L(X, S, B, \eta, \lambda)$ is the same for any X;

2. Complementary slackness ensures $L(X^*, S, B, \eta, \lambda) = \langle W, X^* \rangle$.

Then we are done with optimality: $\forall$ feasible X,

$$\langle W, X \rangle \le L(X, S, B, \eta, \lambda) = L(X^*, S, B, \eta, \lambda) = \langle W, X^* \rangle.$$

Indeed, let's rewrite the Lagrangian multiplier as

$$L(X, S, B, \eta, \lambda) = \langle X, W + S + B - \eta \mathbf{I} - \lambda \mathbf{J} \rangle + k\eta + k^2 \lambda.$$

By the first order condition, the first term is 0 thanks to the first order condition. Hence $L(X, S, B, \eta, \lambda) = k\eta + k^2\lambda$ does not depend on X. By (6.6), for any feasible X we have

$$\langle W, X \rangle \le k\eta + k^2 \lambda.$$

By the complementary slackness condition X^* achieves the above with equality

$$\langle W, X^* \rangle = k\eta + k^2 \lambda.$$

This shows X^* is a maximizer.

To prove X^* is the maximizer (uniqueness), suppose for some feasible X' we have

$$\langle W, X' \rangle = \langle W, X^* \rangle.$$

Since $L(X', S, B, \eta, \lambda) = \langle W, X' \rangle + \langle S, X' \rangle + \langle B, X' \rangle \leq \langle W, X^* \rangle$, we must have $\langle S, X' \rangle = 0$. Note that the positive semidefiniteness of S implies that $\langle S, X^* \rangle = 0$ is equivalent to $S\xi = 0$, *i.e.*, ξ is an eigenvector for the smallest eigenvalue $\lambda_n = 0$ of S. Since X' is positive semidefinite and S has a strictly positive second smallest singular value, $\langle S, X' \rangle = 0$ forces

$$X' = c\xi\xi^\top = cX^*$$

for some constant c. The trace constraint ensures that $\mathsf{Tr}(X) = \mathsf{Tr}(X^*)$. Hence c has to be one, meaning $X' = X^*$. □

Proof of Theorem 6.2. From Lemma 6.1, it suffices to construct $B \geq 0$, $\eta\lambda \in \mathbb{R}$ such that

$$S = \eta \mathbf{I} + \lambda \mathbf{J} - B - W \succeq 0,$$

and $S\xi = 0$, $\langle B, X^* \rangle = 0$, $\lambda_{n-1}(S) > 0$.

The condition $S\xi = 0$ is equivalent to

$$\eta\xi + \lambda k \mathbf{1} = B\xi + W\xi. \tag{6.7}$$

Recall that $X^* = \xi\xi^\top$. The condition $\langle B, X^* \rangle = 0$ is equivalent to $B_{ij} = 0$ for all $(i, j) \in K \times K$. Therefore for all $i \in K$, the i'th entry of $B\xi$ is zero. Let $y = W\xi$. This vector records the (centered) number of neighbors of each vertex has in the clique. For $i \in K$, pull out the i'th place in the vector equality (6.7).

$$\eta + k\lambda = (B\xi)_i + y_i = k - 1.$$

Deduce that $\eta = k(1 - \lambda) - 1$.

For $i \notin K$, we have

$$k\lambda = (B\xi)_i + y_i. \tag{6.8}$$

Construct B from $B = \xi b^\top + b\xi^\top$ for some $b \in \mathbb{R}^n$ such that $b_i = 0$ for all $i \in K$. Such matrix B is of rank 2 and takes the block form

$$B = \begin{bmatrix} 0 & \text{column-wise constant} \\ \text{row-wise constant} & 0 \end{bmatrix}.$$

For B defined as such, we have $B\xi = kb$. Hence (6.8) can be rewritten as

$$k\lambda = kb_i + y_i,$$

implying that the choice of b satisfies $b_i = \lambda - y_i/k$ for all $i \notin K$.

We still need to ensure that $B \geq 0$. Equivalently, we need $b_i \geq 0$ for all i. This entails

$$\lambda \geq \frac{1}{k} \max_{i \notin K} y_i,$$

Note that each y_i is a sum of k i.i.d. *Rademacher*(1/2). Thus we can choose $\lambda = 1/2$ (in reality $\lambda = o(1)$ works) to ensure the above displayed equation holds with high probability.

It remains to verify $S \succeq 0$ and $\lambda_{n-1}(S) > 0$. In other words, we need

$$x^T S x > 0 \quad \text{for all } x \in \mathcal{S}^{n-1} \text{ s.t. } \langle x, \xi \rangle = 0.$$

Use the first-order condition to write

$$x^T S x = \eta + \lambda x^\top \mathbf{J} x - x^T B x - x^T W x.$$

The second term in the right-hand side is nonnegative by positive semidefiniteness of the all ones matrix. We have $x^T Bx = 0$ from $\langle x, \xi \rangle = 0$. Write $W = \mathbb{E}W + (W - \mathbb{E}W)$ to deduce that

$$x^T Sx \geq \frac{k}{2} - 1 - x^\top \mathbb{E}Wx - \|W - \mathbb{E}W\|_{\text{op}}.$$

Note that $\mathbb{E}W = \xi\xi^\top - \text{diag}(\xi)$, therefore $x^\top \mathbb{E}Wx \leq 0$ for all $x \perp \xi$. In (5.2) we have shown that $\|W - \mathbb{E}W\|_{\text{op}} \leq C_0\sqrt{n}$ with high probability. Hence $x^T Sx \geq (k - C_0\sqrt{n}) > 0$ as long as $k \geq C\sqrt{n}$ for some large enough constant c. □

6.2.3 A negative result

We showed exact recovery (with high probability) of the clique of the SDP (6.5) under the assumption that k is at least some large multiple of $\sqrt{n}$. We will show the requirement on the clique size cannot be improved. In fact if k is smaller than some small constant multiple of $\sqrt{n}$, there will typically be a spurious maximizer for (6.5) that does not provide any information on the location of the clique.

Theorem 6.3. *If $k \leq c\sqrt{n}$ (for example $c = 1/2$ works), then with high probability, the SDP (6.5) has a maximizer $X \in \mathbb{R}^{n\times n}$ that is supported by $K^C \times K^C$. In other words,*

1. *X is in the feasible set of (6.5).*
2. *$\langle W, X \rangle = \langle W, X^* \rangle = k^2 - k$.*

Proof. Define $Z = W_{K^C, K^C} \in \mathbb{R}^{m\times m}$, where $m = n - k$. We will construct the feasible alternative solution X as follows. Let

$$X = \begin{bmatrix} 0 & 0 \\ 0 & Y \end{bmatrix},$$

where $Y \in \mathbb{R}^{m\times m}$ is a matrix we need to specify so that X satisfies the feasibility conditions. In terms of Y, that means

1. $Y \succeq 0$,
2. $Y \geq 0$,
3. $\mathsf{Tr}(Y) = k$,
4. $\langle Y, \mathbf{J} \rangle = k^2$.

Let

$$Y = \frac{k}{m}\mathbf{I} + \alpha Z + \beta(\mathbf{J} - \mathbf{I}).$$

In other words, $Y_{ii} = \frac{k}{m}$ and $Y_{ij} = \alpha Z_{ij} + \beta$. It is easy to check that conditions 2 is satisfied as long as $\beta \geq \alpha \geq 0$. Condition 3 is satisfied by specifying the diagonal entries of Y to be k/m. Condition 4 translates to

$$k + \alpha\langle Z, \mathbf{J} \rangle + \beta m(m-1) = k^2. \tag{6.9}$$

We also need $\langle X, W \rangle = k(k-1)$, i.e.

$$\alpha m(m-1) + \beta\langle Z, \mathbf{J} \rangle = k(k-1). \tag{6.10}$$

Solve the linear system (6.9) with (6.10) to obtain the pair

$$\alpha = \beta = \frac{k^2 - k}{m^2 - m + \langle Z, \mathbf{J} \rangle} = \frac{k^2}{n^2}(1 + o_P(1))$$

for $k \leq \sqrt{n}/2$.

It remains to ensure condition 1 holds. To show that $Y \succeq 0$ it suffices to show

$$\frac{k}{n} \geq \alpha \|Z\|_{\mathrm{op}}.$$

The operator norm of the *i.i.d.* Radamacher matrix Z is smaller than $2\sqrt{n}$ with high probability. Therefore, the right-hand size can be upper bounded by $(k^2/n^2)\sqrt{n}(2 + o_P(1))$. This means $k \leq \sqrt{n}(1/2 + o_P(1)$.[1] □

6.2.4 Rounding an approximate optimal solution

Unlike linear programs, there are no known solver for semidefinite programs that outputs the exact optimizer in polynomial time. Via the ellipsoid method, an SDP can only be solved in polynomial time up to some accuracy. To be exact, it is possible to produce a feasible solution, in poly$(n, m, 1/\epsilon)$ time, e.g. by the ellipsoid method, whose objective value is at least $(1 - \epsilon)$OPT, where $n =$ number of variables, $m =$ number of constraints and ϵ is the relative accuracy. Thus, in order to obtain a genuinely polynomial-time algorithm, we need to that show, as a sanity check, that the same statistical guarantee can be attained if we only solve the SDP relaxation up to certain relative accuracy $\epsilon = \frac{1}{\mathsf{poly}(n)}$ followed by some simple post processing (rounding); otherwise, it defeats the purpose considering this relaxation.

In the context of the planted clique problem, we showed in Theorem 6.2 that as long as $k \geq C\sqrt{n}$ for some constant C, then with high probability, X^* is the unique optimizer of (6.5). Suppose we found a feasible solution X to (6.5), such that

$$\langle W, X \rangle \geq (1 - \epsilon)\langle W, X^* \rangle.$$

Next, we can apply simple rounding scheme to convert X to $\widehat{X} \in \{0, 1\}^{n \times n}$, where

$$\widehat{X}_{ij} = \begin{cases} 0 & \text{if } X_{ij} \leq \frac{1}{2}, \\ 1 & \text{if } X_{ij} > \frac{1}{2}. \end{cases}$$

Theorem 6.4 (Rounding)**.** *Under the assumptions of Theorem 6.2, if $\epsilon \leq c_1\sqrt{n}/k^3$ for some $c_1 > 0$, then $\widehat{X} = X^*$ with high probability.*

Proof. Suppose, for the sake of contradiction, that $\widehat{X} \neq X^*$. By assumption $\langle W, X \rangle \geq \langle W, X^* \rangle - \delta$, with $\delta = k(k-1)\epsilon$. From the definition of $\widehat{X}$, we know that $\widehat{X} \neq X^*$ means

$$\begin{aligned} \exists (i_0, j_0) \in K \times K, \quad \text{s.t. } X^*_{i_0, j_0} = 1, \quad \text{but } X_{i_0, j_0} \leq \frac{1}{2}; \\ \text{or } \exists (i_1, j_1) \notin K \times K, \quad \text{s.t. } X^*_{i_1, j_1} = 0, \quad \text{but } X_{i_0, j_0} > \frac{1}{2}; \end{aligned}$$

[1] This should be contrasted with the positive result $k \leq \sqrt{n}(2 + o_P(1)$. In fact, this can be improved to $k \leq \sqrt{n}(1 + o_P(1))$ by choosing $\lambda = o(1)$ in the proof of Theorem 6.2.

As a consequence, we have $\|X^* - X\|_F^2 > 1/4$. However, there is no contradiction to the optimality gap yet, because when (i_1, j_1) is an edge, it might balance the loss of the objective value inside the clique.

Next, we use the dual variables to assess the suboptimality. Recall the set of dual certificates constructed in the proof of Theorem 6.2:

$$S \succeq 0, \quad \text{s.t. } S\xi = 0, \quad B \geq 0, \quad \eta, \lambda \in \mathbb{R}.$$

Under the assumptions of Theorem 6.2 we showed

$$\lambda_{n-1}(S) \geq c_2 \sqrt{n}$$

for some $c_2 > 0$ with high probability. Recall that if the first order conditions on the Lagrangian multiplier are satisfied, the Lagrangian $L(X, S, B, \eta, \lambda) = k\eta + k^2\lambda$ does not depend on X. Deduce that

$$\langle W, X^* - X \rangle = \langle S, X \rangle + \langle B, X \rangle.$$

Denote the optimality gap $\delta = \langle W, X^* - X \rangle$, which satisfies $\delta \leq k(k-1)\epsilon$, we have from the above $\langle S, X \rangle \leq \delta$.

Let $u = \xi/\sqrt{k}$ be the (unit) eigenvector of S corresponding to zero eigenvalue. We have

$$S \succeq \lambda_{n-1}(S) \left(\mathbf{I} - uu^\top \right).$$

Therefore by positive semidefiniteness of X,

$$\langle S, X \rangle \geq c_2 \sqrt{n} \langle X, \mathbf{I} - \frac{1}{k} X^* \rangle.$$

Together with the upper bound on $\langle S, X \rangle$ we have

$$\langle X, \mathbf{I} - \frac{1}{k} X^* \rangle \leq \frac{\delta}{c_2 \sqrt{n}} \xLeftrightarrow{\mathsf{Tr}(X) = k} \langle X, X^* \rangle \geq k^2 - \frac{k\delta}{c_2\sqrt{n}}.$$

We are now ready to obtain an upper bound on $\|X^* - X\|_F^2$:

$$\|X^* - X\|_F^2 = \|X^*\|_F^2 + \|X\|_F^2 - 2\langle X, X^* \rangle.$$

The truth $X^* = \xi\xi^\top$ is with Frobenius norm k^2. Again thanks to $X \succeq 0$,

$$\|X\|_F \leq \|X\|_* = \mathsf{Tr}(X) = k.$$

Therefore

$$\|X^* - X\|_F^2 \leq k^2 + k^2 - 2\left(k^2 - \frac{k\delta}{c_2\sqrt{n}} \right) = \frac{2k\delta}{c_2\sqrt{n}} \leq 0.2$$

as long as $\delta < 0.1 c_2 \sqrt{n}/k$, which is satisfied if $\epsilon < 0.1 c_2\sqrt{n}/k^3$. We have arrived a contradiction with $\|X^* - X\|_F^2 \geq \frac{1}{4}$. □

6.2.5 Robustness against monotone adversary

In this section we consider a semi-random model to address the robustness of the methods we analyzed so far, following [FK00, FK01]. we show that the standard SDP relaxation of the hidden clique problem, unlike previously discussed methods like the degree test or the spectral method, is robust with respect to any monotone adversary. An adversary is a (possibly random) modification on the observed data designed to derail certain estimators. A *monotone adversary* in the context of the planted clique problem takes the observed adjacency matrix $A \sim G(n, 1/2, k)$ and it is allowed to arbitrarily delete edges while leaving the clique intact.

An important observation is that although monotone adversary seems to be helpful as it only makes the clique more pronounced, it can break the consistency of both the degree test and spectral method.

- We proved in Section 2.3 that as long as the size of the clique is of order $k \asymp \sqrt{n \log n}$, the degree test consistently recovers the location of the true clique K. To design a monotone adversary that breaks the degree test, simply remove all edges between $i \in K$ and $j \in K^c$. The modified adjacency matrix takes the form

$$\widetilde{A} \sim \begin{bmatrix} 1 & 0 \\ 0 & G(n-k, \frac{1}{2}) \end{bmatrix}.$$

  After modification, the degree of vertex i for $i \in K$ is k; while the degree of i for $i \notin K$ is distributed $Bin(n-k, 1/2)$. The degree test will obviously fail at picking out the location of K.

- To break the spectral method, design the monotone adversary as follows:

$$\widetilde{A} \sim \begin{bmatrix} k\text{-clique} & 0 & 0 \\ 0 & G(\frac{n-k}{2}, \frac{1}{2}) & 0 \\ 0 & 0 & G(\frac{n-k}{2}, \frac{1}{2}) \end{bmatrix}.$$

  The top two eigenvectors of $\widetilde{A}$ would typically be aligned with the two blocks contained in K^c (with $\lambda_1 \approx \lambda_2 \approx \frac{n-k}{4}$), and they are no longer informative about the hidden clique K.

Next we argue that $\widehat{X}^{\text{SDP}}$ is automatically robust against monotone adversary. Denote the corresponding W after modification as $\widetilde{W}$. Then for all feasible $X \neq X^*$, we have

$$\langle \widetilde{W}, X \rangle \leq \langle W, X \rangle < \langle W, X^* \rangle = \langle \widetilde{W}, X^* \rangle.$$

where the strict inequality is because X^* is the unique global maximizer of the original problem, and the last equality is because the adversary kept the planted clique intact. In other words, whenever X^* is the maximizer under W, it remains the maximizer under $\widetilde{W}$.

6.3 Convexified MLE

Recall the lifted version of the MLE (6.4). The trace constraint $\mathsf{Tr}(X) = k$ is equivalent to $\sum_i \lambda_i(X) = k$. By positive semideniteness of X all eigenvalues of X are nonnegative and are also singular values. Therefore the nuclear norm $\|X\|_* = \sum_i \lambda_i(X) = k$.

Following [CX16], the convexified MLE is obtained by relaxing the trace constraint in (6.4) to an inequality constraint and dropping the rank-one constraint, hence making the feasible set convex. Define

$$\begin{aligned}\widehat{X}_{\text{convex}} = \arg\max_{X} \ & \langle W, X \rangle \\ \text{s.t.} \ & \|X\|_* \le k, \\ & 0 \le X \le \mathbf{J}, \\ & \langle X, \mathbf{J} \rangle = k^2. \end{aligned} \tag{6.11}$$

Remark 6.4 (Matrix norm revisited)**.** We continue the discussion of matrix norms in Section 3.1.3. Recall that the nuclear norm $\|X\|_*$ is the ℓ_1 norm of the singular values of X, which when X is symmetric, equals the summation of the absolute values of the eigenvalues of X.

More generally the Schatten p-norm of a matrix X (denoted as $\|X\|_{S_p}$) is defined as the ℓ_p norm of the singular values of X. For instance, we have

$$\|X\|_{S_2} = \|X\|_F, \quad \text{the Frobenius norm of } X;$$

$$\|X\|_{S_\infty} = \sigma_{\max}(X) = \|X\|_{\text{op}}, \quad \text{the operator norm of } X;$$

$$\|X\|_{S_1} = \|X\|_*, \quad \text{the nuclear norm of } X.$$

In general, the dual norm of a norm $\|\cdot\|$ is defined as

$$\|\cdot\|_* = \max_{y:\|y\|\le 1} \langle \cdot, y \rangle . \tag{6.12}$$

The subscript $_*$ here stands for the dual norm, not to be confused with the nuclear norm.

From duality between the usual ℓ_p vector norms, it is easy to derive that

$$\left(\|\cdot\|_{S_p}\right)_* = \|\cdot\|_{S_q},$$

where $1/p + 1/q = 1$. In particular, we have for that the nuclear norm is dual with the operator norm. Hence for the nuclear norm of X,

$$\|X\|_* = \max_{\|Y\|_{\text{op}}\le 1} \langle X, Y \rangle .$$

6.3.1 SDP formulation

The following proposition allows us to rewrite the program (6.11) as an SDP in the standard form:

Proposition 6.3. *For a matrix $X \in \mathbb{R}^{m\times n}$, $\|X\|_* \le 1$ if and only if there exists $W_1 \in \mathbb{R}^{m\times m}$, $W_2 \in \mathbb{R}^{n\times n}$, such that*

$$\mathsf{Tr}(W_1) + \mathsf{Tr}(W)_2 \le 2, \quad \textit{and} \tag{6.13}$$

$$\begin{bmatrix} W_1 & X \\ X^\top & W_2 \end{bmatrix} \succeq 0.$$

Proof. (*"if" part*) the PSD assumption (6.13) implies that

$$\begin{bmatrix} u^\top, & -v^\top \end{bmatrix} \begin{bmatrix} W_1 & X \\ X^\top & W_2 \end{bmatrix} \begin{bmatrix} u \\ -v \end{bmatrix} \ge 0 \quad \forall u, v.$$

Choose $u = u_i$ to be the i'th left singular vector of X, and $v = v_i$ to be the i'th right singular vector of X. We have

$$2u_i^\top X v_i \le u_i^\top W_i u_i + v_i^T W_2 V_i.$$

Note that the left-hand side is equal to twice the i'th singular value of X. Take summation of the inequality above over $1 \le i \le r \triangleq \min\{m, n\}$ to deduce that

$$2\|X\|_* \le \langle W_1, \sum u_i u_i^\top \rangle + \langle W_2, \sum v_i v_i^\top \rangle.$$

Note that W_1 is positive semidefinite (simply take $v = 0$). Similarly we have $W_2 \succeq 0$. Combining with the fact that $\sum u_i u_i^\top \preceq \mathbf{I}$, $\sum v_i v_i^\top \preceq \mathbf{I}$, and $\langle A, B\rangle \ge 0$ when $A, B \succeq 0$, we have that the right-hand side of the display above is bounded by $\mathsf{Tr}(W_1) + \mathsf{Tr}(W_2) \le 2$.

(*"only if" part*) Suppose $X = U\Sigma V^\top$ is the singular value decomposition of X, where $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n \times r}$, and Σ is a $r \times r$ diagonal matrix with non-negative entries. Choose

$$W_1 = U\Sigma U^\top, \quad W_2 = V\Sigma V^\top.$$

First check the trace condition:

$$\mathsf{Tr}(W_1) + \mathsf{Tr}(W_2) = 2\mathsf{Tr}(\Sigma) = 2\|X\|_* \le 2.$$

Next check the positive semidefinite condition. For all $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$, we have

$$\begin{aligned} &\begin{bmatrix} u^\top, & -v^\top \end{bmatrix} \begin{bmatrix} W_1 & X \\ X^\top & W_2 \end{bmatrix} \begin{bmatrix} u \\ -v \end{bmatrix} \\ =&\begin{bmatrix} u^\top, & -v^\top \end{bmatrix} \begin{bmatrix} U\Sigma U^\top & U\Sigma V^\top \\ V\Sigma U^\top & V\Sigma V^\top \end{bmatrix} \begin{bmatrix} u \\ -v \end{bmatrix} \\ =&(U^\top u - V^T v)^T \Sigma (U^\top u - V^T v) \ge 0, \end{aligned}$$

since Σ is a diagonal matrix with nonnegative diagonal entries. □

6.3.2 Subgradient and norms

Suppose a function $f : \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable. Then the gradient $\nabla f(x_0)$ at each x_0 defines an affine minorant:

$$f(x) \ge f(x_0) + \langle \nabla f(x_0), x - x_0\rangle, \quad \forall x.$$

When f is not differentiable, we could work with the *sub-gradients* of f. A sub-gradient of f at x_0 is defined as any element of

$$\partial f(x_0) = \left\{ u \in \mathbb{R}^d : f(x) \ge f(x_0) + \langle u, x - x_0\rangle, \quad \forall x \right\}.$$

For example, $\partial|\cdot|(0) = [-1, 1]$. Note that for smooth f, subgradient and gradient coincide, i.e., $\partial f(x_0) = \{\nabla f(x_0)\}$.

Next we find the sub-gradient of the nuclear norm $\|\cdot\|_*$. Being the ℓ_1 norm of the singular values of X, it is not differentiable. In fact, it is easy to show that the subgradients of any norm are precisely those vectors for which the duality (6.12) is tight:

Proposition 6.1 (Subgradient of norms). *Let $\|\cdot\|$ and $\|\cdot\|_*$ be a pair of dual norms. Then*

$$\partial\|\cdot\|_*(x_0) = \begin{cases} \{y : \langle x_0, y\rangle = \|x_0\|_*, \|y\| = 1\} & x_0 \ne 0 \\ \{y : \|y\| \le 1\} & x_0 = 0 \end{cases}. \tag{6.14}$$

Proof. "$\supset$": For any y in the RHS, and for any x, $\|x_0\|_* + \langle y, x - x_0\rangle = \langle y, x\rangle \leq \|y\|\|x\|_* \leq \|x\|_*$.

"$\subset$": Suppose $\|x\|_* \geq \|x_0\|_* + \langle y, x - x_0\rangle$ for all x. Then

$$\langle y, x_0\rangle - \|x_0\|_* \geq \sup_{x \in \mathbb{R}^d} \langle y, x\rangle - \|x\|_* \overset{(6.12)}{=} \begin{cases} 0 & \|y\| \leq 1 \\ +\infty & \|y\| > 1 \end{cases}$$

This implies that, necessarily, $\|y\| \leq 1$ and $\langle y, x_0\rangle = \|x_0\|_*$. If $x_0 \neq 0$, then this further means $\|y\| = 1$. □

Specializing to the nuclear norm, we have the following (we only need $\supset$ direction for analyzing convexified MLE next):

Corollary 6.1. *For any $X \neq 0$, denote its SVD as $X = U\Sigma V^\top$. Then*

$$\partial\|\cdot\|_*(X) = \{UV^\top + P^\perp(Y) : \|Y\|_{\text{op}} \leq 1\},$$

where $P^\perp(Y) = (I - UU^\top)Y(I - VV^\top)$ is the projection onto the orthogonal complement of the linear subspace $T = \{UA^\top + BV^\top : A \in \mathbb{R}^{n\times r}, B \in \mathbb{R}^{n\times r}\}$. Note the orthogonal complement of T consists of matrices whose column span is orthogonal to that of X (i.e., span(U)) and whose row span is orthogonal to that of X (i.e., span(V)).

Proof. Exercise. □

6.3.3 Statistical guarantee: primal proof

Theorem 6.5. *Assume that $k \geq C_0\sqrt{n}$ for some constant C_0. Whp, the unique solution to (6.11) is given by $= X^*$.*

Proof. We will show that whp, the objective function of any feasible $X \neq X^*$ is inferior, i.e., $\langle X, W\rangle < \langle X^*, W\rangle$. To this end we will show that whp, for any feasible X,

$$\langle X^* - X, W\rangle \gtrsim \|X - X^*\|_{\ell_1} \tag{6.15}$$

As before, $\xi = \mathbb{1}_{K^*}$ be the indicator vector of the hidden clique and $u = \frac{1}{\sqrt{k}}\xi$. Then $X^* = \xi\xi^\top$, which is almost the same as $\mathbb{E}W = X^* - \text{diag}(\xi)$. Let $E = uu^\top = \frac{1}{k}X^*$ denote the projection matrix onto span(ξ). As in Corollary 6.1, define the projection operator $P^\perp(Y) = (I - E)Y(I - E)$ and $P(Y) = Y - P^\perp(Y) = EY + YE - EYE$.

Write

$$\langle X^* - X, W\rangle = \underbrace{\langle X^* - X, X^*\rangle}_{(a)} + \underbrace{\langle X^* - X, P^\perp(W - X^*)\rangle}_{(b)} + \underbrace{\langle X^* - X, P(W - X^*)\rangle}_{(c)}.$$

Then

(a): This term dominates:

$$(a) = \sum_{(i,j)\in K^*\times K^*} (1 - X_{ij}) = \frac{1}{2}\|X - X^*\|_{\ell_1}$$

where the last step is due to $\langle X - X^*, \mathbf{J}\rangle = 0$ (from feasibility) so that $\sum_{(i,j)\in K^*\times K^*}(1 - X_{ij}) = \sum_{(i,j)\notin K^*\times K^*} X_{ij}$.

(b): We find a subgradient of $\|\cdot\|_*$ at X^*. Note that X^* is rank-one, so by Corollary 6.1, $\partial\|\cdot\|_*(X^*) = \{E + P^\perp(Y) : \|Y\|_{\mathrm{op}} \le 1\}$. Now,

$$\begin{aligned} 0 &\ge \|X\|_* - \|X^*\|_* && \text{by feasibility} \\ &\ge \underbrace{\langle X - X^*, E\rangle}_{= -\frac{1}{2k}\|X-X^*\|_{\ell_1}} + \langle X - X^*, P^\perp(Y)\rangle. && \text{by subgrad.} \end{aligned}$$

Taking $Y = \pm \frac{W - X^*}{\|W - X^*\|_{\mathrm{op}}}$, we get

$$|(b)| \le \frac{\|W - X^*\|_{\mathrm{op}}}{2k}\|X - X^*\|_{\ell_1}$$

(c): By duality, we have

$$|(c)| \le \|P(W - X^*)\|_{\ell_\infty}\|X - X^*\|_{\ell_1}.$$

Combining (a), (b) and (c), we get

$$\langle X^* - X, W\rangle \ge \left(\frac{1}{2} - \frac{\|W - X^*\|_{\mathrm{op}}}{2k} - \|P(W - X^*)\|_{\ell_\infty}\right)\|X - X^*\|_{\ell_1}.$$

Here $\|W - X^*\|_{\mathrm{op}} \le \|W - \mathbb{E}[W]\|_{\mathrm{op}} + \|\mathrm{diag}(\xi)\|_{\mathrm{op}} \le C\sqrt{n} + 1$ for some constant C whp, by Theorem 4.2. Since $k = C_0\sqrt{n}$ for some sufficiently large C_0, it suffices to show that $\|P(W - X^*)\|_{\ell_\infty} = o(1)$ whp.

Note that $W - X^* = W - \mathbb{E}[W] - \mathrm{diag}(\xi)$ and $P(\mathrm{diag}(\xi)) = E\mathrm{diag}(\xi) + \mathrm{diag}(\xi)E - E\mathrm{diag}(\xi)E = E$ so

$$\|P(W - X^*)\|_{\ell_\infty} \le \|P(W - \mathbb{E}[W])\|_{\ell_\infty} + \underbrace{\|P(\mathrm{diag}(\xi)\|_{\ell_\infty}}_{1/k}$$

Next, for any Y, $P(Y) = EY + YE - EYE$, where $\|EYE\|_{\ell_\infty} \le \|EY\|_{\ell_\infty}\|E\|_{\ell_\infty \to \ell_\infty} \le \|EY\|_{\ell_\infty}$. Thus

$$\|P(Y)\|_{\ell_\infty} \le 3\|EY\|_{\ell_\infty},$$

where we used the fact that for any symmetric Y (e.g. $W - \mathbb{E}[W]$) whose support is disjoint from that of E,

$$\|EY\|_{\ell_\infty} = \|YE\|_{\ell_\infty} = \frac{1}{k}\max_{i \notin K^*}\sum_{j \in K^*} Y_{ij}.$$

Furthermore, for each i, $\mathbb{P}\left[|\sum_{j\in K^*}(W_{ij} - \mathbb{E}W_{ij})| \ge \sqrt{k}t\right] \le \exp(-ct^2)$, by Hoeffding's inequality (Lemma 2.2). Therefore whp, $\|P(W - \mathbb{E}[W])\|_{\ell_\infty} \le 3\|E(W - \mathbb{E}[W])\|_{\ell_\infty} \le \sqrt{\frac{C\log n}{k}}$, and we are done. □

Exercise: Give a dual-based proof of Theorem 6.5 by identifying the appropriate dual certificates.

Part II

Planted partition model

§ 7. Detection threshold for SBM

7.1 Planted partition model and overview

In the second part of the course, we will study the problem of *community detection* in a broad sense. Consider the following abstract *planted partition model*, where a matrix $A = (A_{ij})_{1 \le i < j \le n}$ is observed whose distribution depends on the latent labels $\sigma = (\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$, such that

$$A_{ij} \sim \begin{cases} P & \sigma_i = \sigma_j \\ Q & \sigma_i \neq \sigma_j \end{cases}. \tag{7.1}$$

Given A, the goal is to recover the labels σ accurately.

Two prominent special cases are the following:

Stochastic block model (SBM) Here $P = \text{Bern}(p)$ and $Q = \text{Bern}(q)$. In this case the set of vertices $[n]$ is partitioned into two communities $V_+ = \{i : \sigma_i = +1\}$ and $V_- = \{i : \sigma_i = -1\}$, and A is the adjacency matrix of a random graph, such that two nodes i and j are connected with probability p if they belong to the same community, and with probability q if otherwise. The case of $p > q$ is referred to as "assortative", such as friendship networks, and $p < q$ as "disassortative", such as predator-prey networks.

The community structure is determined by the vector σ, which, depending on the problem formulation, could either be fixed or random. We will frequently consider special cases:

- iid model: Each σ_i is equally likely to be ± (Rademacher) and independently.
- exact bisection: $|V_+| = |V_-| = n/2$ (when n is even) and the partition is chosen uniformly at random from all bisections.

Typically these two models behave very similarly.

Spiked Wigner model (Rank-one deformation) Here $P = N(\sqrt{\frac{\lambda}{n}}, 1)$ and $Q = N(-\sqrt{\frac{\lambda}{n}}, 1)$. In matrix notation,

$$A = \sqrt{\frac{\lambda}{n}} \sigma \sigma^\top + Z, \tag{7.2}$$

where Z is such that $\{Z_{ij} : 1 \le i < j \le n\}$ are iid $N(0, 1)$. Therefore A can be viewed is a rank-one perturbation of a Gaussian Wigner matrix.

As opposed to the treatment of the planted clique problem in Part I, we will be focusing on

- Sharp threshold, i.e., finding the exact constant in the fundamental limit (and achieving them with fast algorithms).

- "Sparse" graphs, where the edge density tends to zero (at different speed), unlike the hidden clique model $G(n, \frac{1}{2}, k)$

We will focus on the following three formulations (recovery guarantees):

Detection Here there is a null model. For example,

- For spiked Wigner model, the null hypothesis is A is iid Gaussian. The sharp threshold is given by $\lambda = 1$, in the sense that for any fixed ϵ, it is possible to test the hypotheses with vanishing error probability if $\lambda \geq 1 + \epsilon$, and impossible if $\lambda \leq 1 - \epsilon$.
- For SBM with bisection, we want to test against the null hypothesis of no community structure, that is, an Erdős-Rényi graph $G(n, \frac{p+q}{2})$ with the same average degree. The most interesting regime is bounded average degree $p = \frac{a}{n}, q = \frac{b}{n}$ for constants a, b, and the sharp threshold is given by $\frac{(a-b)^2}{2(a+b)} = 1$.

Correlated (weak) recovery Here and below, there is no null model. The goal is to recover the community structure (labels) strictly better than random guessing. Let $\widehat{\sigma} = \widehat{\sigma}(A)$ be the estimator. Its overlap with the true labels σ is $|\langle \widehat{\sigma}, \sigma \rangle|$ and the number of misclassification errors (up to a global sign flip) is expressed as

$$\ell(\sigma, \widehat{\sigma}) = \min_{s \in \{\pm 1\}} \|\widehat{\sigma} - s\sigma\|_1 = n - |\langle \widehat{\sigma}, \sigma \rangle| \,.$$

In the iid setting, random guessing would yield, by CLT, $|\langle \widehat{\sigma}, \sigma \rangle| = O_P(\sqrt{n})$ and $\mathbb{E}[|\langle \widehat{\sigma}, \sigma \rangle|] = o(n)$. The goal of weak recovery is to achieve a positive correlation, namely

$$\mathbb{E}[|\langle \widehat{\sigma}, \sigma \rangle|] = \Omega(n).$$

Although in general detection and correlated recovery are two different problems, for both SBM and spiked Wigner the thresholds coincide. In fact, for certain models one can have a generic reduction between the problems (e.g. spiked Wigner, see Homework).

(Almost) exact recovery *Almost exact recovery* means achieving a vanishing misclassification rate: $\mathbb{E}\ell(\sigma, \widehat{\sigma}) = o(n)$. Typically the sharp threshold is expressed in terms of Hellinger distance as $H^2(P, Q) \gg \frac{1}{n}$, where $H^2(P, Q) \triangleq \mathbb{E}_Q \left[\left(\sqrt{P/Q} - 1 \right)^2 \right] = \int (\sqrt{P} - \sqrt{Q})^2$.

Exact recovery means $\ell(\sigma, \widehat{\sigma}) = 0$ with probability tending to 1. Typically the sharp threshold is given by $H^2(P, Q) = \frac{(2+\epsilon)\log n}{n}$.

A more statistical flavored question is to characterize the optimal (in the sense of minimax) misclassification rate $\frac{1}{n}\ell(\sigma, \widehat{\sigma})$, which typically behaves as $\exp(-\frac{H^2(P,Q)}{2})$.

7.2 Detection threshold for SBM

It will be useful to recall the basics on hypothesis testing and the relevant information measures from Appendix A. Here, we want to test the hypothesis

$$H_0 : G \sim G\left(n, \frac{p+q}{2}\right) \quad \text{vs.} \quad H_1 : G \sim \text{SBM}(n, p, q).$$

Under the SBM model, we assume the the labels $\sigma = (\sigma_1, \dots, \sigma_n)$ are either iid $\mathrm{Rad}(\frac{1}{2})$, or drawn uniformly at random from all bisections. The detection problem is non-trivial in the regime of bounded average degree:

$$p = \frac{a}{n},\ q = \frac{b}{n}, \tag{7.3}$$

where a, b are constants.

Theorem 7.1. *If $\frac{(a-b)^2}{2(a+b)} > 1$, detection is possible, in the sense of total variation distance that*

$$\mathrm{TV}(\mathrm{Law}(G|H_0), \mathrm{Law}(G|H_1)) \to 1. \tag{7.4}$$

Conversely, if $\frac{(a-b)^2}{2(a+b)} \le 1$, detection is impossible, in the sense that

$$\mathrm{TV}(\mathrm{Law}(G|H_0), \mathrm{Law}(G|H_1)) \le 1 - \Omega(1). \tag{7.5}$$

We start with the impossibility part. For non-detection it is enough to show

$$\chi^2(\mathrm{Law}(G|H_0)||\mathrm{Law}(G|H_1)) = O(1). \tag{7.6}$$

The calculation for SBM can be carried out in the general setting of (7.1), where P and Q denote the weight distribution for edges within and without communities. For each label $\sigma \in \{\pm 1\}^n$, the distribution of the adjacency matrix is

$$P_\sigma = \mathrm{Law}(A|\sigma) = \prod_{i<j} (P\mathbb{1}_{\{\sigma_i = \sigma_j\}} + Q\mathbb{1}_{\{\sigma_i \neq \sigma_j\}}) = \prod_{i<j} \left(\frac{P+Q}{2} + \frac{P-Q}{2} \sigma_i \sigma_j \right). \tag{7.7}$$

Let P_1 denote the marginal distribution of A, namely, P_σ averaged over the random labels σ:

$$P_1 = \sum_\sigma P(\sigma) \cdot P_\sigma$$

where $P(\sigma)$ denotes the PMF of σ. We aim to show P_1 is not perfectly distinguishable from the null distribution $P_0 \equiv \prod_{i<j} \frac{(P+Q)}{2}$. To this end, we bound their χ^2-divergence by applying Lemma A.3. Fix two assignment $\sigma, \widetilde{\sigma} \in \{\pm 1\}^n$. Then

$$\begin{aligned}
G(\sigma, \widetilde{\sigma}) &\equiv \int \frac{P_\sigma P_{\widetilde{\sigma}}}{P_0} \\
&= \int \prod_{i<j} \frac{\left(\frac{P+Q}{2} + \frac{P-Q}{2}\sigma_i\sigma_j\right)\left(\frac{P+Q}{2} + \frac{P-Q}{2}\widetilde{\sigma}_i\widetilde{\sigma}_j\right)}{\frac{P+Q}{2}} \\
&= \prod_{i<j} \left[\underbrace{\int \frac{P+Q}{2}}_{=1} + \underbrace{\int \frac{P-Q}{2}}_{=0} \sigma_i\sigma_j + \underbrace{\int \frac{P-Q}{2}}_{=0} \widetilde{\sigma}_i\widetilde{\sigma}_j + \underbrace{\int \frac{(P-Q)^2}{2(P+Q)}}_{\triangleq \rho} \sigma_i\sigma_j\widetilde{\sigma}_i\widetilde{\sigma}_j \right] \\
&= \prod_{i<j} (1 + \rho\sigma_i\sigma_j\widetilde{\sigma}_i\widetilde{\sigma}_j) \\
&\le \exp\left(\rho \sum_{i<j} \sigma_i\sigma_j\widetilde{\sigma}_i\widetilde{\sigma}_j \right) \le \exp\left(\frac{\rho}{2} \langle \sigma, \widetilde{\sigma} \rangle^2 \right)
\end{aligned}$$

Thus, by Lemma A.3, we have

$$\chi^2(P_1\|P_0) + 1 \le \mathbb{E}_{\sigma,\widetilde{\sigma}}\left[\exp\left(\frac{\rho}{2}\langle\sigma,\widetilde{\sigma}\rangle^2\right)\right],$$

where $\widetilde{\sigma}$ is an iid copy ("replica") of σ.

For SBM(n,p,q), we have $P = \text{Bern}(p)$ and $Q = \text{Bern}(q)$. Under the scaling (7.3), we have

$$\rho = \frac{(p-q)^2}{2(p+q)} + \frac{((1-p)-(1-q))^2}{2(1-p+1-q)} = \frac{\tau + o(1)}{n}, \qquad \tau \triangleq \frac{(a-b)^2}{2(a+b)}.$$

Next we consider two situations:

Independent labels: $\sigma, \widetilde{\sigma} \overset{iid}{\sim} \{\pm1\}^n$. By CLT, $\frac{1}{\sqrt{n}}\langle\sigma,\widetilde{\sigma}\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i\widetilde{\sigma}_i \xrightarrow{\text{D}} Z \sim N(0,1)$. Assuming convergence of MGF (see Lemma 7.1 next) and invoking the MGF of $\chi^2(1)$ distribution, we have

$$\begin{aligned}\chi^2(P_1\|P_0) + 1 &= \mathbb{E}\exp\left(\frac{\tau+o(1)}{2n}\langle\sigma,\widetilde{\sigma}\rangle^2\right)\\ &\to \mathbb{E}\left(\frac{\tau}{2}Z^2\right)\\ &= \begin{cases}\infty & \text{if } \tau \ge 1\\ \frac{1}{\sqrt{1-\tau}} & \text{if } \tau < 1.\end{cases}\end{aligned} \tag{7.8}$$

Exact bisection: Let us consider the case where $\sigma, \widetilde{\sigma}$ are drawn iid and uniformly at random from the set $\{\theta \in \{\pm1\}^n : \sum \theta_i = 0\}$. For simplicity, write

$$\sigma = 2\xi - \mathbf{1}, \quad \widetilde{\sigma} = 2\widetilde{\xi} - \mathbf{1},$$

Then both $\xi, \widetilde{\xi}$ are iid uniform random $\frac{n}{2}$-sparse binary vectors, and $\langle\sigma,\widetilde{\sigma}\rangle = 4\langle\xi,\widetilde{\xi}\rangle - n$. Therefore,

$$\langle\xi,\widetilde{\xi}\rangle \sim \text{Hypergeometric}(n, \frac{n}{2}, \frac{n}{2}),$$

which means (check!)[1]

$$\frac{\langle\xi,\widetilde{\xi}\rangle - \frac{n}{4}}{\sqrt{\frac{n}{16}}} \xrightarrow{\text{D}} Z \sim N(0,1).$$

Thus the dichotomy (7.8) applies to bisection as well.

To pass from weak convergence to convergence of the MGF, the following lemma is useful:

Lemma 7.1 (Convergence of MGF)**.** *Assume that $X_n \xrightarrow{\text{D}} X$. Let $M_n(t) = \mathbb{E}\exp(tX_n)$ and $M(t) = \mathbb{E}\exp(tX)$. If there exists some constant $\alpha > 0$ such that*

$$\sup_n P(|X_n| > x) \le \exp(-\alpha x)$$

for all $x > 0$, then $M_n(t) \to M(t)$ for all $|t| < \alpha$.

Remark 7.1.

- The critical case of $\frac{(a-b)^2}{2(a+b)} = 1$ also implies non-detection. Proving this is outside the scope of this section as the χ^2 truly blows up.
- The threshold of the spiked Wigner model (7.2) is given by $\lambda = 1$. This can be proved by the same second moment method (homework).

[1] Note that the variance of Hypergeometric$(n, \frac{n}{2}, \frac{n}{2})$ is exactly half of its counterpart Binom$(\frac{n}{2}, \frac{1}{2})$. Why? Think about sampling with and without replacements. See, e.g., [Fel70, p. 194]) for the central limit theorem for hypergeometric distributions.

7.3 Test by counting cycles

Below we describe a test for

$$H_0 : G \sim G(n, \frac{p+q}{2}) \text{ vs. } H_1 : G \sim SBM(n, p, q).$$

that achieves the sharp threshold in Theorem 7.1, following [MNS15]. We will consider the labels being iid Rad($\frac{1}{2}$). Note that the average vertex degree is matched under H_0 and H_1. The test is based on counting "short" cycles – by short we mean much shorter than the longest cycle, but the length still need to be slowly growing with n. As there is no generic polynomial-time algorithm for counting k-cycles for growing k, in the next section we make it polynomial-time relying on the randomness of the graph.

Consider the number of k-cycles (not induced cycles) as the test statistic, denoted by X_k. As a warmup, consider the behavior of X_k in $G(n, \frac{d}{n})$. Then by union bound,

$$\mathbb{P}(X_k > 0) \le \mathbb{E}[X_k] = \binom{n}{k} k! \frac{1}{2k} \left(\frac{d}{n}\right)^k \le d^k,$$

where the overcounting factor $2k$ is the number of symmetries (automorphisms) of C_k, namely, cyclic shift and flip. Thus there are no cycles of growing length if $d < 1$. Of course, this first-moment calculation does not tell us about the existence of k-cycles. Nevertheless, it is known that if $d \ge 1$, the longest cycle is of length $\Omega(n)$ [Bol01, Chap. 8].

Now let us get back to the original problem of testing $G(n, \frac{a+b}{2n})$ versus SBM$(n, \frac{a}{n}, \frac{b}{n})$ in Theorem 7.1. Assume that $a > b$. Define

$$s = \frac{a-b}{2}, \quad d = \frac{a+b}{2}.$$

Then the condition $\frac{(a-b)^2}{2(a+b)} > 1$ is the same as $s^2 > d$. Since $d > s$, this implies that $s > 1$ and $a > 2$.

Intuition: For k not too big (we will see that $k = o(\log n / \log\log n)$ is OK), X_k has a Poisson limit under both model with different parameters:

$$\text{Under } H_0\text{:} \quad X_k \xrightarrow{\text{D}} \text{Poi}\left(\frac{d^k}{2k}\right)$$

$$\text{Under } H_1\text{:} \quad X_k \xrightarrow{\text{D}} \text{Poi}\left(\frac{d^k + s^k}{2k}\right).$$

This suggests that we can distinguish the two hypothesis using X_k as a test statistic. Specifically, consider the test $\mathbb{1}\left\{X_k \le \frac{d^k + s^k/2}{2k}\right\}$. We can simply use Chebyshev's inequality to bound the Type-I and Type-II errors. As such, it suffices to compute the mean and variance of X_k under H_0 and H_1. We will show that

$$\text{Under } H_0\text{:} \quad \mathbb{E}X_k = (1+o(1))\frac{d^k}{2k}, \quad \text{Var}X_k = (1+o(1))\frac{d^k}{2k},$$

$$\text{Under } H_1\text{:} \quad \mathbb{E}X_k = (1+o(1))\frac{d^k + s^k}{2k}, \quad \text{Var}X_k = (1+o(1))\frac{d^k + s^k}{2k}.$$

Under the condition $s^2 > d$, we have

$$\mathbb{E}_1[X_k] - \mathbb{E}_0[X_k] \gg \sqrt{\text{Var}_0(X_k) + \text{Var}_1(X_k)},$$

provided that $k = \omega(1)$. Hence, it follows from the Chebyshev's inequality that, for the test $\mathbb{1}\left\{X_k \le \frac{d^k + s^k/2}{2k}\right\}$, the sum of Type-I and Type-II errors is $o(1)$.

7.3.1 First moment calculation

Under H_0. First we note that

$$X_k = \frac{1}{2k} \sum_{\substack{v_1,\ldots,v_k;\\ \text{all ordered } k\text{-tuple} \\ \text{from } V(G)}} \mathbb{1}_{\{v_1 \sim v_2, v_2 \sim v_3, \ldots, v_k \sim v_1\}},$$

which implies

$$\mathbb{E}X_k = \frac{1}{2k} \underbrace{\binom{n}{k} k!}_{\triangleq [n]_k} \underbrace{\mathbb{P}\{v_1 \sim v_2, v_2 \sim v_3, \ldots, v_k \sim v_1\}}_{=(\frac{d}{n})^k} \approx \frac{1+o(1)}{2k} d^k \tag{7.9}$$

under H_0, where the last equality holds provided $k = o(\sqrt{n})$ (Why? Think about birthday problem).

Under H_1. We just need to recompute the probability in (7.9), which now depends on the labels of the vertices. Consider the adjacency matrix A. Then given any two vertices v_i, v_{i+1}, we have

$$A_{v_i, v_{i+1}} \sim \begin{cases} \text{Bern}(p) \text{ if } \sigma_i = \sigma_{i+1} \\ \text{Bern}(q) \text{ if } \sigma_i \neq \sigma_{i+1}. \end{cases}$$

Given any k-tuple $\{v_1, v_2, \ldots, v_k\}$ of vertices, suppose N denotes the number of disagreements of adjacent labels, given by

$$N = \sum_{i=1}^{k} \mathbb{1}_{\{\sigma(v_i) \neq \sigma(v_{i+1})\}}$$

with $k+1$ understood as 1 circularly. Write

$$N = \underbrace{\sum_{i=1}^{k-1} \mathbb{1}_{\{\sigma(v_i) \neq \sigma(v_{i+1})\}}}_{\triangleq T} + \underbrace{\mathbb{1}_{\{\sigma(v_k) \neq \sigma(v_1)\}}}_{\triangleq S}.$$

Then we have $T \sim \text{Binom}(k-1, \frac{1}{2})$ and

$$S = \begin{cases} 0 & T \text{ is even} \\ 1 & T \text{ is odd} \end{cases}$$

is a parity bit, so that $N = S + T$ is always even.

It is clear that conditioned on $N = m$, the probability of $v_1, \ldots, v_k$ forming a cycle is $q^m p^{k-m}$. Note that

$$\mathbb{P}(N = m) = \begin{cases} 0 & m \text{ odd} \\ \mathbb{P}(\text{Binom}(k-1, \frac{1}{2}) = m-1 \text{ or } m) = \binom{k}{m} 2^{-k+1} & m \text{ even} \end{cases}.$$

Thus

$$\begin{aligned}
\mathbb{P}(v_1 \sim v_2, v_2 \sim v_3, \ldots, v_k \sim v_1) &= \sum_{m=0}^{k} q^m p^{k-m} \cdot \mathbb{P}(N = m) \\
&= \sum_{\substack{m=0 \\ m \text{ even}}}^{k} q^m p^{k-m} \binom{k}{m} 2^{-k+1} \\
&= \sum_{m=0}^{k} \frac{(-q)^m p^{k-m} + q^m p^{k-m}}{2} \binom{k}{m} 2^{-k+1} \\
&= \left(\frac{p+q}{2}\right)^k + \left(\frac{p-q}{2}\right)^k = n^{-k}(s^k + d^k).
\end{aligned}$$

Thus, under H_1,

$$\begin{aligned}
\mathbb{E}(X_k) &= \frac{[n]_k}{2k} \left\{ \left(\frac{p+q}{2}\right)^k + \left(\frac{p-q}{2}\right)^k \right\} \\
&\overset{k=o(\sqrt{n})}{=} \frac{1+o(1)}{2k} \left(s^k + d^k\right).
\end{aligned}$$

7.3.2 Second moment calculation

We only consider the variance under the null model, as the computation under the planted model is similar. Given an ordered k-tuple of vertices (viewed as a k-cycle in the complete graph) $T = (v_1, \ldots, v_k)$, define $b_T = \mathbb{1}\{v_1 \sim v_2, \cdots, v_k \sim v_1\}$. Then under H_0, we have

$$\operatorname{Var}(X_k) = \frac{1}{4k^2} \sum_{T,T'} \operatorname{Cov}(b_T, b_{T'}) = \frac{1}{4k^2} \left(\sum_{T,T':T=T'} \underbrace{\operatorname{Var}(b_T)}_{\leq \mathbb{E}[b_T] = (d/n)^k} + \sum_{\substack{T \neq T' \\ T \cap T' \neq \emptyset}} \operatorname{Cov}(b_T, b'_T) \right),$$

where $T = T'$ means they form the same k-cycle in the complete graph, and the last equality holds because if $T \cap T' = \emptyset$ (no common edge), then $\operatorname{Cov}(b_T, b'_T) = 0$.

Consider two distinct k-cycles T and T' that are overlapping. Let

$$\ell = \text{number of common edges}, \quad v = \text{number of common vertices}.$$

Note that

- $\operatorname{Cov}(b_T, b_{T'}) \leq \mathbb{E}[b_T b_{T'}] = p^{2k-\ell}$.
- Crucially,
$$v \geq \ell + 1.$$
This is because the intersection of two cycles is a forest (each connected component is a path), so that $v = \ell + \mathrm{cc} \geq \ell + 1$, where cc denotes the number of connected components of $T \cap T'$.
- Given v, the number of such pairs of (T, T') is at most $[n]_k \binom{k}{v}\binom{n-k}{k-v} k! = [n]_{2k-v} [k]_v \binom{k}{v}$. To see this, note that there are at most $[n]_k$ different choices of T. Also, given T, there are $\binom{k}{v}\binom{n-k}{k-v}$ different choices of the vertex set of T'. Finally, $k!$ counts all the possible orderings of vertices in T'.

Combining these, we get

$$
\begin{aligned}
\sum_{\substack{T\neq T'\\ T\cap T'\neq\emptyset}} \mathrm{Cov}(b_T, b'_T) &\le \sum_{\ell=1}^{k-1}\sum_{v\ge \ell+1} [n]_{2k-v}[k]_v \binom{k}{v}\left(\frac{d}{n}\right)^{2k-\ell} \\
&\le \sum_{\ell=1}^{k-1}\sum_{v\ge \ell+1} n^{2k-\ell-1}k^k \binom{k}{v}\left(\frac{d}{n}\right)^{2k-\ell} \\
&\le \frac{1}{n}k^{k+1}2^k d^{2k} \\
&= o(1), \quad \text{provided that } k = o(\log n/\log\log n),
\end{aligned}
$$

where the third inequality applies $\sum_{v\ge\ell+1}\binom{k}{v} \le 2^k$. So we get

$$
\begin{aligned}
\mathrm{Var}(X_k) &= \frac{1}{4k^2}\sum_{T,T':T=T'} \mathrm{Var}(b_T) + o(1) \\
&= \frac{1}{4k^2}\sum_{T,T':T=T'} \Big(\mathbb{E}\left[b_T\right] - \mathbb{E}\left[b_T\right]^2\Big) + o(1) \\
&= \frac{1}{4k^2}[n]_k \times 2k \times \left(\frac{d}{n}\right)^k\left(1-\left(\frac{d}{n}\right)^k\right) + o(1) \\
&= \frac{1+o(1)}{2k}d^k.
\end{aligned}
$$

Remark 7.2. In fact, under H_0, we can show that for any fixed integer $m \ge 1$,

$$
\mathbb{E}\left[[X_k]_m\right] = \mathbb{E}\left[X_k(X_k-1)\cdots(X_k-m+1)\right] = (1+o(1))\left(\frac{d^k}{2k}\right)^m .
$$

It follows from the method of moments (quoted below) that $X_k \xrightarrow{\mathrm{D}} \mathrm{Poi}\left(\frac{d^k}{2k}\right)$.

Lemma 7.2 (Method of moments)**.** *Let X_n be a sequence of random variables. Suppose there exists $\lambda > 0$ such that for every fixed $m \ge 1$, $\mathbb{E}\left[[X_n]_m\right] \to \lambda^m$ as $n\to\infty$. Then $X_n \xrightarrow{\mathrm{D}} \mathrm{Poi}(\lambda)$.*

7.4 Approximately counting cycles in polynomial time

A caveat: The naive way of counting (exhaustive search) k-cycles takes n^k time, which is not polynomial in n if $k\to\infty$. From the previous analysis, we see that we need to count k-cycles with slowly growing k.

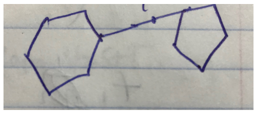
Fix: The trick is to use the sparsity of the random graph and approximately count the number of k-cycles.

Definition 7.1 (ℓ-tangle free)**.** An ℓ-tangle is a connected subgraph of diameter at most 2ℓ that contains at least two cycles.

A graph G is called ℓ-tangle free if no subgraph of G is an ℓ-tangle. In other words, for all $v \in V(G)$, its ℓ-hop neighborhood $N_\ell(v)$ contains at most one cycle.

Lemma 7.3. *If $G \sim G(n, \frac{d}{n})$ and d is a constant, then G is ℓ-tangle free if $\ell = o(\log n)$ (In general $\ell \log d = c\log n$ for small constant c suffices).*

Proof. Suppose G contains an ℓ-tangle. Then G must contain a subgraph H of the following form



with m edges and v vertices, such that $m \le 4\ell$ and $m \ge v+1$. Note that $m = 4\ell$ happens when H consists of two 2ℓ-cycles sharing a common vertex.

There are $O(\ell^3)$ such graphs H up to isomorphism (i.e., unlabelled graphs), as there are $O(\ell)$ choices for the length of each cycle and the connecting path. Also, given a unlabelled graph H, there are at most n^v different vertex labelings, as each vertex in H can have at most n different labels.

Then by union bound, we get that

$$\mathbb{P}\left[G \text{ contains an } \ell\text{-tangle }\right] = \mathbb{P}\left[\exists H \text{ of the above form and } H \text{ is a subgraph of } G\ \right]$$

$$\le O(\ell^3) \times n^v \left(\frac{d}{n}\right)^m \le O(\ell^3) \times n^{m-1} \left(\frac{d}{n}\right)^m \le O(\ell^3) \frac{d^{4\ell}}{n} = o(1),$$

where the last equality holds as long as $\ell \log d \ll \log n$. □

Next we discuss the connection between counting and linear algebra. Let's start with triangles ($k = 3$):

Example 7.1 (Counting triangles)**.** Suppose that A is the adjacency matrix of G. Given any vertex v in G,

$$(A^3)_{vv} = \sum_{a,b} A_{va} A_{ab} A_{bv}$$

is in fact twice the number of triangles incident to v. Therefore, $\mathsf{Tr}(A^3) = 6\times$ the number of triangles in G.

To count k-cycles one can consider computing $\mathsf{Tr}(A^k)$, which can be done in the time of eigenvalue decomposition. But

$$\mathsf{Tr}(A^k) = \text{ number of closed walks of length } k \gg \text{ number of } k\text{-cycles }.$$

The strategy next is use the tangle-free structure and count the number of *non-backtracking* (NB) walks (vertices and edges may repeat).

Definition 7.2 (Non-backtracking walk)**.** We say

- $(v_1, v_2, \ldots, v_k)$ is a NB walk if $v_t \sim v_{t+1}$ and $v_t \ne v_{t-2}$ for all t.
- $(v_1, v_2, \ldots, v_k)$ is a closed NB walk if $v_t \sim v_{t+1}$ and $v_t \ne v_{t-2}$ for all t and $v_1 = v_k$.

For example,

Consequences: Conditioned on G being $2k$-tangle free, any closed NB walk of k steps is either a k-cycle, or an m cycle traversed (in the same direction) for $\frac{k}{m}$ times. Otherwise, we have a $2k$-tangle such as two short cycles sharing a vertex (see Fig. 7.1 above). This reduces the problem of counting k-cycles to counting the number of closed NB walks of length m, for *all* $m = 1, \ldots, k$. Specifically, let N^m_{uv} be the number of NB walks from u to v of length m. Let

$$n_m \triangleq \sum_{v \in V(G)} N^m_{vv}$$

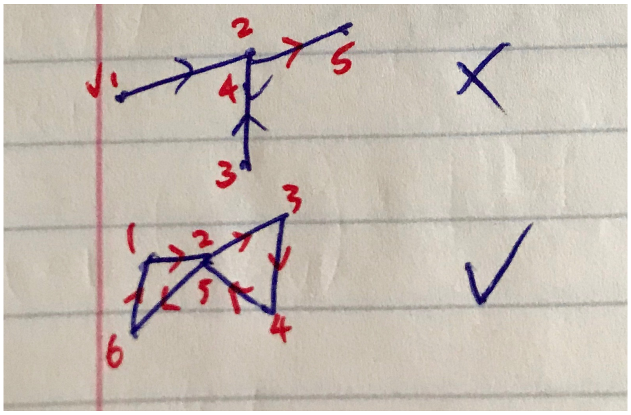


Figure 7.1: Examples of backtracking and non-backtracking walks.

denote the number of closed NB walks of length k. Recall that X_k denote the number of k-cycles. Then

$$\begin{aligned} 2kX_k &= n_k - \sum_{m:m \text{ divides } k} 2mX_m \\ 6X_3 &= n_3. \end{aligned}$$

It remains to compute n_k, which further reduces to computing N_{uv}^m for all pairs u, v. It turns out N_{uv}^m can be counted recursively and given by the following three-term recursion:

$$N_{uv}^{m+1} = \sum_{w \sim v} N_{uw}^m - (d_v - 1)N_{uv}^{m-1}. \tag{7.10}$$

In matrix notation: let $N^{(m)} = (N_{uv}^m)$ and $D = \text{diag}(d_v)$. Then we have[2]

$$\begin{cases} N^{(m+1)} = N^{(m)} \cdot A - N^{(m-1)}(D - \mathbf{I}), \\ N^{(1)} = A, \quad N^{(2)} = A^2 - D \end{cases} \tag{7.11}$$

which means we can compute all N_{uv}^m's using matrix multiplication.

Finally, to justify (7.10), simply notice that the first term on the RHS counts all NB walks of length m from u to a neighbor w of v, which, followed by another step from w to v, constitute a walk of length $m + 1$ from u to v. But, it can be backtracking. So we need to subtract those backtracking walks out, which are precisely given by the second term: fix any NB walk from u to v of length $m - 1$, say, $u, \ldots, v', v$, where $v' \in N(v)$. Append this walk by $w \in N(v)\backslash\{v'\}$ constitutes a NB walk from u to w in m steps.

[2] In the special case of d-regular graphs, (7.11) becomes $N^{(m+1)} = N^{(m)} \cdot A - (d-1)N^{(m-1)}$. This means $N^{(m)}$ is a polynomial of A, in fact, the Chebyshev polynomial, which satisfies the same three-term recurrence. See [ABLS07] for more.

Recall the stochastic block model

$$G \sim SBM(n, p, q)$$
$$\sigma = (\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$$
$$\mathbb{P}[i \sim j] = \begin{cases} p \text{ if } \sigma_i = \sigma_j \\ q \text{ if } \sigma_i \neq \sigma_j. \end{cases}$$

Goal: As described in Section 7.1, an estimator $\widehat{\sigma} = \widehat{\sigma}(G)$ achieves correlated recovery if the overlap is strictly better than random guessing, that is,

$$\mathbb{E}\left[|\langle \widehat{\sigma}, \sigma \rangle|\right] = \Omega(n) \Leftrightarrow \mathbb{E}\left[\min_{s \in \pm 1} \|\widehat{\sigma} - s\sigma\|_1\right] \leq (1 - \Omega(1))\, n.$$

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 $x_1, x_2 \in \{\pm\}$,

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

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 $X \sim \mathrm{Rad}(\frac{1}{2})$ and Y. Then

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

where

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

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

Indeed,

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

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

$$\begin{aligned} I(X;Y) &\leq \frac{1}{2}\left[\chi^2(P_+\|\bar{P}) + \chi^2(P_-\|\bar{P})\right] \\ &= \frac{1}{2}\left[\int \frac{(P_+ - \bar{P})^2}{\bar{P}} + \int \frac{(P_- - \bar{P})^2}{2}\right] \\ &= \int \frac{(P_- - P_+)^2}{2(P_+ + P_-)} \leq \frac{1}{2}\int |P_+ - P_-| = \mathrm{TV}(P_+, P_-). \end{aligned}$$

.

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 $I(\sigma_1, \sigma_2; G) \geq \epsilon$. Then by symmetry $I(\sigma_i, \sigma_j; G) \geq \epsilon$ for all $i \neq j$. Therefore, by Remark 8.1 and (8.1), for all $i \neq j$, $\exists \widehat{T}_{ij} = \widehat{T}_{ij}(G)$, such that

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

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

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

Then the expected number of correctly classified nodes is

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

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

$$\begin{aligned} 2n^2 - \mathbb{E}|\langle \sigma, \widehat{\sigma}\rangle|^2 &= \mathbb{E}\|\sigma\sigma^\top - \widehat{\sigma}\widehat{\sigma}^T\|_F^2 \\ &= 4 \cdot \sum_{i\neq j} \mathbb{P}(\sigma_i\sigma_j \neq \widehat{\sigma}_i\widehat{\sigma}_j) \\ &= 2n^2 - o(n^2), \end{aligned}$$

which means $\mathbb{E}|\langle \widehat{\sigma}, \sigma\rangle|^2 = o(n^2)$. It follows that $\mathbb{E}\left[|\langle \sigma, \widehat{\sigma}\rangle|\right] = o(n)$. □

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:

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

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

To apply this variational representation, take $Q =$ 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 7 for detection. The difference is that here there is no null model and we have to compute the exact asymptotic instead of just an upper bound. 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

$$\int \frac{P_z P_{\widetilde{z}}}{Q} = C(\tau) + o(1), \forall z, \widetilde{z} \in \{\pm 1\},$$

where $C(\tau)$ is some constant independent of z and $\widetilde{z}$. Consider the case of iid labels, where $\sigma_i \overset{\text{i.i.d.}}{\sim} Rad(1/2)$. By the same argument in Section 7.2, we have

$$\begin{aligned}
\int \frac{P_z P_{\widetilde{z}}}{Q} &= \int \frac{\mathbb{E}_\sigma \left[P_\sigma \mid \sigma_1 \sigma_2 = z \right] \mathbb{E}_{\widetilde{\sigma}} \left[P_{\widetilde{\sigma}} \mid \widetilde{\sigma}_1 \widetilde{\sigma}_2 = \widetilde{z} \right]}{Q} \\
&\overset{Fubini}{=} \mathbb{E}_{\sigma \perp \widetilde{\sigma}} \left[\int \frac{P_\sigma P_{\widetilde{\sigma}}}{Q} \mid \sigma_1 \sigma_2 = z, \widetilde{\sigma}_1 \widetilde{\sigma}_2 = \widetilde{z} \right] \\
&= \mathbb{E}_{\sigma \perp \widetilde{\sigma}} \left[\prod_{i<j} (1 + \rho \sigma_i \sigma_j \widetilde{\sigma}_i \widetilde{\sigma}_j) \mid \sigma_1 \sigma_2 = z, \widetilde{\sigma}_1 \widetilde{\sigma}_2 = \widetilde{z} \right] \\
&= \mathbb{E}_{\sigma \perp \widetilde{\sigma}} \left[e^{\sum_{i<j} \log(1 + \rho \sigma_i \sigma_j \widetilde{\sigma}_i \widetilde{\sigma}_j)} \mid \sigma_1 \sigma_2 = z, \widetilde{\sigma}_1 \widetilde{\sigma}_2 = \widetilde{z} \right] \\
&= \mathbb{E}_{\sigma \perp \widetilde{\sigma}} \left[e^{\sum_{i<j} \left(\rho \sigma_i \sigma_j \widetilde{\sigma}_i \widetilde{\sigma}_j - \frac{\rho^2}{2} (\sigma_i \sigma_j \widetilde{\sigma}_i \widetilde{\sigma}_j)^2 + O(\rho^3) \right)} \mid \sigma_1 \sigma_2 = z, \widetilde{\sigma}_1 \widetilde{\sigma}_2 = \widetilde{z} \right] \\
&= e^{-\frac{\rho n}{2} - \frac{\rho^2}{2} \binom{n}{2} + O(\rho^3 n^2)} \mathbb{E}_{\sigma \perp \widetilde{\sigma}} \left[e^{\frac{1}{2} \rho \langle \sigma, \widetilde{\sigma} \rangle^2} \mid \sigma_1 \sigma_2 = z, \widetilde{\sigma}_1 \widetilde{\sigma}_2 = \widetilde{z} \right] \\
&\overset{(b)}{=} e^{-\frac{\tau}{2} - \frac{\tau^2}{4} + o(1)} \mathbb{E}_{\sigma \perp \widetilde{\sigma}} \left[\exp \left(\frac{\tau + o(1)}{2} \underbrace{\boxed{\frac{1}{n} \langle \sigma, \widetilde{\sigma} \rangle^2}}_{N(0,1)^2} \right) \mid \sigma_1 \sigma_2 = z, \widetilde{\sigma}_1 \widetilde{\sigma}_2 = \widetilde{z} \right] \\
&\overset{(c)}{=} (1 + o(1)) \, e^{-\frac{\tau}{2} - \frac{\tau^2}{4}} \frac{1}{\sqrt{1 - \tau}} \triangleq C(\tau) + o(1),
\end{aligned}$$

where (b) follows from $\rho = \frac{\tau}{n} + \frac{(a-b)^2}{4n^2} + O(1/n^3)$; (c) follows from CLT almost the same as before: $\frac{1}{\sqrt{n}}\langle \sigma, \widetilde{\sigma}\rangle = \frac{1}{\sqrt{n}}\sum_{j=3}^{n} \sigma_j \widetilde{\sigma}_j + \frac{1}{\sqrt{n}}(\sigma_1\widetilde{\sigma}_1 + \sigma_2\widetilde{\sigma}_2)$, where the first term is asymptotically $N(0,1)$ and independent of $(\sigma_1, \sigma_2, \widetilde{\sigma}_1, \widetilde{\sigma}_2)$, and the second term is negligible.

More generally,

- For exact bisection the same statement holds true, except that one should be more careful with the conditioning.
- For the spiked Wigner model (7.2), 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 splitting reduction) that impossibility of detection $\implies$ impossibility of correlated recovery (Homework).

8.2 Correlated recovery via spectral methods: first attempt

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 (7.2) 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 $\widehat{u} = u_1(W)$ of the matrix W corresponding to the largest eigenvalue λ_1, and report $\text{sign}(u_1)$ as the estimate $\widehat{\sigma}$. Let $u = \frac{1}{\sqrt{n}}\sigma$. This method succeeds in correlated recovery if and only (why?) if $|\langle u, \widehat{u}\rangle|$ is bounded away from 0.

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

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

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

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

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

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

except for the zero diagonal, and $\mathrm{Var}(A_{ij} - \mathbb{E}A_{ij}) = \frac{a+o(1)}{n}$ or $\frac{b+o(1)}{n}$. It follows that $\sum_j \mathrm{Var}(A_{ij} - \mathbb{E}A_{ij}) = \frac{d+o(1)}{n}$, 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]

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

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 $\|A\| \geq \frac{\|Ae_i\|}{\|e_i\|} = \sqrt{d_{\max}}$. As the matrix has all non-negative entries, by Perron Frobenius 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 $d_i = d_{\max}$. Then

$$\begin{aligned}(A^{2k})_{ii} &= \sum_{i_2,\ldots,i_{2k}} A_{ii_2}A_{i_2i_3}\cdots A_{i_{2k}i} \\ &= \text{ number of closed walks from } i \text{ to } i \text{ of length } 2k \\ &\geq d_{\max}^k,\end{aligned} \tag{8.4}$$

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

$$\|A\|^{2k} \geq \|A^k e_i\|_2^2 = e_i^\top A^{2k} e_i \geq d_{\max}^k.$$

Thus $\lambda_1(A) = \|A\| \geq \sqrt{d_{\max}}$, 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,

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

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 $s^2 \geq d$. 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 explain in Section 9.3. 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.

3. Turn to SDP. This can resolve the high-degree issue by imposing the diagonal constraint $X_{ii} = 1$, but is difficult to achieve the sharp correlated recovery threshold. We will study SDP in the sparse graphs later and show SDP achieves the almost exact recovery threshold.

In this lecture, we first derive the belief propagation (BP) for community detection under the stochastic block model. Then we study the noise sensitivity of BP around the trivial fixed point. This gives rise to the conjectured threshold for the correlated recovery and motivates a new type of spectral method based on the so-called non-backtracking matrix.

9.1 BP algorithm

We beging by showing that the SBM graph with constant average degree is locally tree like. Then we derive the exact belief propagation algorithm for an infinite tree. Finally, we apply the same algorithm on the original SBM graph.

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

$$M = \begin{bmatrix} \frac{a}{2} & \frac{b}{2} \\ \frac{b}{2} & \frac{a}{2} \end{bmatrix}. \tag{9.1}$$

More formally, we can prove the following coupling lemma. For each vertex $i \in [n]$ and t, let G_i^t denote the subgraph of G induced by the vertices whose distance from i is at most t. Let T_i^t denote the two-type Galton Watson tree T rooted at vertex i of depth t, and τ denote the labels of vertices in T, where $\tau_i \sim \mathrm{Unif}(\{\pm\})$.

Lemma 9.1 (Locally-tree like property)**.** *Consider SBM$(n, \frac{a}{n}, \frac{b}{n})$ graph G with average degree $d = \frac{a+b}{2}$ and the underlying community σ uniformly generated at random. Assume $t \log d = o(\log n)$. For any fixed vertex i, there exists a coupling between (G, σ) and (T, τ) such that*

$$\mathbb{P}\left[\left(G_i^t, \sigma_{G_i^t}\right) = \left(T_i^t, \tau_{T_i^t}\right)\right] \geq 1 - n^{-1+o(1)}.$$

Proof. The proof basically follows by applying the Poisson approximation of the Binomial distribution and showing that the local neighborhood G_i^t does not contain a cycle with high probability. See e.g. [HWX15, Appendix C] for a formal proof. □

9.1.1 BP on a two-type Galton Watson tree

Using the recursive tree structure, it turns out that the posterior probability $p(\tau_i | T_i^t)$ can be computed recursively. This leads to the so-called belief propagation algorithm. Specifically, let ∂i denote the set of childen of vertex i and $\pi(i)$ denote the parent of i.

Let

$$\Lambda^t_{i\to\pi(i)} \triangleq \frac{1}{2}\log\frac{\mathbb{P}\left[\tau_i=+|T_i^t\right]}{\mathbb{P}\left[\tau_i=-|T_i^t\right]}.$$

The following lemma gives a recursive formula to compute $u^t_{i\to\pi(i)}$.

Lemma 9.2. *Let $\beta=\frac{1}{2}\log(a/b)$. For $t\ge 1$,*

$$\Lambda^{t+1}_{i\to\pi(i)}=\sum_{\ell\in\partial i} f\left(\Lambda^{(t)}_{\ell\to i}\right), \quad f(x)\triangleq \tanh^{-1}[\tanh(\beta)\tanh(x)]. \tag{9.2}$$

Proof. Let N_i denote the number of childen of vertex i.

$$\begin{aligned}
\Lambda^{t+1}_{i\to\pi(i)} &\overset{(a)}{=} \frac{1}{2}\log\frac{\mathbb{P}\left[T_i^{t+1}|\tau_i=+\right]}{\mathbb{P}\left[T_i^{t+1}|\tau_i=-\right]}\\
&\overset{(b)}{=} \frac{1}{2}\log\frac{\mathbb{P}\left[N_i|\tau_i=+\right]}{\mathbb{P}\left[N_i|\tau_i=-\right]}+\frac{1}{2}\sum_{\ell\in\partial i}\log\frac{\mathbb{P}\left[T_\ell^t|\tau_i=+\right]}{\mathbb{P}\left[T_\ell^t|\tau_i=-\right]}\\
&\overset{(c)}{=} \frac{1}{2}\sum_{\ell\in\partial i}\log\frac{\sum_{x\in\pm}\mathbb{P}\left[T_\ell^t|\tau_\ell=x\right]\mathbb{P}\left[\tau_\ell=x|\tau_i=+\right]}{\sum_{x\in\pm}\mathbb{P}\left[T_\ell^t|\tau_\ell=x\right]\mathbb{P}\left[\tau_\ell=x|\tau_i=-\right]}\\
&\overset{(d)}{=} \frac{1}{2}\sum_{\ell\in\partial i}\log\frac{\mathbb{P}\left[T_\ell^t|\tau_\ell=+\right]\frac{a}{a+b}+\mathbb{P}\left[T_\ell^t|\tau_\ell=-\right]\frac{b}{a+b}}{\mathbb{P}\left[T_\ell^t|\tau_\ell=+\right]\frac{b}{a+b}+\mathbb{P}\left[T_\ell^t|\tau_\ell=-\right]\frac{a}{a+b}}\\
&\overset{(e)}{=} \frac{1}{2}\sum_{\ell\in\partial i}\log\frac{\exp\left(2\beta+2\Lambda^t_{\ell\to i}\right)+1}{\exp\left(2\Lambda^t_{\ell\to i}\right)+\exp\left(2\beta\right)}\\
&\overset{(f)}{=} \sum_{\ell\in\partial i} f\left(\Lambda^{(t)}_{\ell\to i}\right),
\end{aligned}$$

where (a) holds due to $\tau_i\sim\text{Unif}(\{\pm\})$; ($b$) follows because N_i and $\{T_\ell^t:\ell\in\partial i\}$ are independent conditional on τ_i; (c) holds as $N_i\sim\text{Poi}(d)$ is independent from τ_i, and T_ℓ^t is independent of τ_i conditonal on τ_ℓ; (d) follows from the sample splitting property of Poisson so that $\tau_\ell=\tau_i$ with probability $\frac{a}{a+b}$ and $\tau_\ell=-\tau_i$ with probability $\frac{b}{a+b}$; (e) holds by plugging in the definition of β and $\Lambda^t_{\ell\to i}$; (f) follows from the following generic identity:

$$\tanh(y)=\tanh(\beta)\tanh(x)\Longleftrightarrow y=\frac{1}{2}\log\frac{\exp(2x)\exp(2\beta)+1}{\exp(2x)+\exp(2\beta)}.$$

□

9.1.2 BP for SBM

To detect which community a given vertex i belongs to, a natural approach is to compare the posterior ratio $\log\frac{\mathbb{P}[\sigma_i=+|G]}{\mathbb{P}[\sigma_i=-|G]}$ to a certain threshold.[1]. As we have seen, when the average degree is d, the neighborhood of vertex i is tree-like with high probablility as long as the radius of the neighborhood satisfies $t\log d=o(\log n)$; moreover, on the tree, the posterior ratio can be exactly computed in a finite recursion via belief propagation. These two observations together suggest

[1] Careful readers may notice that due to the symmetry between + and −, this posterior ratio is always equal to 1. Let us not worry about this, as this symmetry can be broken easily, for example by fixing the community label of an arbitrary vertex to be +.

the following belief propagation algorithm for approximately computing the posterior ratio for the community recovery problem on SBM. Define the message transimitted from vertex i to one of its neighbor j at $(t+1)$-th iteration as

$$u_{i\to j}^{t+1} = \sum_{\ell\in\partial i\setminus\{j\}} f\left(u_{\ell\to i}^{(t)}\right), \tag{9.3}$$

where ∂i denotes the set of neighbors of i in G. Note that motivated by the tree structure, when computing the outgoing message from i to j, crucially we exclude the incoming message from j to i. This prevents the unuseful echoes of messages that bounces back and forth between i and j and allows for gathering useful information over multi-hop neighbors. Then we can approximate the posterior ratio $\log\frac{\mathbb{P}[\sigma_i=+|G]}{\mathbb{P}[\sigma_i=-|G]}$ by the belief of vertex i at $(t+1)$-th iterations, u_i^t, which is determined by combining incoming messages from its neighbors as follows:

$$u_i^{t+1} = \sum_{\ell\in\partial i} f\left(u_{\ell\to i}^{(t)}\right).$$

Remark 9.1 (Computational efficiency of BP)**.** Note that in total there are $2|E|$ messages, which are updated according to (9.3) for every iteration. Thus the total computational complexity per iteration is linear in the number of edges, rendering BP particularly attracting in sparse graphs where $|E| \ll |V|^2$.

Remark 9.2 (BP in general contexts)**.** In the above, since the community label is binary, it is convenient to derive the BP in terms of the posterior ratio. With multiple community labels, we can derive the BP directly in terms of the posterior probability $\mathbb{P}\left[\tau_i|T_i^t\right]$ in a similar fashion. In fact, more generally, BP (a.k.a. sum-product algorithm) is an iterative algorithm for approximately computing the marginals for graphical models (cf. [MM09, Chapter 14] for a detailed exposition).

9.2 Trivial fixed point and noise sensitivity of BP

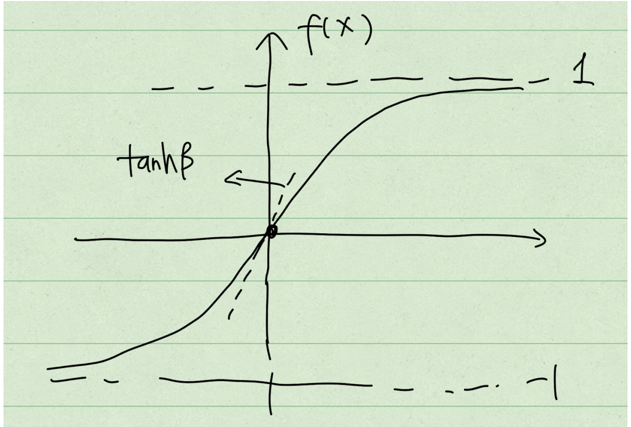
In this section, we study the BP update rule (9.3). To begin with, we collect some simple yet useful properties of f as below.

Properties of f:

- $f(0)=0$, $f(-\infty)=-1$, and $f(+\infty)=+1$.
- $f'(x)=\frac{(1-\tanh^2(x))\tanh(\beta)}{1-\tanh^2(\beta)\tanh^2(x)}$, so $0\le f'(x)\le\tanh(\beta)$. In particular, $f'(0)=\tanh(\beta)$.
- $f''(x)\le 0$, so $f(x)$ is concave

It immediately follows from $f(0)=0$ that $u_{i\to j}^{(t)}\equiv 0$ is always a fixed point, in which every node is equally likely to be in the $+1$ or -1 community. This fixed point is trivial, in the sense that if BP starts from this fixed point, then it will get stuck and never do better than random guessing.

Question: If we initially perturb the BP messages away from the trivial fixed point, should BP fly away from it - hopefully toward the truth- or fall back in?



Answer: It turns out that the trivial fixed point is unstable if and only if $\tau = \frac{(a-b)^2}{2(a+b)} > 1$. To see this, let us assume the BP messages are independently perturbed away from the trivial fixed point by some small random noise, i.e., $u^{(0)}_{\ell\to i} = \epsilon_\ell$, where ϵ_ℓ's are i.i.d. with mean 0 and a sufficiently small variance κ. Since the messages $u^{(t)}_{\ell\to i}$ are expected to be small, we can apply Taylor expansion and approximate

$$f\left(u^{(t)}_{\ell\to i}\right) \approx f(0) + f'(0)u^{(t)}_{\ell\to i} = \tanh(\beta)u^{(t)}_{\ell\to i},$$

where the last equality holds because $f(0) = 0$ and $f'(0) = \tanh(\beta)$. This give rises to a linearized BP:

$$u^{(t+1)}_{i\to j} \approx \tanh(\beta) \sum_{\ell\in\partial i\backslash j} u^{(t)}_{\ell\to i}. \tag{9.4}$$

Let us investigate the behavior of this LBP on the two-type Galton-Watson tree T^t_i rooted at i of depth t. We get that

$$u^{(t)}_{i\to j} \approx \sum_{\text{leaves } \ell \text{ in } T^{(t)}_i} \tanh^t(\beta)\epsilon_\ell.$$

Since the perturbations have mean 0, the mean of $u^{(t)}_{i\to j}$ is also 0. For the variance, however,

$$\mathrm{Var}\left(u^{(t)}_{i\to j}\right) \approx \sum_{\text{leaves } \ell \text{ in } T^{(t)}_i} \tanh^{2t}(\beta)\kappa \approx \tanh^{2t}(\beta)d^t\kappa,$$

where the first approximation holds as the perturbations are independent across different leaves, and the last approximation holds as the number of leaves in $T^{(t)}_i$ is roughly d^t according to the standard branching process. Note that the variance diverges to ∞ as t tends to ∞, when

$$\tanh^2(\beta)d > 1 \iff \frac{(a-b)^2}{2(a+b)} > 1, \tag{9.5}$$

suggesting that the BP messages fly away from the trivial fixed point. In fact, [DKMZ11] further conjectured that under condition (9.5), the BP messages $u^{(t)}_{i\to j}$ are positively correlated with the true community label σ_i for sufficiently large iterations t.

Conjecture 9.1. *Consider the BP algorithm (9.3) with random initialization and let* $\widehat{\sigma}_i = \arg\max_{x\in\pm} u^t_i(x)$. *Then for sufficiently large* t*, with high probability*

$$|\langle\widehat{\sigma}_i, \sigma_i\rangle| = \Omega(n).$$

While convincing numerical evidence is provided in [DKMZ11], a rigorous proof of the conjecture remains open.

9.3 Spectrum of non-backtracking matrices

While a rigorous analysis of the BP is challenging due to its non-linear dynamic, the linearized BP given in (9.4) turns out to be easier to analyze, by relating it to the spectrum of a so-called non-backtracking matrix.

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 $e = (e_1, e_2), f = (f_1, f_2) \in \vec{E}$,

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

Then we can rewrite (9.4) in a compact matrix notation as:

$$u^{(t+1)} \approx \tanh(\beta) B^\top u^{(t)}.$$

This can be viewed as a power method applied to the NB matrix $B^\top$. Thus we expect $u^{(t)}$ converges to the leading eigenvectors of $B^\top$.

Remark 9.3. The stability of the trivial fixed point can be also understood through the eigenvalues of B. As we will see, the largest eigenvalue $\lambda_1(B) = d + o(1)$. The corresponding eigenvector is asymptotically aligned with the all-one vector. Since the perturbations have mean 0 so that $\sum_\ell u^{(t)}_{\ell\to i} \approx 0$, the relevant eigenvalue is the second largest eigenvalue of B. We will show later that $\lambda_2(B) = \max\left\{\sqrt{d}, \frac{a-b}{2}\right\} + o(1)$, which is $\frac{a-b}{2} + o(1)$ when $\frac{a-b}{2} > \sqrt{d}$. Thus the trivial fixed point of the BP is unstable if and only if

$$\frac{a-b}{2} > \sqrt{d} \quad \text{and} \quad \tanh(\beta)\frac{a-b}{2} > 1.$$

It turns out (somewhat magically) that the two above conditions coincide and reduce to $\tau = \frac{(a-b)^2}{2(a+b)} > 1$. Moreover, as we will show later, when $\tau > 1$, the eigenvector corresponding to the second-largest eigenvalue of B is correlated with the true community label σ, thereby achieving the correlated recovery.

To warm up, let us first study some basic properties of NB matrices.

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

1. Row sum: $\forall e = (u, v), \sum_{e'\in\vec{E}} B_{ee'} = d_v - 1$.

2. 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
$$(B^\top)_{ef} = B_{e^{-1}f^{-1}}. \tag{9.6}$$
In matrix notation, let $P = (\mathbb{1}\left\{e = f^{-1}\right\})$ 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 $B^k = PB^kP$.

3. B is a $2m \times 2m$ matrix, and can be partitioned into four $m \times m$ blocks (The first m rows and columns correspond to edges in one direction, and the next m rows and columns correspond to edges in the reversed direction):

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

To see this, by (9.6),

$$B^\top = \begin{bmatrix} B_{11}^\top & B_{21}^\top \\ B_{12}^\top & B_{22}^\top \end{bmatrix} = \begin{bmatrix} B_{22} & B_{21} \\ B_{12} & B_{11} \end{bmatrix}.$$

It follows that $B_{11} = B_{22}^\top$, and B_{12}, B_{21} are symmetric.

Note: Since B is not symmetric, i.e., $B_{ef} \neq B_{fe}$, the eigenvalues of B may be complex-valued.

4. Singular values of B are $\{d_v - 1 : v \in V\} \cup \{1\}$ and thus not informative.

There are two different approaches to see this. One way is to find eigenvalues of $BB^\top$. Here we follow a second approach using the fact that BP is symmetric. In particular, denote the eigenvalue decomposition of BP as $BP = \sum_{j=1}^{2m} s_j x_j x_j^\top$. It follows that $B = \sum_{j=1}^{2m} s_j x_j (P x_j)^\top = \sum_{j=1}^{2m} |s_j| x_j y_j^\top$, where $y_j = \operatorname{sgn}(s_j) P x_j$. This gives rise to the singular value decomposition of B. Hence, it reduces to determining the spectrum of BP. To this end, note that

$$(BP\xi)_e = \sum_f (BP)_{ef} \xi_f = \sum_f \sum_{e'} B_{ee'} P_{e'f} \xi_f = \sum_f B_{ef^{-1}} \xi_f,$$

where the second equality holds by the definition of involution P. Now, we can see that if we let $\xi_e = 1$ for all incoming edges e to v and $\xi_e = 0$ otherwise, then $(BP\xi)_e = (d_v - 1)\xi_e$. This shows that ξ is an eigenvector of BP with eigenvalue $d_v - 1$. By varying the node v, we have n of them in total. Further, if we let $\xi_e = 1$ and $\xi_f = -1$ for some f such that $B_{ef^{-1}} = 1$ and $\xi_{e'} = 0$ otherwise, then $(BP\xi)_e = -\xi_e$. Thus ξ is an eigenvector of BP with eigenvalue -1. Since for every node v, we can fix an incoming edge to be e and choose any other $d_v - 1$ incoming edges as f, we have in total $\sum_v (d_v - 1) = 2m - n$ such eigenvectors.

More generally, denote the eigenvalue decomposition of $B^k P$ as $B^k P = \sum_{j=1}^{2m} s_{j,k} x_{j,k} x_{j,k}^\top$. We have $B^k = \sum_{j=1}^{2m} |s_{j,k}| x_{j,k} y_{j,k}^\top$ with $y_{j,k} = \operatorname{sgn}(s_{j,k}) P x_{j,k}$, which is the SVD of B^k. It turns out that the SVD of B^k for large k reveals the eigen-structure of B. In particular, we have the following classical Perron-Frobenius theorem.

Theorem 9.1 (Perron-Frobenius Theorem)**.** *Suppose B is irreducible, i.e., $\forall (i,j)$, $\exists k$ such that $(B^k)_{ij} > 0$. Then the following holds.*

- *$\lambda_1(B)$ is real and positive, and $|\lambda_i(B)| < \lambda_1(B)$ for all $i \geq 2$.*
- *Let ξ denote the right eigenvector of B with eigenvalue $\lambda_1(B)$, i.e., $B\xi = \lambda_1(B)\xi$. We have $\xi_i > 0$ for all i.*
- *Let $B^k = \sum_j s_{j,k} x_{j,k} y_{j,k}^\top$ denote the SVD of B^k. Then*

$$\lambda_1(B) = \lim_{k \to \infty} (s_{1,k})^{1/k} \quad \text{and} \quad \lim_{k \to \infty} \|\xi - x_{1,k}\|_2 = 0.$$

The above theorem suggests that we can study the eigenvalue of B through the SVD of B^k or equivalently the EVD of $B^k(B^\top)^k$. This is exactly how we will analyze the spectrum of NB matrix B.

5. Ihara-Bass identity [Ter10, p. 89]: Let A, D, B denote the adjacency matrix, diagonal degree, and non-backtracking matrices. Then for any complex number $z \in \mathbb{C}$,
$$\det(\mathbf{I} - zB) = (1-z^2)^{m-n}\det(\mathbf{I} - zA + z^2(D - \mathbf{I})), \tag{9.7}$$
where $D = \operatorname{diag}(d_v)$. This means B has $2(m-n)$ useless eigenvalues that are equal to ± 1, and the rest of the $2n$ eigenvalues are useful, and are given by the $2n$ roots of the polynomial $\det(\lambda^2 I - \lambda A + D - \mathbf{I})$. Note that $\lambda^2\mathbf{I} - \lambda A + D - \mathbf{I}$ is known as the "Bethe Hessian" matrix [SKZ14] and gives a more efficient way to compute the eigenvalues of B.

   For a d-regular graph with $D = d\mathbf{I}$. Then λ is an eigenvector of B if and only if $\lambda = \pm 1$ or $\frac{\lambda^2+d-1}{\lambda}$ is an eigenvalue of A.

   Proof of Ihara-Bass identity:

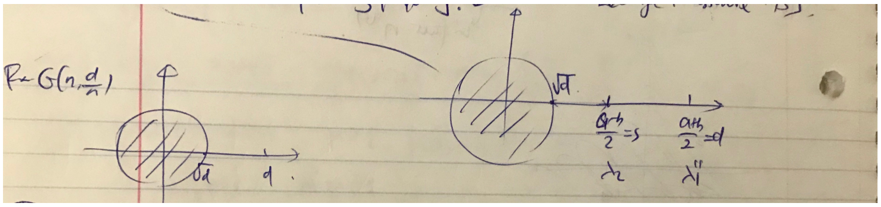
   Proof. Let $S \in \mathbb{R}^{n\times 2m}$ and $T \in \mathbb{R}^{2m\times n}$ be defined as
$$S_{u,v\to w} = \begin{cases} 1 & \text{if } u = v \\ 0 & \text{o.w.} \end{cases} \qquad T_{u\to v,w} = \begin{cases} 1 & \text{if } v = w \\ 0 & \text{o.w.} \end{cases}$$
   Let
$$\xi_u^{\text{out}} = (S\xi)_u = \sum_{v\to w} S_{u,v\to w}\xi_{v\to w} = \sum_{w:w\sim u} \xi_{u\to w}$$
   and
$$\xi_w^{\text{in}} = (\xi^\top T)_w = \sum_{u\to v} \xi_{u\to v}T_{u\to v,w} = \sum_{u:u\sim w} \xi_{u\to w}.$$
   Recall the involution operator $P_{ef} = \mathbb{1}\left\{e = f^{-1}\right\}$. Then $(P\xi)_{u\to v} = \xi_{v\to u}$, $P^2 = \mathbf{I}$ and $P = P^\top$. Note that

   - $(ST)_{uv} = \mathbb{1}\{u \sim v\} = A_{uv}$;
   - $(TS)_{u\to v,s\to t} = \mathbb{1}\{v \to s\}$, so $TS - P = B$;
   - $SPT = D$.

   Note that P (by putting rows and columns corresponding to e and e^{-1} next to each other) consists of blocks $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ along the diagonal. Thus P has eigenvalues ± 1. For any complex $z \notin \{\pm 1\}$, we have
$$\det(\mathbf{I} - zB) = \det(\mathbf{I} + zP - zTS) = \det(\mathbf{I} + zP)\det\left(\mathbf{I} - (\mathbf{I} + zP)^{-1}zTS\right)$$
   Note that $\det(\mathbf{I} + zP) = (1-z^2)^m$ and $(\mathbf{I} + zP)^{-1} = \frac{1}{1-z^2}\mathbf{I} - \frac{z}{1-z^2}P$. It follows that
$$\begin{aligned}\det(\mathbf{I} - zB) &= (1-z^2)^m \det\left(\mathbf{I} - \left(\frac{1}{1-z^2}\mathbf{I} - \frac{z}{1-z^2}P\right)zTS\right) \\ &= (1-z^2)^m \det\left(\mathbf{I} - zS\left(\frac{1}{1-z^2}\mathbf{I} - \frac{z}{1-z^2}P\right)T\right) \\ &= (1-z^2)^m \det\left(\mathbf{I} - \frac{z}{1-z^2}A + \frac{z^2}{1-z^2}D\right) \\ &= (1-z^2)^{m-n}\det\left((1-z^2)\mathbf{I} - zA + z^2D\right),\end{aligned}$$

where the second equality holds by Sylvester's determinant identity: $\det(\mathbf{I}+AB) = \det(\mathbf{I}+BA)$. By continuity, this must also hold for $z \to \pm 1$. □

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 9.2 ([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*

$$\widehat{\sigma}_v = \operatorname{sign}\left(\sum_{e:e_2=v} (u_2)_e \right).$$

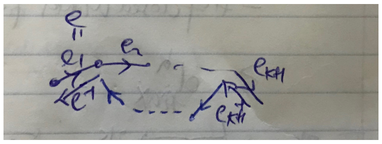
Then $\widehat{\sigma}$ achieves correlated recovery if $s^2 > d$.

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. We claim that for each oriented edge e,

$$(B^k(B^\top)^k)_{ee} \tag{9.8}$$
$$= \#\text{ NB walks starting with } e \text{ in } k \text{ steps then reversing the last step and returning to } e \text{ in } k \text{ steps}$$

such as



Indeed, using the symmetry property,

$$\begin{aligned}
(B^k(B^\top)^k)_{ee} &= \sum_{e_2 \ldots e_{2k}} B_{e_1e_2}B_{e_2e_3}\ldots B_{e_ke_{k+1}}B^\top_{e_{k+1}e_{k+2}}\ldots B^\top_{e_{2k}e_{2k+1}} \qquad e_1 = e, e_{2k+1} = e \\
&\overset{(9.6)}{=} \sum B_{ee_2}B_{e_2e_3}\ldots B_{e_ke_{k+1}}B^\top_{e^{-1}_{k+1}e^{-1}_{k+2}}\ldots B^\top_{e^{-1}_{2k}e^{-1}}
\end{aligned}$$

To simplify the counting in (9.8), 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 (9.8), the path must reverse itself (otherwise there will be a cycle). Thus, on the event that locally tree-like structure holds, we have

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

even if the degree of u is as large as $\frac{\log n}{\log\log n}$!

The last step follows from basic results in branching process, which states that the total number of kth-generation children grows exponentially as d^k. More formally, we have the following theorem.

Theorem 9.3. *Let $Z_0 = 1$ and $Z_k = \sum_{i=1}^{Z_{k-1}} Y_i$, where $Y_i \overset{i.i.d.}{\sim} \text{Poi}(d)$ with $d > 1$. Let $\mathcal{F}_k$ denote the σ-field generated by $\{Z_1, \ldots, Z_k\}$. Then there exists a random variable X such that*

$$\frac{Z_k}{d^k} \xrightarrow{L^2} X$$

and $\mathbb{E}[X|\mathcal{F}_k] = Z_k/d^k$.

Proof. Let $X_k = Z_k/d^k$. Then we have $\mathbb{E}[X_k|\mathcal{F}_k] = X_{k-1}$ and hence X_k is a martingale w.r.t. the filtration $\mathcal{F}_k$. We can further show $\text{Var}(X_k) = \text{Var}(X_{k-1}) + d^{-k}$ and hence $\text{Var}(X_k) = \sum_{i=1}^k d^{-i} \le \frac{1}{1-1/d}$. Thus the theorem follows from the martingale convergence theorem. □

Finally,

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

which suggests that the *bulk* of the eigenvalues belong to the disk of radius $\sqrt{d}$, i.e., all but a $o(1)$ fraction of eigenvalues of B should be within the disk $\{z \in \mathbb{C} : |z| \le \sqrt{d} + \epsilon\}$ for some arbitrary constant $\epsilon > 0$ and all large n. In fact, [BLM18] shows that all but (1 for Erdős-Rényi and 2 for SBM) eigenvalues are within the disk.

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 v, we estimate its label $\sigma(v)$ by $\widehat{\sigma}_v = \text{sign}(\sum_{e:e_2=v} \xi_e)$.

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.[2] Fix an edge $e = (e_1, e_2)$ with $e_2 = v$.

$$\begin{aligned}(B^k\xi^*)_e &= \sum_f (B^k)_{ef}\xi^*_f \\ &= \underbrace{\sum_{f:\sigma(f_2)=+} (B^k)_{ef}}_{\#\text{ of } k\text{th-gen children of type } + \triangleq Z_k^+} - \underbrace{\sum_{f:\sigma(f_2)=-} (B^k)_{ef}}_{\#\text{ of } k\text{th-gen children of type } - \triangleq Z_k^-}\end{aligned}$$

where 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 (9.1), whose eigenvalues are $\lambda_1 = d$ and $\lambda_2 = s$. More formally,

[2] The 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.

Theorem 9.4. *If $\lambda_2^2 > \lambda_1$, then there exists a random variable X such that*

$$\frac{Z_k^+ - Z_k^-}{\lambda_2^k} \xrightarrow{L^2} X,$$

$\mathbb{E}\left[X|\mathcal{F}_k\right] = (Z_k^+ - Z_k^-)/\lambda_2^k$*, and X is correlated with the label of the root.*

Proof. The proof is similar to that of Theorem 9.3. Let $X_k = (Z_k^+ - Z_k^-)/\lambda_2^k$. Then we can show $\mathbb{E}\left[X_k|\mathcal{F}_{k-1}\right] = X_{k-1}$ and hence X_{k-1} is a martingale w.r.t. the filtration $\mathcal{F}_k$. We can further show that $\mathrm{Var}(X_k) = \mathrm{Var}(X_{k-1}) + (\lambda_1/\lambda_2^2)^k$. Thus under the assumption that $\lambda_2^2 > \lambda_1$, we have $\mathrm{Var}(X_k) = O(1)$. The theorem then follows from the martingale convergence theorem. □

Theorem 9.4 implies that $(B^k\xi^*)_e$ has non-trivial correlation with σ_v, and correlated recovery can be achieved by majority vote.

Now, let's intuitively verify that $(B^k\xi^*)/\lambda_2^k$ is approximately an eigenvector of B with eigenvalue $\lambda_2 = s = (a-b)/2$. Indeed,

$$B\frac{B^k\xi^*}{\lambda_2^k} = \lambda_2\frac{B^{k+1}\xi^*}{\lambda_2^{k+1}} \approx \lambda_2\frac{B^k\xi^*}{\lambda_2^k},$$

where the last approximation holds because by Theorem 9.4, $(B^k\xi^*)_e/\lambda_2^k$ converges as $k\to\infty$ so that $(B^k\xi^*)_e/\lambda_2^k \approx (B^{k+1}\xi^*)_e/\lambda_2^{k+1}$. In a similar vein, we can also argue that $(B^k\mathbf{1})_e/\lambda_1^k$ is approximately the eigenvector of B with eigenvalue $\lambda_1 = d = \frac{a+b}{2}$.

Nevertheless, the above plan is too simplistic as B is asymmetric so the straightforward power method does not work. In reality, to apply the power method properly, one needs to study the EVD of B by considering the SVD of B^k or equivalently the EVD of $B^k(B^k)^\top$. 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.[3] For details, see [BLM18, Sec 8].

9.4 Reconstruction on a two-type Galton-Watson tree process

While a rigorous proof of Conjecture 9.1 is still lacking [DKMZ11], we can perform a rigorous analysis of BP in a closely related, but simpler reconstruction problem on the two-type Galton-Watson tree.

Question: Given the labels at the depth t of the GW tree T, can we estimate the label of the root better than random guessing, as $t\to\infty$? More formally, let $T_i^{(t)}$ denote the tree rooted at i up to depth t and $L_i^{(t)}$ denote the set of leaves at depth t. Does $\lim_{t\to\infty}\mathbb{P}\left[\sigma_i = +1|T_i^{(t)}, \sigma_{L_i^{(t)}}\right] = 1/2$? If yes, then we cannot do better than random guessing.

Answer: It turns out that the reconstruction is possible if and only if $\lambda_2^2 > \lambda_1$, i.e., $\tau = \frac{(a-b)^2}{2(a+b)} > 1$. In particular, we can achieve the threshold using the simple majority voting, that is, $\widehat{\sigma}_i = \mathrm{sgn}\left(\sum_{\ell\in L_i^{(t)}}\sigma_\ell\right)$. This follows from the Kesten-Stigum theorem stated in Theorem 9.4. We can do even better in terms of the error probability. It is not hard to show that the estimator that minimizes the error probability is given by the *maximum a posterior* (MAP) estimator:

$$\widehat{\sigma}_i = \arg\max_{s\in\{\pm1\}}\mathbb{P}\left[\sigma_i = s|T_i^{(t)}, \sigma_{L_i^{(t)}}\right].$$

[3] To 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 ith-gen children.

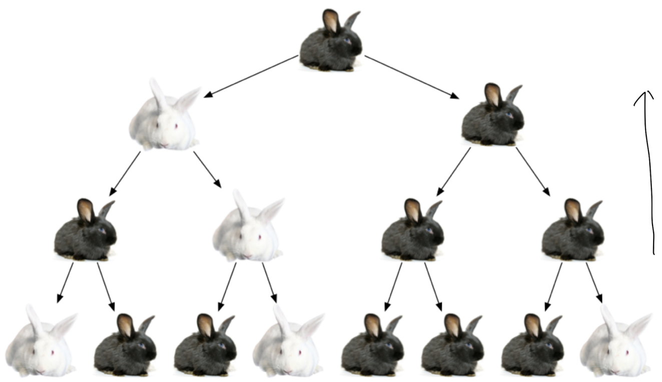


Figure 5: The reconstruction problem on trees. With probability λ, the color is copied from parent to child; with probability $1-\lambda$, the child's color is uniformly random. Many generations later, the majority of the population has the same color as the root with high probability if and only if $c\lambda^2 > 1$. The rabbits in this experiment have $c = 2$ children at each step and $\lambda = 0.28$, below the Kesten-Stigum threshold.

Here $\lambda = \frac{a-b}{a+b}$ and $c = \frac{a+b}{2}$

Figure 9.1: Extracted from [Moo17]

Moreover, this MAP estimator can be efficiently computed via BP (9.2) initialized as

$$\Lambda^{(0)}_{\ell\to i} = \begin{cases} +\infty & \text{if } \sigma_\ell = + \\ -\infty & \text{if } \sigma_\ell = - \end{cases} \qquad \forall \ell \in L_i^{(t)}$$

9.4.1 Density evolution and Gaussian approximation

In this section, we analyze the belief propagation (9.2) by characterizing the density evolution of the BP messages. Define

$$U_\pm^{(t)} \overset{d}{=} \Lambda^{(t)}_{i\to j} \text{ conditional on } \sigma_i = \pm 1.$$

By symmetry, the above holds for $i \to j$ replaced by any children-parent pair $\ell \to i$. Moreover, for all different children nodes $\ell \in \partial i \setminus j$, since the subtrees rooted at ℓ are disjoint, it follows that $\Lambda^{(t)}_{\ell\to i}$ are independent conditional on σ_i. Therefore, according to the BP update rule (9.2),

$$U_\pm^{(t)} \overset{d}{=} \sum_{\ell=1}^{N_+} f\left(U^{(t-1)}_{\pm,\ell}\right) + \sum_{\ell=1}^{N_-} f\left(U^{(t-1)}_{\mp,\ell}\right), \tag{9.9}$$

where $N_+ \sim \text{Poi}(a/2)$, $N_- \sim \text{Poi}(b/2)$, $U^{(t)}_{\pm,\ell}$ are i.i.d. copies of $U^{(t)}_\pm$, and they are mutually independent.

Remark 9.4. Note that (9.9) gives a recursion of the distribution of $U^t_\pm$, which is known as *density evolution*. We make two important remarks in order. First, the above derivation crucially exploits the recursive structure of the trees. It turns out that such derivation can be generalized to locally-tree graphs, see e.g. [HWX15]. Second, characterizing the density evolution is a key to the analysis of BP, which is hard in general, because the recursion maps between (infinite-dimensional) probability distributions on $\mathbb{R}$. Sometimes, if we are lucky, the density evolution can collapse to a recursion of a finite dimensional object. One way to obtain such dimension reduction is via Gaussian approximation.

Gaussian approximation in large degree asymptotic

Let us first state informally the high-level idea. If the average degree parameters a, b are relatively large, then $U_{\pm}^{(t)}$ are Poisson sums of many i.i.d. random variables, which are approximately Gaussian due to the central limit theorem. Since Gaussian density is determined by mean and variance, we only need to characterize the recursions of mean and variance.

To this end, let us assume $a, b \to \infty$ while keeping $\tau = \frac{(a-b)^2}{2(a+b)}$ to be a fixed constant in the sequel. The following lemma characterizes the recursions of mean and variance of $U_{\pm}^{(t)}$.

Lemma 9.3. *For all $t \geq 1$, it holds that*

$$\mathbb{E}\left[U_{\pm}^{(t)}\right] = \pm\tau\mathbb{E}\left[\tanh(U_{+}^{(t-1)})\right] + O\left(a^{-1/2}\right), \tag{9.10}$$

$$\mathrm{Var}\left(U_{\pm}^{(t)}\right) = \tau\mathbb{E}\left[\tanh(U_{+}^{(t-1)})\right] + O\left(a^{-1/2}\right). \tag{9.11}$$

The next lemma from [KS12, Theorem 3] is an analog of Berry-Esseen theorem for Poisson sum.

Lemma 9.4. *Let $S_v = X_1 + \ldots + X_{N_v}$, where X_i's are i.i.d. with finite second moment and $\mathbb{E}\left[|X_i|^3\right] \leq \rho^3$, and $N_v \sim \mathrm{Poi}(v)$ for some $v > 0$. Then*

$$\sup_x \left|\mathbb{P}\left[\frac{S_v - v\mathbb{E}\left[X_1\right]}{\sqrt{v\mathbb{E}\left[X_1^2\right]}} \leq x\right] - \mathbb{P}\left[Z \leq x\right]\right| \leq \frac{C_{\mathrm{BE}}\rho^3}{\sqrt{v\left(\mathbb{E}\left[X_1^2\right]\right)^3}},$$

where $C_{\mathrm{BE}} = 0.3041$ and $Z \sim \mathcal{N}(0,1)$.

The above two lemmas together imply that

$$U_{\pm}^{(t)} \overset{(d)}{\approx} \mathcal{N}\left(\pm\tau\mathbb{E}\left[\tanh(U_{+}^{(t-1)})\right], \tau\mathbb{E}\left[\tanh(U_{+}^{(t-1)})\right]\right).$$

More formally, we have the following theorem.

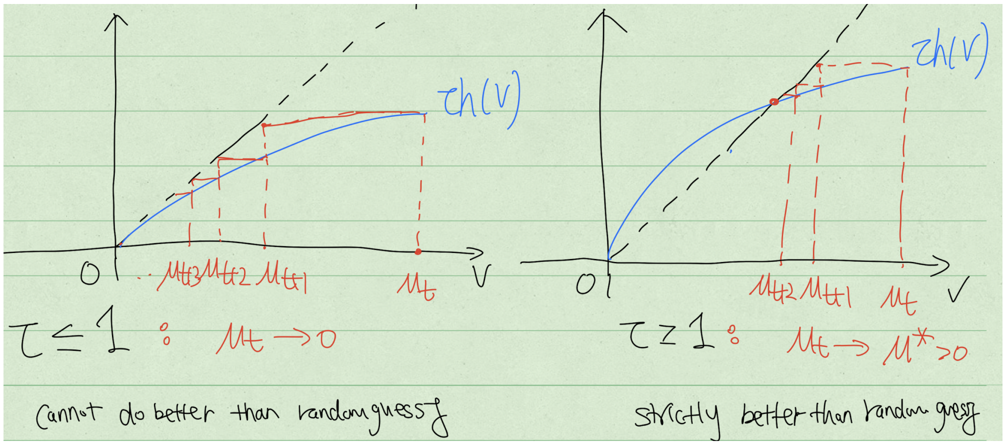
Theorem 9.5. *Define $(\mu_t : t \geq 0)$ recursively by*

$$\mu_0 = \infty \text{ and } \mu_t = \tau\mathbb{E}\left[\tanh\left(\mu_{t-1} + \sqrt{\mu_{t-1}}Z\right)\right]$$

for $Z \sim \mathcal{N}(0,1)$. Then for any $t \geq 1$:

$$\sup_x \left|\mathbb{P}\left[\frac{U_{\pm}^{(t)} - \pm\mu_t}{\sqrt{\mu_t}} \leq x\right] - \mathbb{P}\left[Z \leq x\right]\right| = O(a^{-1/2}).$$

The above theorem shows that $U_{\pm}^{(t)}$ is approximately distributed as $\mathcal{N}(\pm\mu_t, \mu_t)$. Let $h(v) \triangleq \mathbb{E}\left[\tanh\left(v + \sqrt{v}Z\right)\right]$. One can show that $h(v)$ is continuous on $[0, \infty)$ and $0 \leq h'(v) \leq 1$ for $v \in (0, +\infty)$ with $h'(0) = 1$. One can also prove that $h(v)$ is strictly concave [DAM15, Lemma 6.1]. Therefore, the recursion $\mu_t = \tau h(\mu_{t-1})$ exhibits a phase transition at threshold $\tau = 1$.



- If $\tau \le 1$, then as $t \to \infty$, $\mu_t \to 0$ and hence $\mathbb{P}\left[\sigma_i = + \mid T_i^{(t)}, \sigma_{L_i^{(t)}}\right] \to \frac{1}{2}$, i.e., it is impossible to do strictly better than random guessing.
- If $\tau > 1$, then as $t \to \infty$, $\mu_t \to \mu^* > 0$ and hence $\mathbb{P}\left[\sigma_i = \pm \mid T_i^{(t)}, \sigma_{L_i^{(t)}}\right] > \frac{1}{2}$ conditional on $\sigma_i = \pm$, achieving the correlated recovery.

Proof of Lemma 9.3

Proof. By definition and symmetry between $+1$ and -1, we have

$$\Lambda_{i\to j}^{(t)} \text{ conditional on } \sigma_i = - \overset{(d)}{=} -u_{i\to j}^{(t)} \text{ conditional on } \sigma_i = +.$$

Therefore, $U_-^{(t)} \overset{(d)}{=} -U_+^{(t)}$ and hence it suffices to prove the lemma for U_-^t. Moreover,

$$\exp\left(2\Lambda_{i\to j}^{(t)}\right) = \frac{\mathbb{P}\left[\sigma_i = +1 | T_i^{(t)}, \sigma_{L_i^{(t)}}\right]}{\mathbb{P}\left[\sigma_i = -1 | T_i^{(t)}, \sigma_{L_i^{(t)}}\right]} = \frac{\mathbb{P}\left[T_i^{(t)}, \sigma_{L_i^{(t)}} | \sigma_i = +1\right]}{\mathbb{P}\left[T_i^{(t)}, \sigma_{L_i^{(t)}} | \sigma_i = -1\right]},$$

where the last equality holds because the prior distribution of σ_i is uniform. Therefore, by change of measure, for any measurable function g,

$$\mathbb{E}\left[g\left(\Lambda_{i\to j}^{(t)}\right) | \sigma_i = -1\right] = \mathbb{E}\left[g\left(\Lambda_{i\to j}^{(t)}\right) \exp\left(-2\Lambda_{i\to j}^{(t)}\right) | \sigma_i = +1\right].$$

It follows that

$$\mathbb{E}\left[g(U_-^{(t)})\right] = \mathbb{E}\left[g(U_+^{(t)}) \exp\left(-2U_+^{(t)}\right)\right]. \tag{9.12}$$

Now, we are ready to compute $\mathbb{E}\left[U_-^{(t)}\right]$. By Taylor expansion,

$$\begin{aligned} f(x) &= \tanh^{-1}[\tanh(\beta)\tanh(x)] \\ &= \frac{1}{2}\log\left(\frac{\exp(2x+2\beta)+1}{\exp(2x)+\exp(2\beta)}\right) \\ &= -\beta + \frac{\exp(4\beta)-1}{2} g(x) - \frac{(\exp(4\beta)-1)^2}{4} g^2(x) + O\left(\left|e^{4\beta}-1\right|^3\right), \end{aligned}$$

where $g(x) = 1/\left[1+e^{-2(x-\beta)}\right]$. Thus it follows from (9.9) that

$$\begin{aligned} U_-(t) = &-\beta\left(N_+ + N_-\right) + \frac{\exp(4\beta)-1}{2}\left[\sum_{\ell=1}^{N_-} g\left(U_{+,\ell}^{(t-1)}\right) + \sum_{\ell=1}^{N_+} g\left(U_{-,\ell}^{(t-1)}\right)\right] \\ &- \frac{(\exp(4\beta)-1)^2}{4}\left[\sum_{\ell=1}^{N_-} g^2\left(U_{+,\ell}^{(t-1)}\right) + \sum_{\ell=1}^{N_+} g^2\left(U_{-,\ell}^{(t-1)}\right)\right] + O\left(\left|e^{4\beta}-1\right|^3\right) \times \left(N_+ + N_-\right). \end{aligned}$$

Taking expectation yields that

$$\begin{aligned} \mathbb{E}\left[U_-^{(t)}\right] = &-\beta\frac{a+b}{2} + \frac{\exp(4\beta)-1}{4}\left(b\mathbb{E}\left[g\left(U_+^{(t-1)}\right)\right] + a\mathbb{E}\left[g\left(U_-^{(t-1)}\right)\right]\right) \\ &- \frac{(\exp(4\beta)-1)^2}{8}\left(b\mathbb{E}\left[g^2\left(U_+^{(t-1)}\right)\right] + a\mathbb{E}\left[g^2\left(U_-^{(t-1)}\right)\right]\right) + O\left(\left|e^{4\beta}-1\right|^3\right) \times (a+b). \end{aligned}$$

In view of (9.12) and $\beta = \frac{1}{2}\log\frac{a}{b}$, we have

$$b\mathbb{E}\left[g\left(U_+^{(t-1)}\right)\right] + a\mathbb{E}\left[g\left(U_-^{(t-1)}\right)\right] = b\mathbb{E}\left[g\left(U_+^{(t-1)}\right) + e^{2\beta}\exp\left(-2U_+^{(t-1)}\right)g\left(U_+^{(t-1)}\right)\right] = b,$$

where the last equality holds due to the definition of g. Similarly,

$$b\mathbb{E}\left[g^2\left(U_+^{(t-1)}\right)\right] + a\mathbb{E}\left[g^2\left(U_-^{(t-1)}\right)\right] = b\mathbb{E}\left[g^2\left(U_+^{(t-1)}\right) + e^{2\beta}\exp\left(-2U_+^{(t-1)}\right)g^2\left(U_+^{(t-1)}\right)\right] = b\mathbb{E}\left[g\left(U_+^{(t-1)}\right)\right].$$

Combining the last three displayed equation gives that

$$\mathbb{E}\left[U_-^{(t)}\right] = -\beta\frac{a+b}{2} + \frac{\exp(4\beta)-1}{4}b - \frac{(\exp(4\beta)-1)^2}{8}b\mathbb{E}\left[g\left(U_+^{t-1}\right)\right] + O\left(\left|e^{4\beta}-1\right|^3\right) \times (a+b).$$

Now, using Taylor expansion, we have

$$\beta = \frac{1}{2}\log\frac{a}{b} = \frac{a-b}{2b} - \frac{(a-b)^2}{4b^2} + O\left(\frac{|a-b|^3}{b^3}\right)$$

and

$$e^{4\beta} - 1 = 4\beta + \frac{1}{2}(4\beta)^2 + O(\beta^3).$$

Since $\tau = \frac{(a-b)^2}{2(a+b)}$ is a fixed constant, it follows that

$$-\beta\frac{a+b}{2} + \frac{\exp(4\beta)-1}{4}b = \frac{(a-b)^2}{2(a+b)} + O(a^{-1/2}) = \tau + O(a^{-1/2})$$

and

$$\frac{(\exp(4\beta)-1)^2}{8}b = 2\tau + O(a^{-1/2})$$

and

$$\left|e^{4\beta}-1\right|^3(a+b) = O(a^{-1/2}).$$

Finally, note that

$$\left|g(x) - \frac{1}{1+\exp(-2x)}\right| \le |\exp(2\beta)-1| = O(a^{-1/2}).$$

Assembling all these together yields that

$$\mathbb{E}\left[U_-^{(t)}\right] = \tau - 2\tau\mathbb{E}\left[\frac{1}{1+\exp\left(-2U_+^{(t-1)}\right)}\right] + O(a^{-1/2}) = -\tau\mathbb{E}\left[\tanh\left(U_+^{(t-1)}\right)\right] + O(a^{-1/2}).$$

Analogously, we can show that

$$\operatorname{Var}\left(U_-^{(t)}\right) = \tau\mathbb{E}\left[\tanh(U_+^{(t-1)})\right] + O\left(a^{-1/2}\right).$$

□

9.5 Concluding remarks

We have been focusing exclusively on the symmetric two-community case for correlated recovery. It turns out that most of the theory developed here can be generalized to multiple communities with extra effort. With multiple k communities, the Kesten-Stigum threshold becomes

$$\frac{(a-b)^2}{k\left(a+(k-1)b\right)} > 1.$$

However, for large k ($k \geq 5$), this no longer coincides with the information-theoretic threshold for correlated recovery. In particular, we observe the following intriguing "easy-hard-impossible" phase transition diagram [DKMZ11, BMNN16, AS15]:

- If
$$\frac{(a-b)^2}{k\left(a+(k-1)b\right)} > 1,$$
correlated recovery is easy, in particular can be achieved via non-backtracking matrix;

- If
$$\frac{\log k}{k} \lesssim \frac{(a-b)^2}{k\left(a+(k-1)b\right)} \leq 1,$$
correlated recovery is information-theoretically possible, but conjectured to be computationally hard;

- If
$$\frac{\log k}{k} \lesssim \frac{(a-b)^2}{k\left(a+(k-1)b\right)} \lesssim \frac{\log k}{k},$$
correlated recovery is information-theoretically impossible.

We consider a two-community SBM model in a very broad sense (cf. Lecture 7). Let $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\} \in \{\pm 1\}^n$ be the community labels of nodes. The weighted adjacency matrix is $A = (A_{ij})$, where $\{A_{ij} : 1 \le i < j \le n\}$ are independent conditioned on σ, such that

$$A_{ij} \sim \begin{cases} P & \sigma_i = \sigma_j \\ Q & \text{o.w.} \end{cases}.$$

Let $d_H(x, y) = \sum_i \mathbb{1}\{x_i \neq y_i\}$ denote the Hamming distance . Consider the loss function

$$\ell(\sigma, \widehat{\sigma}) = \min\{d_H(\sigma, \widehat{\sigma}), d_H(\sigma, -\widehat{\sigma})\},$$

which is the number of misclassified vertices (up to a global relabeling). We say

1. $\widehat{\sigma}$ achieves *almost exact* recovery if $\mathbb{E}\ell(\sigma, \widehat{\sigma}) = o(n)$;
2. $\widehat{\sigma}$ achieves *exact* recovery if $\ell(\sigma, \widehat{\sigma}) = 0$ w.h.p.

Under mild assumptions on the distributions P, Q, we will see that the requirements for these two types of recovery are $H^2(P, Q) \gg \frac{1}{n}$ for almost exact recovery, and $H^2(P, Q) \ge \frac{(2+\epsilon)\log n}{n}$ for some $\epsilon > 0$ for exact recovery. In this lecture, we will focus on the exact recovery.

10.1 MLE and SDP relaxation

The log-likelihood is

$$\begin{aligned} \log p(A|\sigma) &= \sum_{i,j} \log p(A_{ij}|\sigma_i, \sigma_j) \\ &= \sum_{i,j} \log p(A_{ij}) \mathbb{1}\{\sigma_i = \sigma_j\} + \log q(A_{ij}) \mathbb{1}\{\sigma_i \neq \sigma_j\} \\ &= \sum_{i,j} \frac{\log p(A_{ij}) + \log q(A_{ij})}{2} + \frac{\log p(A_{ij}) - \log q(A_{ij})}{2} \sigma_i \sigma_j, \end{aligned}$$

where p, q are the densities of P, Q w.r.t some dominating measure. And we assume $P \ll Q$. Define the likelihood ratio (LLR) matrix W with $W_{ij} = \log \frac{p}{q}(A_{ij})$ for $i \neq j$ and $W_{ii} = 0$ for convenience. Then MLE is equivalently formulated as solving

$$\max_{\sigma \in \{\pm 1\}^n} \langle W, \sigma\sigma^\top \rangle. \tag{10.1}$$

This is closely related with the "min-cut" problem. For a weighted graph $G = (V, E, W)$, a cut is a partition of V into two disjoint subsets S and S^c. The value of a cut (S, S^c) is the total edge weights between the two subsets:

$$\sum_{i \in S, j \in S^c} W_{ij}. \tag{10.2}$$

We can define $\sigma_i = 2\mathbb{1}\{i \in S\} - 1 \in \{\pm 1\}$ to be the label of cut, then the cut value is

$$\sum_{i,j} W_{ij} \left(\frac{1 - \sigma_i \sigma_j}{2} \right).$$

So we can see that, solving for MLE of SBM (10.1) is equivalent to searching for the min-cut for the complete graph weighted with the LLR matrix W.

Like what we did in the clique problem (Lecture 6), we can reformulate (10.1) as

$$\begin{aligned} \max \ & \langle W, X \rangle \\ \text{s.t. } & X \succeq 0 \\ & \operatorname{diag}(X) = \mathbf{I} \\ & \operatorname{rank}(X) = 1. \end{aligned}$$

Dropping the rank-one constraint, we arrive at the following SDP relaxation:

$$\begin{aligned} \max \ & \langle W, X \rangle \\ \text{s.t. } & X \succeq 0 \\ & \operatorname{diag}(X) = \mathbf{I}. \end{aligned} \tag{10.3}$$

For bisection (two equally sized communities), we have $\langle \sigma, \mathbf{1} \rangle = 0$ and we can further strengthen the above min-cut SDP (10.3) by adding a hard constraint $\langle X, \mathbf{J} \rangle = 0$ and get that

$$\begin{aligned} \max \ & \langle W, X \rangle \\ \text{s.t. } & X \succeq 0 \\ & \operatorname{diag}(X) = \mathbf{I} \\ & \langle X, \mathbf{J} \rangle = 0. \end{aligned} \tag{10.4}$$

Theorem 10.1 (Min-cut SDP)**.** *Assume that*

$$H^2(P, Q) \geq \frac{2(1+\epsilon)\log n}{n}, \quad \textit{for a fixed constant } \epsilon > 0 \tag{10.5}$$

$$\|W - \mathbb{E}W\|_{\text{op}} = o_p(\log n) \tag{10.6}$$

$$\operatorname{Var}(W_{ij}) = o\left(\frac{\log^2 n}{n} \right) \tag{10.7}$$

$$D(P\|Q) - D(Q\|P) = O\left(H^2(P, Q) \log n \right). \tag{10.8}$$

Then w.h.p., the unique solution to the min-cut SDP (10.3) is $\sigma\sigma^\top$.

Theorem 10.2 (Min-bisection SDP)**.** *Consider bisection $\langle \sigma, \mathbf{1} \rangle = 0$. Suppose the conditions (10.5) and (10.6) hold. Then w.h.p., the unique solution to the min-bisection SDP (10.4) is $\sigma\sigma^\top$.*

Remark 10.1. The assumptions of the preceeding theorem may be interpreted as follows:

- The assumption (10.5) is fundamental since it is the information-theoretical limit that guarantees the MLE to succeed, and it is also necessary under further regularity conditions on the distributions P, Q, e.g., for Bernoulli and Gaussians.[1]

- The assumption (10.6) on the spectral deviation of the weight matrix is the main assumption for SDP relaxation to succeed. To appreciate this assumption, let us compute a lower bound to $\|\mathbb{E}[W]\|_{\text{op}}$. Specifically, $\mathbb{E}[W]$ (plus multiples of diagonal) is a rank-two matrix, with
$$\mathbb{E}[W_{ij}] = \begin{cases} D(P\|Q) & \sigma_i = \sigma_j \\ -D(Q\|P) & \sigma_i \neq \sigma_j. \end{cases} \tag{10.9}$$
Thus $(\mathbb{E}[W]\sigma)_i = kD(P\|Q) + (n-k)D(Q\|P)$ if $\sigma_i = +1$ and $-kD(Q\|P) - (n-k)D(P\|Q)$ if $\sigma_i = -1$, where k is the number of $+1$'s in σ. Note that the KL divergence dominates the Hellinger distance[2]:
$$\min\{D(P\|Q), D(Q\|P)\} \geq 2H^2(P\|Q). \tag{10.10}$$
Thus $\|\mathbb{E}[W]\|_{\text{op}} \geq n \min\{D(P\|Q), D(Q\|P)\} \overset{(10.5)}{=} \Omega(\log n)$. This means $\mathbb{E}[W]$ is on the scale of $\log n$ which dominates $(W - \mathbb{E}[W])$, thanks to (10.6).

- The assumptions (10.7) and (10.8) are for technical reasons. Note that (10.7) is consistent with (10.6), because for Wigner matrix we expect the spectral norm is $\sqrt{n \cdot \text{Var}}$.

Remark 10.2 (Specialization to Gaussian weights)**.** Suppose $P = \mathcal{N}(\mu, 1)$ and $Q = \mathcal{N}(0, 1)$ with $\mu > 0$. Then
$$H^2(P, Q) = 2 - 2\int \sqrt{dPdQ} = 2 - 2\exp(-\mu^2/8).$$
For μ small, by Taylor's expansion the above is approximately $\mu^2/4$. Hence assumption (10.5) roughly translates to
$$\mu^2 \geq \frac{8(1+\epsilon)\log n}{n}.$$

On the other hand, we argue that $\mu^2 \geq 8 \log n/n$ is necessary for the MLE to achieve exact recovery (both for the min-cut MLE and the min-bisection MLE). Consider an assignment vector $\widetilde{\sigma}$ that reverses the labels of i, j in σ for some $\sigma_i = +1$ and $\sigma_j = -1$. Then we have
$$\langle A, \sigma\sigma^T\rangle - \langle A, \widetilde{\sigma}\widetilde{\sigma}^T\rangle \sim \mathcal{N}(4n\mu, 32n).$$
There are $\Theta(n^2)$ such pairs of i, j, which yield $\Theta(n^2)$ random variables distributed $\mathcal{N}(4n\mu, 32n)$ that are almost mutually independent.[3] For the MLE to achieve exact recovery, with high probability all these r.v.'s must be nonnegative. Therefore we must have $4n\mu - \sqrt{32n}\sqrt{2\log(n^2)} \geq 0$, or equivalently, $\mu^2 \geq 8 \log n/n$.

[1] Related calculations are carried out in [HWX17] for the single-community model with general distributions.

[2] To see this, note that
$$D(P\|Q) = -2\mathbb{E}_P\left[\log\sqrt{\frac{dQ}{dP}}\right] \geq -2\mathbb{E}_P\left[\frac{\sqrt{dQ}-\sqrt{dP}}{\sqrt{dP}}\right] = 2\left(1 - \int\sqrt{dPdQ}\right) = 2H^2(P, Q),$$
where the inequality holds due to $\log(1+x) \leq x$.

[3] It is also easy to select $\Theta(n^2)$ pairs of i, j for which the Gaussian random variables are exactly independent.

Remark 10.3 (Specialization to SBM)**.** For SBM(n, p, q), set $p = \frac{\alpha \log n}{n}, q = \frac{\beta \log n}{n}$. Then

$$H^2(P, Q) = (\sqrt{p} - \sqrt{q})^2 + (\sqrt{1-p} - \sqrt{1-q})^2 = (\sqrt{\alpha} - \sqrt{\beta})^2 \frac{\log n}{n}(1 + o(1)).$$

So the condition for Hellinger distance in Theorem 10.1 is $\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$.

For the LLR:

$$\begin{aligned} W_{ij} &= \log \frac{p}{q} \mathbb{1}\{A_{ij} = 1\} + \log \frac{1-p}{1-q} \mathbb{1}\{A_{ij} = 0\} \\ &= \log \frac{p(1-q)}{q(1-p)} A_{ij} + \log \frac{1-p}{1-q}. \end{aligned}$$

Thus

$$W = \log \frac{p(1-q)}{q(1-p)} \cdot A + \log \frac{1-p}{1-q} \cdot \mathbf{J}. \tag{10.11}$$

where $\mathbf{J}$ is the all-ones matrix. Thus the SDP (10.3) is equivalent with the following penalized form:

$$\begin{aligned} \max \ & \langle A, X \rangle + \tau \langle \mathbf{J}, X \rangle \\ \text{s.t. } & X \succeq 0 \\ & \operatorname{diag}(X) = \mathbf{I}. \end{aligned}$$

where $\tau = \frac{\log \frac{1-p}{1-q}}{\log \frac{p(1-q)}{q(1-p)}}$. For bisection, the SDP (10.4) is equivalent to

$$\begin{aligned} \max \ & \langle A, X \rangle \\ \text{s.t. } & X \succeq 0 \\ & \operatorname{diag}(X) = \mathbf{I} \\ & \langle X, \mathbf{J} \rangle = 0. \end{aligned}$$

Finally, from (10.11) assuming $\log \frac{p(1-q)}{q(1-p)}$ is bounded from above and away from 0, we have

$$\|W - \mathbb{E}W\|_{\text{op}} = o(\log n) \Leftrightarrow \|A - \mathbb{E}A\|_{\text{op}} = o(\log n).$$

Indeed, $p, q = O(\frac{\log n}{n})$, it can be shown (cf. e.g. [HWX16, Theorem 5]) that w.h.p,

$$\|A - \mathbb{E}A\|_{\text{op}} \lesssim \sqrt{\log n}.$$

This is consistent with the Gaussian heuristic in Lecture 4 that $\|A - \mathbb{E}A\|_{\text{op}} \lesssim \sqrt{n \cdot \text{variance}}$.

10.2 Proof of Theorem 10.2

In this section, we give the proof of Theorem 10.2 for min-bisection SDP; Theorem 10.1 for min-cut SDP is proved analogously but with a bit more detailed calculations in the next section.

The following deterministic lemma gives a sufficient condition for (10.4) to achieve the exact recovery.

Lemma 10.1 (Duality)**.** *$X^* = \sigma\sigma^\top$ is the unique maximizer of (10.4) if $\exists D = \operatorname{diag}(d_i)$, $S \succeq 0$, $\lambda \in \mathbb{R}$, s.t.*

$$S = D - W + \lambda \mathbf{J}, \tag{10.12}$$

$$S\sigma = 0, \tag{10.13}$$

$$\lambda_{n-1}(S) > 0.$$

Proof of Lemma 10.1. The proof follows the usual route of KKT conditions (cf. Section 6.2). Define the following function as the Lagrangian form of (10.4) with specified parameters

$$L(X, D, S, \lambda) = \langle W, X\rangle + \langle S, X\rangle + \mathrm{Tr}(D) - \langle D, X\rangle - \lambda\langle \mathbf{J}, X\rangle.$$

For any feasible X in (10.4), $\mathrm{Tr}(D) - \langle D, X\rangle - \lambda\langle \mathbf{J}, X\rangle = 0$, $\langle S, X\rangle \geq 0$. So

$$\langle W, X\rangle \leq L(X, D, S, \lambda) \overset{(10.12)}{=} L(X^*, D, S, \lambda) \overset{(10.13)}{=} \langle W, X^*\rangle.$$

Thus X^* is an optimal solution to (10.4).

Finally, we prove the uniqueness. Note that if $\langle W, X\rangle = \langle W, X^*\rangle$, then $\langle S, X\rangle = 0$. But $\lambda_{n-1}(S) > 0$, so the column space of X is spanned by σ, which means $X = cX^*$. Finally, $c = 1$ since $\mathrm{diag}(X) = \mathbf{I}$. □

With Lemma 10.1, to prove Theorem 10.2, it suffices to construct (D, λ) and verify that the conditions of Lemma 10.1 are satisfied with high probability.

Proof of Theorem 10.2. From (10.12) and (10.13) and $\langle \sigma, \mathbf{1}\rangle = 0$, we know $D\sigma = W\sigma$, which means

$$d_i = \sum_j W_{ij}\sigma_i\sigma_j.$$

It remains to show S as defined in (10.12) satisfies w.h.p. $\lambda_{n-1}(S) > 0$, that is

$$\inf_{x\perp\sigma, \|x\|_2=1} x^\top(D - W + \lambda\mathbf{J})x > 0. \tag{10.14}$$

Note that by the assumption (10.6), $|x^\top(W - \mathbb{E}W)x| \leq \|W - \mathbb{E}W\|_{\mathrm{op}} = o_p(\log n)$. Thus it suffices to show

$$\inf_{x\perp\sigma, \|x\|_2=1} x^\top(D - \mathbb{E}[W] + \lambda\mathbf{J})x = \Omega_p(\log n). \tag{10.15}$$

Write $s = D(P\|Q), t = D(Q\|P)$ as shorthand. By (10.9), we have

$$\mathbb{E}W = \frac{s-t}{2}\mathbf{J} + \frac{s+t}{2}\sigma\sigma^\top - s\mathbf{I}.$$

Thus to show (10.15), it suffices to show

$$\inf_{x\perp\sigma, \|x\|_2=1} x^\top\left(D - \frac{s-t}{2}\mathbf{J} + \lambda\mathbf{J}\right)x = \Omega_p(\log n). \tag{10.16}$$

We finish the proof in two steps. First, we choose $\lambda \geq \frac{s-t}{2}$ so that $\left(-\frac{s-t}{2} + \lambda\right)x^\top\mathbf{J}x \geq 0$. Then, we show that under the information-theoretic condition (10.5), $\min_{i\in[n]} d_i \geq \epsilon(1+\epsilon)\log n$ with high probability. This is where the condition (10.5) is from.

Lemma 10.2. *Consider two-community SBM with possibly different community sizes. Assume that (10.5) holds. Then w.h.p,* $\min_{i\in[n]} d_i \geq \frac{1}{2}\epsilon n H^2(P, Q) \geq \epsilon(1+\epsilon)\log n$.

□

10.2.1 Proof of Lemma 10.2: Large deviation

This is a good exercise on the (information-theoretic flavored) large deviation analysis. Assume that $H^2(P,Q) \geq \frac{2(1+\epsilon)\log n}{n}$. By union bound, it suffices to show

$$\mathbb{P}(d_i \leq c\log n) = o(1/n)$$

for each $i \in [n]$ and $c = \frac{\epsilon n H^2(P,Q)}{2\log n}$. Let us focus on d_i for a node i with $\sigma_i = +1$ and simply call it d; the case for a node i with $\sigma_i = -1$ can be proved analogously.

Define X and Y be distributed as the law of $\log\frac{dP}{dQ}$ under P and Q, respectively. Then

$$d \overset{\mathcal{D}}{=} \sum_{i=1}^{k} X_i - \sum_{i=1}^{n-k} Y_i, \quad \mathbb{E}d = kD(P\|Q) + (n-k)D(Q\|P),$$

where X_i are iid copies of X and Y_i are iid copies of Y. To apply Chernoff bound, denote the log moment generating function (log MGF) of X and Y as

$$\psi_P(\theta) = \log\mathbb{E}[e^{\theta X}] = \log\int p^{1+\theta}q^{-\theta}, \quad \psi_Q(\theta) = \log\mathbb{E}[e^{\theta Y}] = \log\int p^{\theta}q^{1-\theta} = \psi_P(\theta-1). \tag{10.17}$$

For any $\theta > 0$,

$$\begin{aligned}
\mathbb{P}\left(d \leq c\log n\right) =& \mathbb{P}\left(\sum_{i=1}^{n-k} Y_i - \sum_{i=1}^{k} X_i \geq -c\log n\right) \\
=& \mathbb{P}\left(\exp\left\{\theta\sum_{i=1}^{n-k} Y_i - \theta\sum_{i=1}^{k} X_i\right\} \geq \exp(-\theta c\log n)\right) \\
\leq& \mathbb{E}\exp\left(\theta\sum_{i=1}^{n-k} Y_i - \theta\sum_{i=1}^{k} X_i + \theta c\log n\right) \\
=& \exp\left((n-k)\psi_Q(\theta) + k\psi_P(-\theta) + \theta c\log n\right) \\
=& \exp\left((n-k)\psi_P(\theta-1) + k\psi_P(-\theta) + \theta c\log n\right).
\end{aligned}$$

Choose $\theta = \frac{1}{2}$,[4] note that $\psi_P(-1/2) = \log\int\sqrt{PQ} = \log(1 - \frac{H^2(P,Q)}{2})$, and recall that $c = \frac{\epsilon n H^2(P,Q)}{2\log n}$. Then we have as desired

$$\begin{aligned}
\mathbb{P}\left[d \leq 0\right] &\leq \exp\left\{n\log\left(1 - \frac{H^2(P,Q)}{2}\right) + \frac{\epsilon n H^2(P,Q)}{4}\right\} \\
&\overset{(a)}{\leq} \exp\left\{-\frac{(1-\epsilon/2)nH^2(P,Q)}{2}\right\} \overset{(b)}{\leq} n^{-(1-\epsilon/2)(1+\epsilon)} = o(1/n),
\end{aligned}$$

where (a) follows from $\log(1-x) \leq -x$ for $x \in [0,1]$ and (b) follows from the assumption (10.5).

Remark 10.4 (Optimality)**.** We explain heuristically the sharp threshold expressed in terms of the Hellinger distance:

[4] We chose $\theta = 1/2$ independent of k. To see why, consider the special case of bisection where $k = n/2$. Due to the convexity of the log MGF ψ_P, for any θ, $\psi_P(\theta-1) + \psi_P(-\theta) \geq 2\psi_P(-1/2)$, with equality if $\theta = 1/2$, which is the optimal choice.

- Exact recovery: Under further assumptions, the condition (10.5) on $H^2(P,Q)$ is necessary for MLE (and hence any method) to achieve exact recovery. Indeed, consider another solution $\widetilde{\sigma}$ that differs from σ only on a single coordinate. Then the potential increment of the likelihood is $\langle W, \widetilde{\sigma}\widetilde{\sigma}^\top - \sigma\sigma^\top \rangle \overset{\text{D}}{=} -4d_i$ defined above. We have showed that $H^2(P,Q) = \frac{2(1+\epsilon)}{n}\log n$ ensures that $\min d_i > 0$. If the above large deviation analysis is tight, then in the opposite condition $H^2(P,Q) = \frac{2(1-\epsilon)}{n}\log n$, we will have $\min d_i < 0$, meaning MLE will make a mistake.
- Almost exact recovery: Consider the *oracle* situation when $\sigma_2, \dots, \sigma_n$ are observed, with roughly half $+$ and half $-$, and the only goal is to estimate σ_1. This is equivalent to testing the hypothesis of $H_0 : \sigma_1 = +$ versus $H_1 : \sigma_1 = -$. In this case, only the first row of A is useful. Hence σ_1 cannot be tested with vanishing probability of error, if

  $$TV(P^{\otimes \frac{n}{2}} \otimes Q^{\otimes \frac{n}{2}}, Q^{\otimes \frac{n}{2}} \otimes P^{\otimes \frac{n}{2}}) \le 1 - c \Leftrightarrow H^2(P \otimes Q, Q \otimes P) = O(1/n) \Leftrightarrow H^2(P,Q) = O(1/n),$$

  where the first equivalence follows from the following two generic facts: $\frac{1}{2}H^2(P,Q) \le TV(P,Q) \le H(P,Q)\sqrt{1 - H^2(P,Q)/4}$ and $H^2(P^{\otimes n}, Q^{\otimes n}) = 2 - 2\left(1 - H^2(P,Q)/2\right)^n$.

10.3 Proof of Theorem 10.1

The following deterministic lemma gives a sufficient condition for (10.3) to achieve the exact recovery.

Lemma 10.3 (Duality)**.** *$X^* = \sigma\sigma^\top$ is the unique maximizer of* (10.3) *if $\exists D = \operatorname{diag}(d_i)$ s.t.*

$$S = D - W, \tag{10.18}$$

$$S\sigma = 0, \tag{10.19}$$

$$\lambda_{n-1}(S) > 0.$$

The proof is identical to the proof of Lemma 10.1 with λ setting to be 0. With Lemma 10.3, to prove Theorem 10.1, it suffices to construct D and verify that the conditions of Lemma 10.3 are satisfied with high probability.

Proof of Theorem 10.1. The proof is almost identical to the proof of Theorem 10.2 with λ setting to be 0. The only difference is in showing

$$\inf_{x \perp \sigma, \|x\|_2 = 1} x^\top \left(D - \frac{s-t}{2}\mathbf{J}\right) x = \Omega(\log n). \tag{10.20}$$

To evaluate this minimum, let us define a unit vector ξ such that

$$\xi_i = \begin{cases} \sqrt{\frac{n-k}{nk}} & \text{if } \sigma_i = +1, \\ \sqrt{\frac{k}{n(n-k)}} & \text{if } \sigma_i = -1, \end{cases}. \tag{10.21}$$

Then it is straightforward to verify that $\operatorname{span}(\sigma, \mathbf{1}) = \operatorname{span}(\sigma, \xi)$, consisting of vectors which takes constant values on each of the two communities. For any feasible x in (10.20), we have $x = \cos\theta\xi + \sin\theta z$ for some unit vector $z \in \operatorname{span}(\sigma, \mathbf{1})^\perp$.

Expanding the quadratic form, we have

$$x^\top \left(D - \frac{s-t}{2}\mathbf{J}\right) x = \cos^2\theta \cdot (\mathrm{I}) + 2\cos\theta\sin\theta \cdot (\mathrm{II}) + \sin^2\theta \cdot (\mathrm{III}),$$

where

- $$(\mathrm{I}) = \xi^\top \left(D - \frac{s-t}{2}\mathbf{J}\right)\xi \geq \xi^\top \left(\mathbb{E}D - \frac{s-t}{2}\mathbf{J}\right)\xi - \|(D-\mathbb{E}D)\xi\|_2 = nt - \|(D-\mathbb{E}D)\xi\|_2.$$

  Here $\xi^\top \mathbf{J}\xi = \langle \xi, \mathbf{1}\rangle^2 = \frac{4k(n-k)}{n}$, $\xi^\top \mathbb{E}D\xi = \sum_i \mathbb{E}[d_i]\xi_i^2 = [(n-k)d_+ + kd_-]/n = \frac{s+t}{2} + \frac{t-s}{2n}(n-2k)^2$, with $\mathbb{E}[d_i] = \frac{s+t}{2}n + \sigma_i \frac{s-t}{2}(2k-n)$, so that (by some miracle) $\xi^\top \left(\mathbb{E}D - \frac{s-t}{2}\mathbf{J}\right)\xi = nt$.

- $$|(\mathrm{II})| = \left|\xi^\top \left(D - \frac{s-t}{2}\mathbf{J}\right) z\right| = \left|\xi^\top (D - \mathbb{E}D) z\right| \leq \|(D-\mathbb{E}D)\xi\|_2,$$

  where we used $z \perp \mathbf{1}$ and $\xi^\top \mathbb{E}Dz = \langle z, \mathbb{E}D\xi\rangle = 0$, since the entries of $\mathbb{E}D\xi$ are constant in each of the two communities.

- $$(\mathrm{III}) = z^\top \left(D - \frac{s-t}{2}\mathbf{J}\right) z = z^\top D z \geq \min_i d_i.$$

Overall, we have

$$\inf_{x\perp\sigma, \|x\|_2=1} x^\top \left(D - \frac{s-t}{2}\mathbf{J}\right) x \geq \min\{nt, \min_i d_i\} - 3\|(D-\mathbb{E}D)\xi\|_2 = \Omega(\log n),$$

where $nt = nD(Q\|P) \overset{(10.10)}{\geq} 2nH^2(P,Q) \overset{(10.5)}{=} \Omega(\log n)$, $\min_i d_i = \Omega(\log n)$ w.h.p by Lemma 10.2, and the following lemma.

Lemma 10.4. *Under (10.7),* $\|(D-\mathbb{E}D)\xi\|_2 = o_p(\log n)$.

This completes the proof of Theorem 10.1. □

10.3.1 Proof of Lemma 10.4

Write

$$\|(D-\mathbb{E}D)\xi\|_2^2 = \xi^\top (D-\mathbb{E}D)^2 \xi = \sum_i \xi_i^2 \left(\sum_j (W_{ij} - \mathbb{E}W_{ij})\sigma_j\right)^2.$$

Take expected value:

$$\mathbb{E}\|(D-\mathbb{E}D)\xi\|_2^2 = \sum_i \xi_i^2 \sum_j \mathrm{Var}(W_{ij}) = o\left(\log^2 n\right),$$

where the last equality is due to (10.7). It follows that

$$\mathbb{E}\|(D-\mathbb{E}D)\xi\|_2 \leq \left(\mathbb{E}\|(D-\mathbb{E}D)\xi\|_2^2\right)^{1/2} = o(\log n),$$

and by Chebyshev's inequality,

$$\|(D-\mathbb{E}D)\xi\|_2 - \mathbb{E}\|(D-\mathbb{E}D)\xi\|_2 = o_p(\log n).$$

Therefore $\|(D-\mathbb{E}D)\xi\|_2 = \mathbb{E}\|(D-\mathbb{E}D)\xi\|_2 + (\|(D-\mathbb{E}D)\xi\|_2 - \mathbb{E}\|(D-\mathbb{E}D)\xi\|_2) = o_p(\log n)$.

§ 11. Ranking from Pairwise Comparisons (by Cheng Mao)

Ranking from comparisons arises in various applications, including recommender systems, social choice and sports tournament. We consider the following setup. Suppose that there are items $1, \dots, n$ associated with unknown ranks $\pi^*(1), \dots, \pi^*(n)$, where $\pi^* : [n] \to [n]$ is a permutation. Observing a set of pairwise comparisons, each of the form $i \succ j$ meaning that "item i beats item j", we aim to recover the ranking π^*.

11.1 Modeling pairwise comparisons

We first give an overview of common models for ranking from pairwise comparisons.

11.1.1 Models for probabilities of outcomes

Each pairwise comparison is a Bernoulli outcome. Let us denote the probability that the item at rank k beats the item at rank ℓ by $M_{k,\ell}$ where $M \in \mathbb{R}^{n\times n}$, so that

$$\mathbb{I}\{i \succ j\} \sim \mathsf{Ber}(M_{\pi^*(i),\pi^*(j)}).$$

In the sequel, we present several models on the matrix M of probabilities. It is vacuous to compare an item to itself, so we assume without loss of generality that $M_{i,i} = 1/2$ for $i \in [n]$. Moreover, we consider the case that there is one and only one winner in a pairwise comparison, so it always holds that $M_{k,\ell} + M_{\ell,k} = 1$.

Parametric models Parametric models assume that for $i \in [n]$, item i is associated with a strength parameter $\theta_i \in \mathbb{R}$, and

$$M_{\pi^*(i),\pi^*(j)} = F(\theta_i - \theta_j)$$

where $F : \mathbb{R} \to (0, 1)$ is a known, increasing link function. Two classical examples are the logistic function $F(x) = \frac{1}{1+e^{-x}}$ and the Gaussian cumulative density function, which correspond to the Bradley-Terry model and the Thurstone model respectively.

Noisy sorting The noisy sorting model [BM08] assumes that

$$M_{k,\ell} = \begin{cases} 1/2 + \lambda & \text{if } k > \ell, \\ 1/2 - \lambda & \text{if } k < \ell. \end{cases} \tag{11.1}$$

This is the model we focus on later, as it is simple yet captures important concepts and tools.

Strong stochastic transitivity Strong stochastic transitivity (SST) means that for any triplet $(k, \ell, m) \in [n]^3$ such that $k < \ell < m$, we have

$$M_{k,m} \geq M_{k,\ell} \vee M_{\ell,m}.$$

In matrix terminology, this is saying that M is bivariate isotonic (bi-isotonic) in addition to the constraint $M + M = \mathbf{1}\mathbf{1}^\top$. More precisely, all the columns of M are nonincreasing while all the rows of M are nondecreasing. Note that any parametric model, as well as the noisy sorting model, satisfies SST.

11.1.2 Sampling models

We consider uniform sampling. Namely, for $m \in [N]$ where N is the sample size, we observe independent outcomes

$$y_m \sim \mathsf{Ber}(M_{\pi^*(i_m),\pi^*(j_m)}), \tag{11.2}$$

where the random pairs (i_m, j_m) are sampled uniformly randomly with replacement from all possible pairs $\{(i,j)\}_{i \neq j}$. Here $y_m = 1$ means that $i_m \succ j_m$ and $y_m = 0$ means that $j_m \succ i_m$. We collect the outcomes of comparisons in a matrix $A \in \mathbb{R}^{n \times n}$ whose entry $A_{i,j}$ is defined to be the number of times item i beats item j.

Note that for parametric models, we have for $m \in [N]$,

$$\mathbb{E}[y_m] = F(\theta_{i_m} - \theta_{j_m}) = F\big(x_m^\top \theta\big),$$

where $x_m = e_{i_m} - e_{j_m}$ is the design point. This is simply the setup of generalized linear regression. Particularly, the Bradley-Terry model is essentially logistic regression with this special design.

11.2 Kendall's tau and minimax rates for noisy sorting

In general, we would like to estimate both π^* and M, but let us focus on estimating π^* under the noisy sorting model (11.1) for the rest of the notes. Full details of the discussion can be found in the paper [MWR18].

Consider the Kendall tau distance, i.e., the number of inversions between permutations, defined as

$$d_{\mathsf{KT}}(\pi, \sigma) = \sum_{i,j \in [n]} \mathbb{I}\big(\pi(i) > \pi(j),\, \sigma(i) < \sigma(j)\big).$$

Note that $d_{\mathsf{KT}}(\pi, \sigma) \in [0, \binom{n}{2}]$ and it is equal to the minimum number of adjacent transpositions required to change from π to σ (think of bubble sort). A closely related distance is the ℓ_1-distance, also known as Spearman's footrule, defined as

$$\|\pi - \sigma\|_1 = \sum_{i=1}^{n} |\pi(i) - \sigma(i)|.$$

It is well known [DG77] that

$$d_{\mathsf{KT}}(\pi, \sigma) \leq \|\pi - \sigma\|_1 \leq 2 d_{\mathsf{KT}}(\pi, \sigma). \tag{11.3}$$

Theorem 11.1. *Consider the noisy sorting model (11.1) with $\lambda \in (0, \frac{1}{2} - c]$ where c is a positive constant. Suppose N independent comparisons are given according to (11.2). Then it holds that*

$$\min_{\widetilde{\pi}} \max_{\pi^*} \mathbb{E}_{\pi^*}[d_{\mathsf{KT}}(\widetilde{\pi}, \pi^*)] \asymp \frac{n^3}{N\lambda^2} \wedge n^2.$$

11.2.1 Inversions and metric entropy

Before proving the theorem, we study the metric entropy of the set of permutations $\mathfrak{S}_n$ with respect to the Kendall tau distance d_{KT}. Let $\mathcal{B}(\pi, r) = \{\sigma \in \mathfrak{S}_n : d_{\mathsf{KT}}(\pi, \sigma) \le r\}$.

The inversion table $b_1, \dots, b_n$ of a permutation $\pi \in \mathfrak{S}_n$ is defined by

$$b_i = \sum_{j:i<j} \mathbb{I}\big(\pi(i) > \pi(j)\big).$$

Note that $b_i \in \{0, 1, \dots, n-i\}$ and $d_{\mathsf{KT}}(\pi, \mathsf{id}) = \sum_{i=1}^n b_i$. On can reconstruct a permutation using its inversion table $\{b_i\}_{i=1}^n$, so the set of inversion tables is bijective to $\mathfrak{S}_n$. (Try the permutation $(3\,5\,2\,4\,1)$ which has inversion table $(4\,2\,0\,1\,0)$.)

Lemma 11.1. *For $0 \le k \le \binom{n}{2}$, we have that*

$$n \log(k/n) - n \le \log |\mathcal{B}(\mathsf{id}, k)| \le n \log(1 + k/n) + n\,.$$

Proof. According to the discussion above, $|\mathcal{B}(\mathsf{id}, k)|$ is equal to the number of inversion tables $b_1, \dots, b_n$ such that $\sum_{i=1}^n b_i \le k$ where $b_i \in \{0, 1, \dots, n-i\}$. On the one hand, if $b_i \le \lfloor k/n \rfloor$ for all $i \in [n]$, then $\sum_{i=1}^n b_i \le k$, so a lower bound is given by

$$\begin{aligned}
|\mathcal{B}(\mathsf{id}, k)| &\ge \prod_{i=1}^n (\lfloor k/n \rfloor + 1) \wedge (n - i + 1) \\
&\ge \prod_{i=1}^{n - \lfloor k/n \rfloor} (\lfloor k/n \rfloor + 1) \prod_{i=n-\lfloor k/n \rfloor + 1}^{n} (n - i + 1) \\
&\ge (k/n)^{n - k/n} \lfloor k/n \rfloor !\,.
\end{aligned}$$

Using Stirling's approximation, we see that

$$\begin{aligned}
\log |\mathcal{B}(\mathsf{id}, k)| &\ge n \log(k/n) - (k/n) \log(k/n) + \lfloor k/n \rfloor \log \lfloor k/n \rfloor - \lfloor k/n \rfloor \\
&\ge n \log(k/n) - n\,.
\end{aligned}$$

On the other hand, if b_i is only required to be a nonnegative integer for each $i \in [n]$, then we can use a standard "stars and bars" counting argument to get an upper bound

$$|\mathcal{B}(\mathsf{id}, k)| \le \binom{n+k}{n} \le e^n (1 + k/n)^n\,.$$

Taking the logarithm finishes the proof. □

For $\varepsilon > 0$ and $S \subseteq \mathfrak{S}_n$, let $N(S, \varepsilon)$ and $D(S, \varepsilon)$ denote respectively the ε-covering number and the ε-packing number of S with respect to d_{KT}.

Proposition 11.1. *We have that for $\varepsilon \in (0, r)$,*

$$n \log\big(\frac{r}{n + \varepsilon}\big) - 2n \le \log N(\mathcal{B}(\pi, r), \varepsilon) \le \log D(\mathcal{B}(\pi, r), \varepsilon) \le n \log\big(\frac{2n + 2r}{\varepsilon}\big) + 2n\,.$$

For $n \lesssim \varepsilon < r \le \binom{n}{2}$, the ε-metric entropy of $\mathcal{B}(\pi, r)$ scales as $n \log \frac{r}{\varepsilon}$. In other words, $\mathfrak{S}_n$ equipped with d_{KT} is a doubling space[1] with doubling dimension $\Theta(n)$.

[1] A metric space (X, d) is called a doubling space with doubling dimension $\log_2 M$, if M is the smallest number such that any ball of radius r in (X, d) can be covered with M balls of radius $r/2$.

Proof. The relation between the covering and the packing number is standard. We employ a volume argument for the bounds. Let $\mathcal{P}$ be a 2ε-packing of $\mathcal{B}(\pi, r)$ so that the balls $\mathcal{B}(\sigma, \varepsilon)$ are disjoint for $\sigma \in \mathcal{P}$. By the triangle inequality, $\mathcal{B}(\sigma, \varepsilon) \subseteq \mathcal{B}(\pi, r+\varepsilon)$ for each $\sigma \in \mathcal{P}$. By the invariance of the Kendall tau distance under composition, Lemma 11.1 yields

$$\begin{aligned}\log D(\mathcal{B}(\pi, r), 2\varepsilon) &\le n \log(1 + r/n) + n - n\log(\varepsilon/n) + n \\ &= n \log\big(\frac{n+r}{\varepsilon}\big) + 2n\,.\end{aligned}$$

In addition, if $\mathcal{N}$ is an ε-net of $\mathcal{B}(\pi, r)$, then the set of balls $\{\mathcal{B}(\sigma, \varepsilon)\}_{\sigma \in \mathcal{N}}$ covers $\mathcal{B}(\pi, r)$. By Lemma 11.1, we obtain

$$\begin{aligned}\log N(\mathcal{B}(\pi, r), \varepsilon) &\ge \log|\mathcal{B}(\pi, r)| - \log|\mathcal{B}(\sigma, \varepsilon)| \\ &\ge n\log(r/n) - n - n\log(1+\varepsilon/n) - n \\ &= n\log\big(\frac{r}{n+\varepsilon}\big) - 2n\,,\end{aligned}$$

as claimed. □

11.2.2 Proof of the minimax upper bound

We only present the proof of the upper bound in Theorem 11.1 with $\lambda = 1/4$ for simplicity. The estimator we use is a sieve maximum likelihood estimator (MLE), meaning that it is the MLE over a net (called a sieve). More precisely, define $\varphi = \frac{n}{N}\binom{n}{2}$. Let $\mathcal{P}$ be a maximal φ-packing (and thus a φ-net) of $\mathfrak{S}_n$ with respect to d_{KT}. Consider the sieve MLE

$$\widehat{\pi} \in \operatorname*{argmax}_{\pi \in \mathcal{P}} \sum_{\pi(i) < \pi(j)} A_{i,j}\,. \tag{11.4}$$

Basic setup Since $\mathcal{P}$ is a φ-net, there exists $\sigma \in \mathcal{P}$ such that $D \triangleq d_{\mathsf{KT}}(\sigma, \pi^*) \le \varphi$. By definition of $\widehat{\pi}$, $\sum_{\widehat{\pi}(i)<\widehat{\pi}(j)} A_{i,j} \ge \sum_{\sigma(i)<\sigma(j)} A_{i,j}$. Canceling concordant pairs (i,j) under $\widehat{\pi}$ and σ, we see that

$$\sum_{\widehat{\pi}(i)<\widehat{\pi}(j),\, \sigma(i)>\sigma(j)} A_{i,j} \ge \sum_{\widehat{\pi}(i)>\widehat{\pi}(j),\, \sigma(i)<\sigma(j)} A_{i,j}\,.$$

Splitting the summands according to π^* yields that

$$\sum_{\substack{\widehat{\pi}(i)<\widehat{\pi}(j),\\ \sigma(i)>\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j} + \sum_{\substack{\widehat{\pi}(i)<\widehat{\pi}(j),\\ \sigma(i)>\sigma(j),\\ \pi^*(i)>\pi^*(j)}} A_{i,j} \ge \sum_{\substack{\widehat{\pi}(i)>\widehat{\pi}(j),\\ \sigma(i)<\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j} + \sum_{\substack{\widehat{\pi}(i)>\widehat{\pi}(j),\\ \sigma(i)<\sigma(j),\\ \pi^*(i)>\pi^*(j)}} A_{i,j}\,.$$

Since $A_{i,j} \ge 0$, we may drop the rightmost term and drop the condition $\widehat{\pi}(i) < \widehat{\pi}(j)$ in the leftmost term to obtain that

$$\sum_{\substack{\sigma(i)>\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j} + \sum_{\substack{\widehat{\pi}(i)<\widehat{\pi}(j),\\ \sigma(i)>\sigma(j),\\ \pi^*(i)>\pi^*(j)}} A_{i,j} \ge \sum_{\substack{\widehat{\pi}(i)>\widehat{\pi}(j),\\ \sigma(i)<\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j}\,. \tag{11.5}$$

To set up the rest of the proof, we define, for $\pi \in \mathcal{P}$,

$$\begin{aligned}L_\pi &= |\{(i,j) \in [n]^2 : \pi(i) < \pi(j), \sigma(i) > \sigma(j), \pi^*(i) > \pi^*(j)\}| \\ &= |\{(i,j) \in [n]^2 : \pi(i) > \pi(j), \sigma(i) < \sigma(j), \pi^*(i) < \pi^*(j)\}|\,.\end{aligned}$$

Moreover, define the random variables

$$X_\pi = \sum_{\substack{\pi(i)>\pi(j),\\ \sigma(i)<\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j}\,, \quad Y_\pi = \sum_{\substack{\pi(i)<\pi(j),\\ \sigma(i)>\sigma(j),\\ \pi^*(i)>\pi^*(j)}} A_{i,j}\,, \quad \text{and} \quad Z = \sum_{\substack{\sigma(i)>\sigma(j),\\ \pi^*(i)<\pi^*(j)}} A_{i,j}.$$

We show that the random process $X_\pi - Y_\pi - Z$ is positive with high probability if $d_{\mathsf{KT}}(\pi,\sigma)$ is large.

Binomial tails For a single pairwise comparison sampled uniformly from the possible $\binom{n}{2}$ pairs, the probability that

1. the chosen pair (i,j) satisfies $\pi(i) > \pi(j)$, $\sigma(i) < \sigma(j)$ and $\pi^*(i) < \pi^*(j)$, *and*
2. item i wins the comparison,

is equal to $\frac{3}{4}L_\pi\binom{n}{2}^{-1}$. By definition, X_π is the number of times the above event happens if N independent pairwise comparisons take place, so $X_\pi \sim \mathsf{Bin}\big(N, \frac{3}{4}L_\pi\binom{n}{2}^{-1}\big)$. Similarly, we have $Y_\pi \sim \mathsf{Bin}\big(N, \frac{1}{4}L_\pi\binom{n}{2}^{-1}\big)$ and $Z \sim \mathsf{Bin}\big(N, \frac{3}{4}D\binom{n}{2}^{-1}\big)$. The tails of a Binomial random variable can be bounded by the following lemma.

Lemma 11.2. *For $0 < r < p < s < 1$ and $X \sim \mathsf{Bin}(N,p)$, we have*

$$\mathbb{P}(X \le rN) \le \exp\Big(-N\frac{(p-r)^2}{2p(1-r)}\Big) \quad \textit{and} \quad \mathbb{P}(X \ge sN) \le \exp\Big(-N\frac{(p-s)^2}{2s(1-p)}\Big).$$

Therefore, we obtain

1. $\mathbb{P}\big(X_\pi \le \frac{5}{8}L_\pi N\binom{n}{2}^{-1}\big) \le \exp\big(-L_\pi N\binom{n}{2}^{-1}/128\big)$,
2. $\mathbb{P}\big(Y_\pi \ge \frac{3}{8}L_\pi N\binom{n}{2}^{-1}\big) \le \exp\big(-L_\pi N\binom{n}{2}^{-1}/128\big)$, and
3. $\mathbb{P}\big(Z \ge 2\varphi N\binom{n}{2}^{-1}\big) \le \exp\big(-\varphi N\binom{n}{2}^{-1}/4\big) = \exp(-n/4)\,.$

Then we have that

$$\mathbb{P}\big(X_\pi - Y_\pi \le \tfrac{1}{4}L_\pi N\tbinom{n}{2}^{-1}\big) \le 2\exp\big(-L_\pi N\tbinom{n}{2}^{-1}/128\big)\,. \tag{11.6}$$

Peeling and union bounds For an integer $r \in [C\varphi, \binom{n}{2}]$ where C is a sufficiently large constant to be chosen, consider the slice $\mathcal{S}_r = \{\pi \in \mathcal{P} : L_\pi = r\}$. Note that if $\pi \in \mathcal{S}_r$, then

$$\begin{aligned} d_{\mathsf{KT}}(\pi,\pi^*) &= |\{(i,j) : \widehat{\pi}(i) < \widehat{\pi}(j), \pi^*(i) > \pi^*(j)\}| \\ &\le |\{(i,j) : \widehat{\pi}(i) < \widehat{\pi}(j), \sigma(i) > \sigma(j), \pi^*(i) > \pi^*(j)\}| \\ &\quad + |\{(i,j) : \sigma(i) < \sigma(j), \pi^*(i) > \pi^*(j)\}| \\ &= L_\pi + d_{\mathsf{KT}}(\sigma,\pi^*) \le r + \varphi\,, \end{aligned} \tag{11.7}$$

showing that $\mathcal{S}_r \subseteq \mathcal{B}(\pi^*, r+\varphi)$. Therefore, Proposition 11.1 gives

$$\log|\mathcal{S}_r| \le n\log\frac{2n+2r+2\varphi}{\varphi} + 2n \le n\log\frac{45r}{\varphi}\,.$$

By (11.6) and a union bound over $\mathcal{S}_r$, we have $\min_{\pi \in \mathcal{S}_r}(X_\pi - Y_\pi) > \frac{1}{4} r N \binom{n}{2}^{-1}$ with probability

$$1 - \exp\left(n \log \frac{45r}{\varphi} + \log 2 - \frac{rN}{128\binom{n}{2}}\right) \geq 1 - \exp(-2n)\,,$$

where the inequality holds by the definition of φ and the range of r. Then a union bound over integers $r \in [C\varphi, \binom{n}{2}]$ yields that

$$X_\pi - Y_\pi > \frac{C}{4}\varphi N \binom{n}{2}^{-1}$$

for all $\pi \in \mathcal{P}$ such that $L_\pi \geq C\varphi$ with probability at least $1 - e^{-n}$. This is larger than the above high probability upper bound on Z, so we conclude that with probability at least $1 - e^{-n/8}$,

$$X_\pi - Y_\pi - Z > 0$$

for all $\pi \in \mathcal{P}$ with $L_\pi \geq C\varphi$. However, (11.5) says that $X_{\widehat{\pi}} - Y_{\widehat{\pi}} - Z \leq 0$, so $L_{\widehat{\pi}} \leq C\varphi$ on the above event. By (11.7), $d_{\mathsf{KT}}(\widehat{\pi}, \pi^*) \leq L_{\widehat{\pi}} + \varphi$ on the same event, which completes the proof.

11.3 An efficient algorithm for noisy sorting

Let us move on to present an efficient algorithm. We continue to assume $\lambda = 1/4$. To recover the underlying order of items, it is equivalent to estimate the row sums $\sum_{j=1}^{n} M_{\pi^*(i),\pi^*(j)}$ which we call scores of the items. Initially, for each $i \in [n]$, we estimate the score of item i by the number of wins item i has. If item i has a much higher score than item j in the first stage, then we are confident that item i is stronger than item j. Hence in the second stage, we know $M_{\pi^*(i),\pi^*(j)} = 3/4$ with high probability. For those pairs that we are not certain about, $M_{\pi^*(i),\pi^*(j)}$ is still estimated by its empirical version. The variance of each score is thus greatly reduced in the second stage, thereby yielding a more accurate order of the items. Then we iterate this process to obtain finer and finer estimates of the scores and the underlying order.

To present the T-stage sorting algorithm formally, we split the sample into T subsamples each containing N/T pairwise comparisons. For $t \in [T]$, we define a matrix $A^{(t)} \in \mathbb{R}^{n \times n}$ by setting $A^{(t)}_{i,j}$ to be the number of times item i beats item j in the t-th sample. The algorithm proceeds as follows:

1. For $i \in [n]$, define $I^{(0)}(i) = [n]$, $I^{(0)}_-(i) = \varnothing$ and $I^{(0)}_+(i) = \varnothing$. For $0 \leq t \leq T$, we use $I^{(t)}(i)$ to denote the set of items j whose ranking relative to i has not been determined by the algorithm at stage t.

2. At stage t, compute the score $S^{(t)}_i$ of item i:

$$S^{(t)}_i = \frac{T\binom{n}{2}}{N} \sum_{j \in I^{(t-1)}(i)} A^{(t)}_{i,j} + \frac{3}{4}\left|I^{(t-1)}_-(i)\right| + \frac{1}{4}\left|I^{(t-1)}_+(i)\right|.$$

3. Set the threshold

$$\tau^{(t)}_i \asymp n\sqrt{|I^{(t-1)}(i)| T N^{-1} \log(nT)}\,,$$

and define the sets

$$\begin{aligned} I^{(t)}_+(i) &= \{j \in [n] : S^{(t)}_j - S^{(t)}_i < -\tau^{(t)}_i\}, \\ I^{(t)}_-(i) &= \{j \in [n] : S^{(t)}_j - S^{(t)}_i > \tau^{(t)}_i\}, \text{ and} \\ I^{(t)}(i) &= [n] \setminus \left(I^{(t)}_-(i) \cup I^{(t)}_+(i)\right). \end{aligned}$$

4. Repeat step 2 and 3 for $t = 1, \dots, T$. Output a permutation $\widehat{\pi}^{\mathsf{MS}}$ by sorting the scores $S_i^{(T)}$ in nonincreasing order, i.e., $S_i^{(T)} \geq S_j^{(T)}$ if $\widehat{\pi}^{\mathsf{MS}}(i) < \widehat{\pi}^{\mathsf{MS}}(j)$.

We take $T = \lfloor \log\log n \rfloor$ so that the overall time complexity of the algorithm is only $O(n^2 \log\log n)$.

Theorem 11.2. *With probability at least $1 - n^{-7}$, the algorithm with $T = \lfloor \log\log n \rfloor$ stages outputs an estimator $\widehat{\pi}^{\mathsf{MS}}$ that satisfies*

$$\|\widehat{\pi}^{\mathsf{MS}} - \pi^*\|_\infty \lesssim \frac{n^2}{N} (\log n) \log\log n$$

and

$$d_{\mathsf{KT}}(\widehat{\pi}^{\mathsf{MS}}, \pi^*) \lesssim \frac{n^3}{N} (\log n) \log\log n\,.$$

The second statement follows from the first one together with (11.3).

11.3.1 Proof (sketch) of Theorem 11.2

Assume that $\pi^* = \mathsf{id}$ without loss of generality. We define a score

$$s_i^* = \sum_{j \in [n] \setminus \{i\}} M_{i,j} = \frac{i}{2} + \frac{n}{4} - \frac{3}{4}$$

for each $i \in [n]$, which is simply the i-th row sum of M minus $1/2$.

Lemma 11.3. *Fix $t \in [T]$, $I \subseteq [n]$ and $i \in I$. Let us define*

$$S = \frac{T\binom{n}{2}}{N} \sum_{j \in I} A_{i,j}^{(t)} + \frac{3}{4}\big|\{j \in [n] \setminus I : j < i\}\big| + \frac{1}{4}\big|\{j \in [n] \setminus I : j > i\}\big|\,.$$

If $|I|$ is not too small, then it holds with probability at least $1 - (nT)^{-9}$ that

$$|S - s_i^*| \lesssim n\sqrt{|I| T N^{-1} \log(nT)}\,.$$

Proof. The probability that a uniform pair consists of item i and an item in $I \setminus \{i\}$, and that item i wins the comparison, is equal to $q \triangleq \big(\sum_{j \in I \setminus \{i\}} M_{i,j}\big)/\binom{n}{2}$. Thus the random variable $X \triangleq \sum_{j \in I} A_{i,j}^{(t)}$ has distribution $\mathsf{Bin}(N/T, q)$. In particular, we have $\mathbb{E}[X] = Nq/T = \frac{N}{T\binom{n}{2}} \sum_{j \in I \setminus \{i\}} M_{i,j}$, so S is an unbiased estimate of s_i^*. Moreover, we have the tail bound

$$\mathbb{P}\Big(\big|X - \mathbb{E}[X]\big| \gtrsim \sqrt{qNT^{-1} \log(nT)}\Big) \leq (nT)^{-9}\,,$$

from which the conclusion follows. □

We apply Lemma 11.3 inductively to each stage of the algorithm. By a union bound over all $i \in [n]$ and $t \in [T]$, all the events studied below hold with high probability. For $t \in [T]$, define

$$\mathcal{E}^{(t-1)} \triangleq \big\{j < i \text{ for all } j \in I_-^{(t-1)}(i) \text{ and } j > i, \text{ for all } j \in I_+^{(t-1)}(i)\big\}.$$

On the event $\mathcal{E}^{(t-1)}$, the score $S_i^{(t)}$ is exactly the quantity S in Lemma 11.3 with $I = I^{(t-1)}(i)$, so

$$|S_i^{(t)} - s_i^*| \lesssim n\sqrt{|I^{(t-1)}(i)| T N^{-1} \log(nT)} = \tau_i^{(t)}/2. \qquad (11.8)$$

For any $j \in I_+^{(t)}(i)$, by definition $S_j^{(t)} - S_i^{(t)} < -\tau_i^{(t)}$, so we have $s_j^* < s_i^*$ and thus $j > i$. Similarly, $j < i$ for any $j \in I_-^{(t)}(i)$. Hence $\mathcal{E}^{(t)}$ occurs with high probability. Moreover, if $|s_j^* - s_i^*| > 2\tau_i^{(t)}$, then $|S_j^{(t)} - S_i^{(t)}| > \tau_i^{(t)}$, so $j \notin I^{(t)}(i)$. Hence if $j \in I^{(t)}(i)$, then $|j - i| \lesssim \tau_i^{(t)}$. Consequently,

$$|I^{(t)}(i)| \lesssim \tau_i^{(t)} \lesssim n\sqrt{|I^{(t-1)}(i)| T N^{-1} \log(nT)}\,. \tag{11.9}$$

Note that if we have $\alpha^{(0)} = n$ and the iterative relation $\alpha^{(t)} \le \beta\sqrt{\alpha^{(t-1)}}$ where $\alpha^{(t)} > 0$ and $\beta > 0$, then it is easily seen that $\alpha^{(t)} \le \beta^2 n^{2^{-t}}$. Consequently, we obtain that

$$|I^{(T-1)}(i)| \lesssim \frac{n^2 T}{N} \log(nT) n^{2^{-T+1}} \lesssim \frac{n^2}{N} (\log n)(\log\log n)$$

for $T = \lfloor \log\log n \rfloor$. Taking T to be larger does not make $|I^{(T-1)}(i)|$ smaller, because Lemma 11.3 requires a lower bound on $|I^{(T-1)}(i)|$. The details are left out. It follows from (11.8) that

$$|S_i^{(T)} - s_i^*| \lesssim \frac{n^2}{N} (\log n) \log\log n =: \delta\,.$$

As the permutation $\widehat{\pi}^{\mathsf{MS}}$ is defined by sorting the scores $S_i^{(T)}$ in nonincreasing order, we see that $\widehat{\pi}^{\mathsf{MS}}(i) < \widehat{\pi}^{\mathsf{MS}}(j)$ for all pairs (i, j) with $s_i^* - s_j^* > 2\delta$, i.e., $j - i > \delta$.

Finally, suppose that $\widehat{\pi}^{\mathsf{MS}}(i) - i > \delta$ for some $i \in [n]$. Then there exists $j > i + \delta$ such that $\widehat{\pi}^{\mathsf{MS}}(j) < \widehat{\pi}^{\mathsf{MS}}(i)$, contradicting the guarantee we have just proved. A similar argument leads to a contradiction if $\widehat{\pi}^{\mathsf{MS}}(i) - i < -\delta$. Therefore, we obtain that $|\widehat{\pi}^{\mathsf{MS}}(i) - i| \le \delta$, completing the proof.

§ 12. Grothendieck Inequality and its Application in SBM

Previously in Lecture 10, we discussed the exact recovery of SBM. In this lecture, we turn to the almost exact recovery. Let σ be the community labels of nodes. Recall that the loss to evaluate a community estimate $\widehat{\sigma}$ is $l(\sigma, \widehat{\sigma}) = \frac{1}{n} \min_{s \in \{\pm\}} d_H(\sigma, s\widehat{\sigma})$, where $d_H(x, y)$ is the Hamming distance $\sum_i \mathbb{1}\{x_i \neq y_i\}$. We call $\widehat{\sigma}$ an almost exact recovery if $\mathbb{E} l(\sigma, \widehat{\sigma}) = o(1)$, and an Exact recovery if $l(\sigma, \widehat{\sigma}) = 0$ w.h.p. We have seen that the requirement is $H^2(P, Q) \geq \frac{(2+\epsilon)\log n}{n}$ for $\forall \epsilon > 0$ for exact recovery. Here we are going to show the necessary and sufficient condition for almost exact recovery is $H^2(P, Q) \gg \frac{1}{n}$, which can be achieved by SDP relaxation.

We first introduce the key technical tool: Grothendieck Inequality (Theorem 12.1). Then we discuss its application to SBM following Guédon-Vershynin [GV16].

12.1 $\|\cdot\|_{\infty\to 1}$ norm

Consider $A \in \mathbb{R}^{n\times m}$. We look at the following optimization

$$\max_{x_i, y_j \in \{\pm\}} \sum_{1\leq i\leq n, 1\leq j\leq m} a_{ij} x_i y_j = \max_{x\in\{\pm\}^n, y\in\{\pm\}^m} \langle A, xy^\top \rangle. \tag{12.1}$$

Remark 12.1. The objective above (12.1) is a norm of A, denoted as $\|A\|_{\infty\to 1} = \max_{\|x\|_\infty \leq 1} \|Ax\|_1$. This is easily seen by writing $\|\cdot\|_1$ in the dual form.

Moreover, when A is PSD, $A = U^\top U$ and hence $\langle A, xy^\top \rangle = \langle Ux, Uy \rangle \leq \|Ux\|_2 \|Uy\|_2$, where the inequality is met with equality when $x = y$. Thus, $\|A\|_{\infty\to 1} = \max_{x\in\{\pm\}^n, y\in\{\pm\}^m} \langle A, xy^\top \rangle = \max_{x\in\{\pm\}^n} \langle A, xx^\top \rangle$.

Remark 12.2. $\|A\|_{\infty\to 1}$ is closely related to the cut norm. The cut norm $\|A\|_{\text{cut}}$ is defined as (cf. the min cut in Lecture 10)

$$\|A\|_{\text{cut}} = \max_{I\subset[n], J\subset[m]} \left| \sum_{i\in I, j\in J} a_{ij} \right|.$$

The relation of the two norms is

$$\|A\|_{\text{cut}} \leq \|A\|_{\infty\to 1} \leq 4\|A\|_{\text{cut}}.$$

The left side inequality can be seen by

$$\|A\|_{\text{cut}} = \max_{I\subset[n], J\subset[m]} \left| \sum_{i\in I, j\in J} a_{ij} \right| \leq \max_{I\subset[n], J\subset[m]} \sum_{i\in I} \left| \sum_{j\in J} a_{ij} \right| \leq \max_J \|Ax_J\|_1 \leq \|A\|_{\infty\to 1}.$$

x_J is the indicator vector of J. The right side inequality can be shown by writing $x = x_I - x_{I^c}$, $y = y_J - y_{J^c}$ in (12.1).

We have the SDP relaxation of $\|A\|_{\infty\to 1}$: for $r \geq n+m$ (otherwise it is nonconvex),

$$\mathrm{SDP}(A) = \max_{u_i, v_j \in \mathbb{R}^r, \|u_i\|=\|v_j\|=1} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \langle u_i, v_j \rangle. \tag{12.2}$$

When A is PSD, since $\|A\|_{\infty\to 1} = \max_{x\in\{\pm\}^n} \langle A, xx^\top \rangle$, we can just take $u_i = v_i$.

Remark 12.3. When $r = 1$, $\mathrm{SDP}(A)$ corresponds to $\|A\|_{\infty\to 1}$. Thus it is indeed a "relaxation" of the norm: $\|A\|_{\infty\to 1} \leq \mathrm{SDP}(A)$.

Remark 12.4 (Dimension-free)**.** $\mathrm{SDP}(A)$ is dimension-free in the sense that the value does not depend on r as long as $r \geq n+m$. In particular, if it helps construction, we are free to consider the infinite-dimensional setting, e.g., the decision variables u_i, v_j take values in the Hilbert space of random variables – and we will do so next.

Remark 12.5 (Standard Form)**.** $\mathrm{SDP}(A)$ can be written into a standard SDP form

$$\mathrm{SDP}(A) = \max_{X \succeq 0, X_{ii}=1} \langle W, X \rangle$$

where $W = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix}$. The correspondence is by writing

$$X = \begin{pmatrix} U^\top \\ V^\top \end{pmatrix} \begin{pmatrix} U & V \end{pmatrix} = \begin{pmatrix} U^T U & U^T V \\ V^T U & V^T V \end{pmatrix}.$$

We can see the role of r in $\mathrm{SDP}(A)$ is $\mathrm{rank}(X) \leq r$. But X is $n+m$ by $n+m$, so as long as $r \geq n+m$, this constrain disappears.

12.2 Grothendieck Inequality

Theorem 12.1 (Grothendieck Inequality)**.**

$$\|A\|_{\infty\to 1} \leq \mathrm{SDP}(A) \leq k\|A\|_{\infty\to 1}.$$

Here the absolute constant k can be chosen as $k = \frac{1}{\frac{4}{\pi}-1} \approx 3.66$ (with the world record ≈ 1.78).

Proof: following Rietz [Rie74]. The left side is obvious and stated in Remark 12.3. We focus on the right side. The main idea is randomized rounding. Let $u_i, v_j \in \mathbb{S}^{d-1}$ achieve the maximum in $\mathrm{SDP}(A)$ (12.2) and $d = n+m$. We hope such u_i, v_j can match (not too far away from in objective) the x_i, y_j in (12.1). If we take some random x_i, y_j, then we can have the lower bound

$$\|A\|_{\infty\to 1} \geq \mathbb{E} \sum_{i,j} a_{ij} x_i y_j = \sum_{i,j} a_{ij} \mathbb{E}(x_i y_j).$$

But a_{ij} can be positive or negative, so we cannot go further directly.

Consider $x_i = \mathrm{sgn}(\langle g, u_i \rangle)$, and $y_j = \mathrm{sgn}(\langle g, v_j \rangle)$, where $g \sim N(0, I_d)$.

Fact 12.1. *Note that $\frac{g}{\|g\|} \sim$ unif $(\mathbb{S}^{d-1})$, so*

$$\mathbb{E} x_i y_j = \frac{2}{\pi} \arcsin\langle u_i, v_j \rangle.$$

Denote by $g_{u_i} = \langle g, u_i\rangle$ and $g_{v_j} = \langle g, v_j\rangle$. We consider the generic setting $x_i = f(g_{u_i}), y_j = f(g_{v_j})$ for some $f : \mathbb{R} \to [-1, 1]$. We have the following facts.

Fact 12.2. *1.* $\mathbb{E} g_u g_v = \langle u, v\rangle$.

2. $\mathbb{E} g_u f(g_v) = \langle u, v\rangle \mathbb{E} Z f(Z) \triangleq \langle u, v\rangle K,\ Z \sim N(0, 1)$.

3. $\mathbb{E}(g_u - f(g_u))^2 = 1 - 2K + L,\ L \triangleq \mathbb{E} f^2(Z)$.

The facts hold noticing each $g_u \sim N(0, 1)$. Then

$$\begin{aligned}
\|A\|_{\infty\to 1} \geq & \sum a_{ij}\mathbb{E} x_i y_j \\
= & \sum a_{ij}\mathbb{E} f(g_{u_i}) f(g_{v_j}) \\
= & \sum a_{ij}\mathbb{E}(g_{u_i} - f(g_{u_i}))(g_{v_j} - f(g_{v_j})) - \sum a_{ij}\mathbb{E} g_{u_i} g_{v_j} \\
& + \sum a_{ij}\mathbb{E}\left(g_{v_j} f(g_{u_i}) + g_{u_i} f(g_{v_j})\right) \\
(\text{def of } u_i, v_j) = & \underbrace{\sum a_{ij}\mathbb{E}(g_{u_i} - f(g_{u_i}))(g_{v_j} - f(g_{v_j}))}_{\star} + (2K - 1)\mathrm{SDP}(A).
\end{aligned}$$

The magical next step is observing $(\star)$ is a feasible representation (after normalizing) in (12.2). More formally, consider the Hilbert space $\mathcal{H}$ consisting of $L^2(\mu)$ square-integrable functions with standard Gaussian measure μ on $\mathbb{R}^d$ and $\langle h, \widetilde{h}\rangle_{\mathcal{H}} \triangleq \mathbb{E}\left[h(g)\widetilde{h}(g)\right]$. Then

$$\begin{aligned}
\mathrm{SDP}(A) = & \max_{h_i, \widetilde{h}_j \in \mathcal{H}: \|h_i\|_{\mathcal{H}} = \|\widetilde{h}_j\|_{\mathcal{H}} = 1} \sum_{i,j} a_{ij}\langle h_i, \widetilde{h}_j\rangle_{\mathcal{H}} \\
\geq & \sum_{i,j} a_{ij} \frac{-\langle \langle u_i, \cdot\rangle - f(\langle u_i, \cdot\rangle), \langle v_j, \cdot\rangle - f(\langle v_j, \cdot\rangle)\rangle_{\mathcal{H}}}{\|\langle u_i, \cdot\rangle - f(\langle u_i, \cdot\rangle)\|_{\mathcal{H}} \| \langle v_j, \cdot\rangle - f(\langle v_j, \cdot\rangle)\|_{\mathcal{H}}} \\
= & \sum_{i,j} a_{ij} \frac{-\mathbb{E}\left[(g_{u_i} - f(g_{u_i}))\left(g_{v_j} - f(g_{v_j})\right)\right]}{\sqrt{\mathbb{E}\left[(g_{u_i} - f(g_{u_i}))^2\right]}\sqrt{\mathbb{E}\left[\left(g_{v_j} - f(g_{v_j})\right)^2\right]}} \\
= & -\frac{1}{1 - 2K + L}\sum_{i,j} a_{ij}\mathbb{E}\left[(g_{u_i} - f(g_{u_i}))\left(g_{v_j} - f(g_{v_j})\right)\right].
\end{aligned}$$

Thus

$$\begin{gathered}
(\star) \geq -(1 - 2K + L)\mathrm{SDP}(A). \\
\Rightarrow \|A\|_{\infty\to 1} \geq (4K - L - 2)\mathrm{SDP}(A).
\end{gathered}$$

Let $f = \mathrm{sgn}$. Then $L = 1$, $K = \mathbb{E}|Z| = \sqrt{\frac{2}{\pi}}$, and hence $4K - L - 2 = 4\sqrt{\frac{2}{\pi}} - 3 > 0.19$. So we proved that we can choose $k = 5.27 > 1/0.19$ in the theorem.

Moreover, there is a natural way to improve the constant. If we replace f by αf in the derivation above, then

$$\begin{aligned}
\alpha^2\|A\|_{\infty\to 1} \geq & \sum a_{ij}\mathbb{E}\alpha f(g_{u_i}) \cdot \alpha f(g_{v_j}) \\
= & \sum a_{ij}\mathbb{E}(g_{u_i} - \alpha f(g_{u_i}))(g_{v_j} - \alpha f(g_{v_j})) - \sum a_{ij}\mathbb{E} g_{u_i} g_{v_j} \\
& + \sum a_{ij}\mathbb{E}\left(\alpha g_{v_j} f(g_{u_i}) + \alpha g_{u_i} f(g_{v_j})\right) \\
= & (\star)(\alpha) + (2\alpha K - 1)\mathrm{SDP}(A).
\end{aligned}$$

And

$$(\star)(\alpha) \geq -(1 - 2\alpha K + \alpha^2 L)\mathrm{SDP}(A).$$

Then the statement would be

$$\|A\|_{\infty\to 1} \geq \left(\frac{4K}{\alpha} - \frac{2}{\alpha^2} - L\right)\mathrm{SDP}(A).$$

The optimal α is $\frac{1}{K}$, and the bound is

$$\|A\|_{\infty\to 1} \geq \left(2K^2 - L\right)\mathrm{SDP}(A)$$

In the case $f = \mathrm{sgn}$, $2K^2 - L = \frac{4}{\pi} - 1 > 0.273$, as suggested in the theorem. □

Remark 12.6 (Optimizing the choice of f)**.** Clearly, the best strategy is to let

$$f = \arg\max_{|f|\leq 1, \mathbb{E}Zf(Z)>0} 2K^2 - L.$$

It is shown in [Rie74] that the solution is given by the bounded linear function, that is,

$$f(x) = \begin{cases} 2Kx & |x| \leq \frac{1}{2K} \\ \mathrm{sgn}(x) & |x| > \frac{1}{2K} \end{cases}.$$

where K is chosen such that $K = \mathbb{E}\left[Zf(Z)\right]$. Plugging in the above choice of f and using the Gaussian integration by parts, we have $\mathbb{E}\left[Zf(Z)\right] = \mathbb{E}\left[f'(Z)\right] = 2K\mathbb{E}\left[\mathbb{1}\left\{|Z| \leq 1/2K\right\}\right]$. Thus the requirement $K = \mathbb{E}\left[Zf(Z)\right]$ is equivalent to $1 = 2\mathbb{P}\left[|Z| \leq 1/2K\right]$. Moreover, the optimal value $2K^2 - L = \mathbb{E}\left[(2KZ - f(Z))f(Z)\right] = \mathbb{E}\left[(2K|Z| - 1)\mathbb{1}\left\{|Z| \leq 1/2K\right\}\right] = 2K\mathbb{E}\left[|Z|\mathbb{1}\left\{|Z| \leq 1/2K\right\}\right] - \mathbb{P}\left[|Z| \leq 1/2K\right] = 2K\sqrt{\frac{2}{\pi}}e^{-1/(8K^2)} - 1/2$, which is bigger than 0.4423, by checking the table of Gaussian integrals. Thus, by the above proof of Theorem 12.1, we get that the Grothendick's constant $k < 2.261$.

Remark 12.7 (PSD A)**.** When A is psd, the optimal bound is

$$\|A\|_{\infty\to 1} \geq \frac{2}{\pi}\mathrm{SDP}(A).$$

In this case, we can lower bound $(\star)$ by 0 instead, since by design $u_i = v_i$ and A is psd in this case. Then

$$\frac{\|A\|_{\infty\to 1}}{\mathrm{SDP}(A)} \geq \sup_{\alpha} \frac{2\alpha K - 1}{\alpha^2} = K^2 = \frac{2}{\pi}.$$

To show the sharpness, we construct examples that (asymptotically) achieve the bound. Let $l_i \overset{\text{i.i.d}}{\sim} \mathrm{unif}(\mathbb{S}^{d-1})$, $A_{ij} = \frac{1}{n^2}\langle l_i, l_j\rangle$. Choose $u_i = v_i = l_i$ in (12.2), we have

$$\mathrm{SDP}(A) \geq \frac{1}{n^2}\sum_{i,j}\langle l_i, l_j\rangle^2 \overset{\text{LLN}}{\approx} \mathbb{E}\langle l, l'\rangle^2 = \frac{1}{d}(1 + o(1)).$$

But

$$\|A\|_{\infty\to 1} \overset{\text{w.h.p}}{\leq} \frac{1 + o(1)}{d}\sqrt{\frac{2}{\pi}}.$$

This is because

$$\|A\|_{\infty\to 1} = \max_{x\in\{\pm\}^n}\langle A, xx^\top\rangle = \max_{x\in\{\pm\}^n}\frac{1}{n^2}\sum_{i,j}x_i x_j\langle l_i, l_j\rangle \leq \frac{1}{n^2}\sum_{i,j}|\langle l_i, l_j\rangle| \overset{\text{LLN}}{\approx} \mathbb{E}|\langle l, l'\rangle| = \frac{1 + o(1)}{d}\sqrt{\frac{2}{\pi}}.$$

Remark 12.8 (Connection to max-cut)**.** Given a weighted graph with positive weight matrix W, similar to min-cut in (10.2), define:

$$\mathrm{maxcut}(W) \triangleq \max_{I \subset [n]} \sum_{i \in I, j \in I^c} W_{ij}.$$

Then on one hand,

$$\begin{aligned} 2\mathrm{maxcut}(W) &= \max_{\sigma \in \{\pm\}^n} \sum W_{ij}(1 - \sigma_i \sigma_j) \\ &= \max_{\sigma \in \{\pm\}^n} \langle W, J - \sigma\sigma^\top \rangle \\ &\leq \max_{X \succeq 0, X_{ii}=1} \langle W, J - X \rangle \triangleq GW(W). \end{aligned}$$

On the other hand, following similarly as the proof of Theorem 12.1, we have

$$\begin{aligned} 2\mathrm{maxcut}(W) &\geq \sum_{ij} W_{ij} \left(1 - \mathrm{sgn}\left(g_{u_i}\right) \mathrm{sgn}\left(g_{u_j}\right)\right) \\ &= \sum W_{ij} \left(1 - \frac{2}{\pi} \arccos\langle u_i, u_j \rangle\right) \\ &\geq 0.878 \sum W_{ij} \left(1 - \langle u_i, u_j \rangle\right) \\ &= 0.878 \times GW(W), \end{aligned}$$

where the last inequality holds because $W_{ij} \geq 0$ and $\frac{2}{\pi} \arccos \rho \geq 0.878(1 - \rho)$ for all $\rho \in [-1, 1]$.

12.3 Application to SBM

Consider $\mathrm{SBM}(n, p, q)$ with $p = \frac{a}{n}, q = \frac{b}{n}$, and bisection $\langle \sigma, \mathbf{1} \rangle = 0$. Define $d = \frac{a+b}{2}$, and $s = a - b$. Recall in the bisection case (see Remark 10.2), the MLE has the following SDP relaxation

$$\begin{aligned} \widehat{X} = \arg\max\ & \langle A, X \rangle. \\ & X \succeq 0 \\ & X_{ii} = 1 \\ & \langle X, J \rangle = 0 \end{aligned}$$

We claim that the necessary and sufficient condition is

$$\frac{(a-b)^2}{a+b} \to \infty.$$

Here $\frac{(a-b)^2}{a+b}$ can be interpreted as the signal-to-noise ratio (snr). In the more general P/Q model, the condition is $H^2(P, Q) \gg \frac{1}{n}$, which recovers the above when $P = \mathrm{Bern}(p)$ and $Q = \mathrm{Bern}(q)$.

Theorem 12.2 ([GV16])**.** *Let $\widehat{v}$ be the top eigenvector of $\widehat{X}$, and $\widehat{\sigma} = sgn(\widehat{v})$. Then*

$$\mathbb{E}l(\widehat{\sigma}, \sigma) \overset{(also\ w.h.p)}{\lesssim} \frac{1}{\sqrt{snr}}.$$

Note: The above misclassification rate is later sharpened to exponential (optimal) by [FC18]:

$$\mathbb{E}l(\widehat{\sigma}, \sigma) \leq \exp(-\Omega(snr)).$$

Proof. Unlike exact recovery, here we do not know too much about the behavior of the optimal solution $\widehat{X}$ and hence it is not easy to apply the dual certificate argument. Instead, we will follow the primal proof.

Define the population solution

$$\begin{aligned} X^* = \arg\max\ & \langle \mathbb{E}A, X \rangle \\ & X \succeq 0 \\ & X_{ii} = 1 \\ & \langle X, J \rangle = 0 \end{aligned}$$

We can calculate

$$\mathbb{E}A = \frac{p+q}{2}J + \frac{p-q}{2}\sigma\sigma^\top - pI,$$

and justify $\sigma\sigma^\top = X^*$.

At a high level, we expect that since A is "close" to $\mathbb{E}[A]$, it follows that the optimal solutions $\widehat{X}$ and X^* are also close. To capture this intuition, we decompose

$$\begin{aligned} \langle \mathbb{E}A, \widehat{X} \rangle &= \langle A, \widehat{X} \rangle - \langle A - \mathbb{E}A, \widehat{X} \rangle \\ &\geq \langle A, X^* \rangle - \langle A - \mathbb{E}A, \widehat{X} \rangle \\ &= \langle \mathbb{E}A, X^* \rangle + \underbrace{\langle A - \mathbb{E}A, X^* \rangle - \langle A - \mathbb{E}A, \widehat{X} \rangle}_{\triangleq -\delta}, \end{aligned}$$

where the first inequality follows from the optimality of $\widehat{X}$. If we can somehow say $\delta \leq 0$, in other words $\langle A - \mathbb{E}A, X^* \rangle - \langle A - \mathbb{E}A, \widehat{X} \rangle \geq 0$, then we can conclude $\langle \mathbb{E}A, \widehat{X} \rangle \geq \langle \mathbb{E}A, X^* \rangle$, and thus $\widehat{X} = X^*$. Though this is not possible in general, we can show δ is not too big to get the conclusion. Let $\widehat{v} = v_1(\widehat{X}), v = v_1(X^*) = \frac{\sigma}{\sqrt{n}}$, then by the Davis-Kahan sinθ theorem,

$$\min \|\widehat{v} \pm v\|_2 \lesssim \frac{\|\widehat{X} - X^*\|_{op}}{\lambda_1(X^*) - \lambda_2(\widehat{X})} \leq \frac{\|\widehat{X} - X^*\|_F}{n - \|\widehat{X} - X^*\|_F},$$

where the last inequality follows by Weyl's theorem so that $\lambda_2(\widehat{x}) \leq \lambda_2(X^*) + \|\widehat{X} - X^*\|_{op} \leq \|\widehat{X} - X^*\|_{op} \leq \|\widehat{X} - X^*\|_F$. Also note that for every $\sigma_i \neq \widehat{\sigma}_i$, $\|\widehat{v} \pm v\|_2^2$ differs at least $\frac{1}{n}$ at this i. Thus

$$l(\widehat{\sigma}, \sigma) \leq \frac{1}{n} \cdot n \min \|\widehat{v} \pm v\|_2^2 \lesssim \frac{\|\widehat{X} - X^*\|_F^2}{n^2}.$$

Suppose $n\sqrt{d} \overset{\text{w.h.p}}{\gtrsim} \delta \geq \langle \mathbb{E}A, X^* \rangle - \langle \mathbb{E}A, \widehat{X} \rangle = \frac{p-q}{2}(n^2 - \langle \sigma\sigma^\top, \widehat{X} \rangle)$. Then

$$\begin{aligned} \|\widehat{X} - X^*\|_F^2 &= \|\widehat{X}\|_F^2 + \|X^*\|_F^2 - 2\langle \widehat{X}, X^* \rangle \\ &= \|\widehat{X}\|_F^2 + n^2 - 2\langle \widehat{X}, \sigma\sigma^\top \rangle \\ &\leq \mathrm{Tr}(\widehat{X})^2 + n^2 - 2\langle \widehat{X}, \sigma\sigma^\top \rangle \\ &= 2(n^2 - \langle \sigma\sigma^\top, \widehat{X} \rangle) \lesssim \frac{\delta}{p-q} = \frac{n\delta}{a-b} \leq \frac{n^2}{\sqrt{snr}}. \end{aligned}$$

This completes the proof. So it remains only to show $\delta \overset{\text{w.h.p}}{\lesssim} n\sqrt{d}$. Denote $W = A - \mathbb{E}A$. We want to show

$$\frac{1}{2}|\delta| \leq \mathrm{SDP}(W) = \max_{X \succeq 0, X_{ii}=1} \langle W, X \rangle \overset{\text{w.h.p}}{\lesssim} n\sqrt{d}.$$

By Grothendieck Inequality,

$$\begin{aligned}
\mathrm{SDP}(W) &= \max_{\|u_i\|=1} \sum_{i,j} W_{ij} \langle u_i, u_j \rangle \\
&\leq \max_{\|u_i\|=\|v_j\|=1} \sum_{i,j} W_{ij} \langle u_i, v_j \rangle \\
&\overset{G.I.}{\lesssim} \|W\|_{\infty \to 1} \\
&= \max_{x,y \in \{\pm\}^n} \langle W, xy^\top \rangle.
\end{aligned}$$

By Hoeffding's inequality Lemma 2.2,

$$\mathbb{P}(|\langle W, xy^\top \rangle| \geq t) \leq \exp(-c \cdot \frac{t^2}{n^2}).$$

To apply union bound on x, y, which in total 4^n, we need to choose $t \sim n^{3/2}$. We apply Bernstein's inequality instead,

$$\mathbb{P}(|\langle W, xy^\top \rangle| \geq t) \leq \exp\left(-\frac{t^2}{2\mathrm{Var}(\langle W, xy^\top \rangle) + 2t/3}\right) \leq \exp\left(-c \cdot \frac{t^2}{nd + t}\right),$$

then we can choose $t \sim n\sqrt{d}$. □

Remark 12.9. Note that the constraint $X_{ii} = 1$ is crucial. If such constraint were replaced by $\mathsf{Tr}(X) = n$ instead, then we would have $\max_{X \succeq 0, \mathsf{Tr}(X)=n} \langle W, X \rangle = n\|W\|_{op}$. In the sparse graphs where $d = o(\log n)$, we have $\|W\|_{op} \asymp \sqrt{d_{\max}} \asymp n\sqrt{\log n / \log(\log(n)/d)} \gg n\sqrt{d} \asymp \|W\|_{\infty \to 1}$. Loosely speaking, the constraint $X_{ii} = 1$ supresses the spiky eigenvectors of W located on high-degree vertices.

Initiated by [BR13], studying the computational limit of statistical problems is a broad topic. Results of varying precision have been obtained for different problems. As a case study, let us consider the problem of submatrix detection (biclustering):

Observe

$$X = M + Z$$

where X is a $N \times N$ matrix, M is the signal amd Z is the noise. Z is i.i.d $N(0,1)$ and $M = \mu \mathbb{1}_S \mathbb{1}_S^T$ where $S \ \subset [N], |S| = K$, and S is chosen uniformly at random. The problem is parametrized by (N, K, μ). We consider the asymptotic regime where $N \to \infty$ and it is convenient to look at the exponents so let $K = N^{\alpha}$ where $\alpha \in [0,1]$ and $\mu = N^{-\beta}$ where $\beta \in \mathbb{R}$. We consider the problem of detection, i.e., testing between

$$\begin{cases} H_0 : & X = Z \quad (\mu = 0) \\ H_1 : & X = M + Z \end{cases}.$$

The main result can be represented in an "easy-hard-impossible" phase transition diagram shown below.

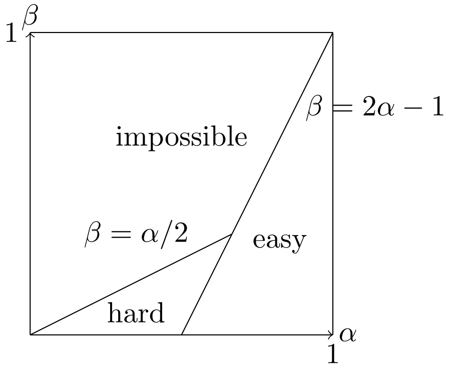


Figure 13.1: Difficulty of the problem in different regimes

In "impossible" regime, i.e., when $\beta > \max\{\alpha/2, 2\alpha - 1\}$. we will show that for any test $\phi : \mathbb{R}^{N \times N} \to \{0,1\}$,

$$\mathbb{P}_0(\phi = 1) + \sup_{S \in \binom{[N]}{k}} \mathbb{P}_S(\phi = 0) \to 1$$

by showing that $\mathrm{TV}(\mathbb{P}_0, \frac{1}{\binom{N}{k}} \sum_{S \in \binom{[N]}{k}} \mathbb{P}_S) \to 0$.

In "easy" regime, i.e., when $\beta < \max\{2\alpha - 1, 0\}$, there exist efficient tests that runs in $O(N^2)$ times. More specifically, we consider the following two test statistics

1. Linear test statistic

$$T_{SUM} = \sum_{i,j} X_{i,j} \sim \begin{cases} N(0, N^2) & \text{under } H_0 \\ N(\mu k K^2, N^2) & \text{under } H_1 \end{cases}.$$

   This test suceef if $\mu K^2 \gg N$, i.e., if $\beta < 2\alpha - 1$.

2. Max test statistic

$$T_{MAX} = \max_{i,j} X_{i,j} \overset{\mu \gg \sqrt{\log N}}{\approx} \begin{cases} \Theta(\sqrt{\log N}) & \text{under } H_0 \\ \mu & \text{under } H_1 \end{cases}.$$

   This test suceef if $\mu \gg \sqrt{\log N}$, i.e., if $\beta < 0$.

Now we return to the proof of the impossibility regime. Let $P_0 = \mathcal{L}(Z)$ denote the law under the null and $P_1 = \mathcal{L}(X)$ denote the law under the alternative if M is draw uniformly at random. We will prove that $\text{TV}(P_0, P_1) \to 0$ by proving that $\chi^2(P_1 || P_0) \to 0$. After some algebra we get

$$\chi^2(P_1||P_0) = \mathbb{E} \exp \left(\langle M, \widetilde{M} \rangle \right) - 1$$

where $M = \mu \mathbb{1}_S \mathbb{1}_S^T$ and $\widetilde{M} = \mu \mathbb{1}_{\widetilde{S}} \mathbb{1}_{\widetilde{S}}^T$ are i.i.d copies. Observe that

$$\langle M, \widetilde{M} \rangle = \mu^2 |S \cap \widetilde{S}|^2 = \mu^2 H^2$$

where $H \triangleq |S \cap \widetilde{S}|$ follows $\text{Hyp}(N, K, K)$ from previous lectures . Putting things together, we get

$$\chi^2(P_1||P_0) = \mathbb{E} \exp \left(\mu^2 H^2 \right) - 1$$

To characterize the behavior of $\mathbb{E} \exp \left(\mu^2 H^2 \right)$, we make use of the following lemma.

Lemma 13.1. *$\mathbb{E} \exp \left(\lambda H^2 \right) \leq C_1$ if $\lambda \leq C_2 \left(\frac{1}{K} \log \frac{eN}{K} \wedge \frac{N^2}{K^4} \right)$.Moreover, $C_1 \to 1$ if $C_2 \to 0$.*

Therefore, Lemma 13.1 is suggesting $\mu^2 \ll \frac{1}{K} \log \frac{eN}{K} \wedge \frac{N^2}{K^4}$, which is equivalent to $\beta > \alpha/2$ or $\beta > 2\alpha - 1$.

Sveral remarks on this Lemma is in order.

1. By Jensen inequality,

$$\mathbb{E} \exp \left(\lambda H^2 \right) \geq \exp(\lambda \frac{N^2}{K^4}).$$

   Therefore, $\lambda = O(\frac{N^2}{K^4})$ is necessary.

2. $$\mathbb{E} \exp \left(\lambda H^2 \right) \geq \exp(\lambda K^2) \mathbb{P}(H = K)$$

   Note that $\mathbb{P}(H = K) = \frac{1}{\binom{N}{K}} \geq (\frac{N}{K})^{-K}$. Therefore, $\lambda = O(\frac{1}{K} \log \frac{eN}{K})$ is also necessary.

3. Sharp constant is known.

In "hard" regime, i.e., $2\alpha - 1 < \beta < \alpha/2$, we first consider the scan test which is not computationally efficient, but works.

$$T_{SCAN} = \max_{|T|=K}, \sum_{i \in T, j \in T} X_{ij}.$$

Under H_1, $T_{SCAN} \geq \sum_{i \in S, j \in S} X_{i,j} \sim N(\mu K^2, K^2)$. Under H_0, $T_{SCAN} \lessapprox \sqrt{K^2 \log \binom{N}{K}} = K^{3/2}\sqrt{\log \frac{eN}{K}}$. Therefore, the test works if $\mu K^2 \gg K^{3/2} \Leftrightarrow \beta < \alpha/2$.

Now we discuss why in the regime $2\alpha - 1 < \beta < \alpha/2$ the problem is computatinally hard. But first we have to discuss wht we mean by polynomial-time algorithm. Conventionally, polynomial-time algorithm means that

$$\text{running time} = O\left(\text{poly}(\#\text{ of bits to describe the input})\right).$$

However, gaussian random variables are theoretical objects that are continuous and takes infinite number of values so takes infinite number of bits to describe precisely. So how do we make sense of the "Gaussian" noise?
Idea: find a discrete model that is *asymptotically equivalent* to Gaussian. To define *asymptotically equivalent* we first introduce the notion of LeCam deficiency of $\mathbb{P}$ w.r.t $\mathbb{Q}$. Let $\mathbb{P} = \{P_\theta : \theta \in \Theta\}$ on X and $\mathbb{Q} = \{Q_\theta : \theta \in \Theta\}$ on Y. Let $T : X \to Y$ be a Markove kernel (conditional distribution) so that $TP(dy) = \int T(dy|x)P(dx)$. Define

$$\delta(\mathbb{P}, \mathbb{Q}) \triangleq \inf_T \sup_{\theta \in \Theta} \text{TV}(TP_\theta, Q_\theta).$$

Define the LeCam distance

$$\Delta(\mathbb{P}, \mathbb{Q}) \triangleq \max\{\delta(\mathbb{P}, \mathbb{Q}), \delta(\mathbb{Q}, \mathbb{P})\}.$$

Then *asymptotically equivalent* simply means $\Delta(\mathbb{P}, \mathbb{Q}) \to 0$. Now let $X = M + Z$ be $\mathbb{P}$. We discretize X into X_t, denoted $\mathbb{P}_t$, by letting $(X_t)_{ij} = \frac{[X_{ij}2^t]}{2^t}$.

Lemma 13.2. *$\Delta(\mathbb{P}^{(N)}, \mathbb{P}_t^{(N)}) \leq c' N^2 \exp(-ct) \to 0$ as long as $t \geq C \log N$.*

Proof sketch. $\delta(\mathbb{P}, \mathbb{P}_t) = 0$ by definition since we can let kernel T to be the discretization procedure. $\delta(\mathbb{P}_t, \mathbb{P}) \lessapprox N^2 \exp(-ct)$ because the total variation between $(X_t)_{ij} + Unif(0, 2^{-t})$ and $(X_t)_{ij}$ is roughly $\exp(-ct)$. □

Therefore, the above discussion implies that it makes sense to talk about polynomial in the dimension N instead of the number of bits to describe the inputs.

Reduction Informally, Problem A is at least as hard as Problem B if B can be reduced to A in polynomial time.

We will show in the following sequel that the problem in "hard" regime is at least as hard as the planted clique problem. Recall that the planted clique problem is a testing problem where $H_0 : G \sim G(n, \frac{1}{2})$ and $H_1 : G \sim G(n, \frac{1}{2}, k)$. We will turn a planted clique problem of size n into a submatrix detection problem of size $N = nl$ where $l \to \infty$. The reduction consists of three steps.

1. We find a kernel $T : 0, 1 \to \mathbb{R}$ such that $\text{Ber}(\frac{1}{2}) \mapsto N(0, \frac{1}{l^2})$ and $\text{Ber}(1) \mapsto N(\mu, \frac{1}{l^2})$. Let

$$\frac{1}{2}(P_0 + P_1) = N(0, \frac{1}{l^2}) = \mathbb{Q} \quad P_1 = N(\mu, \frac{1}{l^2}) = \mathbb{P}.$$

Then we get $P_0 = 2\mathbb{Q} - \mathbb{P}$. Indeed, we can choose P_0 and P_1 such that

$$\frac{1}{2}(P_0 + P_1) = N(0, \frac{1}{l^2}) \quad P_1 \approx N(\mu, \frac{1}{l^2}).$$

Here the approximation denotes that the total variation distance is small.

2. Randomize the block.

3. Reduction scheme. For any $s, t \in [n]$, generate a $l \times l$ block T by

$$T = \begin{cases} P_0 & \text{if } A_{st} = 0 \\ P_1 & \text{if } A_{st=1} \end{cases}.$$

This is possible because if $X_1, \dots, X_n$ i.i.d $N(\mu, 1)$, then $\overline{X}$ is sufficient statistic. In other words, we can simulate $X_1, \dots, X_n$ given $\overline{X}$.

Observe that under H_0, $X \overset{d}{=} Z$. Under H_1, $\mathrm{TV}(\mathcal{L}(X), \mathcal{L}(M + Z)) \to 0$. The approximation error comes from two places. a) $P_1 \neq N(\mu, \frac{1}{l^2})$ and b) diagonal $l \times l$ blocks are always $N(0, 1)$.

So the reduction scheme works for $N = nl, K = kl, l \ll \frac{1}{\mu}$. Pick $l = \frac{1}{\mu} \frac{1}{\log N}$, we see that $k \ll \sqrt{n} \Leftrightarrow \mu \ll \frac{N}{K^2} \Leftrightarrow \beta > 2\alpha - 1$.

§ A. Preliminary on binary hypothesis testing

Consider a binary hypothesis testing problem:

$$H_0 : X \sim Q \text{ vs. } H_1 : X \sim P.$$

The goal is to tell whether X is generated from Q or P based on observation of X. Given a test $\phi : X \to \{0, 1\}$, where $\phi(X) = 0$ means "deciding on H_0 is true" and $\phi(X) = 1$ means "deciding on H_1 is true", its Type-I error (a.k.a. false positive) is $Q(\{\phi(X) = 1\})$ and Type-II error (a.k.a. false negative) is $P(\{\phi(X) = 0\})$.

Lemma A.1. *The minimum of Type-I + Type-II errors is* $1 - \mathrm{TV}(P, Q)$*, that is*

$$\min_{\phi} \{Q(\{\phi(X) = 1\}) + P(\{\phi(X) = 0\})\} = 1 - \mathrm{TV}(P, Q),$$

where $\mathrm{TV}(P, Q) \triangleq \frac{1}{2}\mathbb{E}_Q [|P/Q - 1|] = \frac{1}{2}\int |P - Q| = \sup_E \{P(E) - Q(E)\}$ *stands for the total variation distance.*

Proof. Exercise. Hint. Let $E = \{\phi(X) = 1\}$ in the definition of TV. □

Remark A.1.

- $\mathrm{TV}(P, Q) = 1 \Leftrightarrow P \perp Q$, minimum Type I+II errors is 0;
- $\mathrm{TV}(P, Q) = 0 \Leftrightarrow P = Q$, minimum Type I+II errors is 1.

In many high-dimensional statistical inference problem, $\mathrm{TV}(P, Q)$ is difficult to compute. Instead, we resort to bound $\mathrm{TV}(P, Q)$ in terms of other distances that are easier to compute. One such distance is the χ^2-divergence, which is the variance of the likelihood ratio.

Definition A.1. The χ^2-divergence between P and Q is given by[1]

$$\chi^2(P\|Q) \triangleq \mathbb{E}_Q\left[\left(\frac{P}{Q} - 1\right)^2\right] = \int \frac{P^2}{Q} - 1.$$

In general, $\chi^2(P\|Q) \neq \chi^2(Q\|P)$. The usefulness of χ^2 in proving negative results for hypothesis testing lies in the following observation:

Lemma A.2. *For any* P, Q*,* $\mathrm{TV}(P, Q) \leq \frac{1}{2}\sqrt{\chi^2(P\|Q)}$*. Thus if* $\chi^2(P\|Q) = o(1)$*, then* $\mathrm{TV}(P, Q) = o(1)$*.*

Proof.

$$2\mathrm{TV}(P, Q) = \mathbb{E}_Q\left[\left|\frac{P}{Q} - 1\right|\right] \leq \sqrt{\mathbb{E}_Q\left[\left(\frac{P}{Q} - 1\right)^2\right]} = \sqrt{\chi^2(P\|Q)}. \tag{A.1}$$

□

[1] Hereafter we assume P is absolutely continuous to Q and let P/Q denote the Radon-Nikodym derivative of P relative to Q.

Remark A.2 (Contiguity)**.** Recall the notion of contiguity (of two sequences of probability measures (P_n) and (Q_n)). We say (P_n) is contiguous to (Q_n) if for any sequence of events E_n, $Q_n(E_n) \to 0 \implies P(E_n) \to 0$. Contiguity implies non-detection, because for any sequence of tests:

$$Q_n(\text{failure}) \to 0 \implies P_n(\text{success}) \to 0,$$

which further implies that the minimum sum of Type-I+II errors is at least $\Omega(1)$, i.e., $\mathrm{TV}(P_n, Q_n) \leq 1 - \Omega(1)$.

A sufficient condition of contiguity is bounded second moment of likelihood, i.e., $\chi^2(P_n \| Q_n) = O(1)$. Indeed, by Cauchy-Schwarz,

$$P_n(E_n) = \mathbb{E}_{Q_n}\left[\frac{P_n}{Q_n}\mathbb{1}\left\{E_n\right\}\right] \leq \underbrace{\sqrt{\mathbb{E}_{Q_n}\left[\left(\frac{P_n}{Q_n}\right)^2\right]}}_{\sqrt{\chi^2+1}} Q_n(E_n) \to 0.$$

In many planted problems, P_n is a mixture distribution and Q_n is a simple distribution. The following lemma is very useful for computing χ^2(mixture distribution||simple distribution). The introduction of two iid copies of randomness is typical in second moment calculation (cf. Section 1.2.2).

Lemma A.3 (Second moment trick)**.** *Suppose we have a parametric family of distributions $\{P_\theta : \theta \in \Theta\}$. Given a prior on the parameter space Θ, define the mixture distribution:*

$$P_\pi \triangleq \int P_\theta \pi(d\theta).$$

Then we have $\chi^2(P_\pi || Q) = \mathbb{E}G(\theta, \widetilde{\theta}) - 1$, where $\theta, \widetilde{\theta} \overset{iid}{\sim} \pi$ and $G(\theta, \widetilde{\theta})$ is defined by

$$G(\theta, \widetilde{\theta}) \triangleq \int \frac{P_\theta P_{\widetilde{\theta}}}{Q}.$$

.

Proof. The proof is just by Fubini:

$$\begin{aligned}\int \frac{P_\pi^2}{Q} &= \int \frac{(\int P_\theta(x)\pi(d\theta))(\int P_{\widetilde{\theta}}(x)\pi(d\widetilde{\theta}))}{Q(x)}\mu(dx) \\ &= \int \pi(d\theta)\pi(d\widetilde{\theta}) \underbrace{\left(\int \frac{P_\theta(x)P_{\widetilde{\theta}}(x)}{Q(x)}\mu(dx)\right)}_{G(\theta,\widetilde{\theta})}.\end{aligned}$$

□

Example A.1 (Gaussian)**.** Consider $P_\theta = N(\theta, I_d)$ and $Q = N(0, I_d)$, and let π be some distribution on $\mathbb{R}^d$. Then $\chi^2(P_\pi \| Q) = \mathbb{E}[e^{\langle\theta,\widetilde{\theta}\rangle}] - 1$, where $\theta, \widetilde{\theta} \overset{\text{i.i.d.}}{\sim} \pi$. To see this, by Lemma A.3, it suffices to compute $G(\theta, \widetilde{\theta})$:

$$G(\theta, \widetilde{\theta}) = \mathbb{E}_{X\sim Q}\left[\frac{P_\theta(X)P_{\widetilde{\theta}}(X)}{Q(X)^2}\right] = \mathbb{E}_Q\left[e^{-\frac{\|X-\theta\|_2^2}{2} - \frac{\|X-\widetilde{\theta}\|_2^2}{2} + \|X\|_2^2}\right] = e^{-\frac{\|\theta\|_2^2 + \|\widetilde{\theta}\|_2^2}{2}} \mathbb{E}_Q\left[e^{\langle X, \theta+\widetilde{\theta}\rangle}\right] = e^{\langle\theta,\widetilde{\theta}\rangle}.$$

Bibliography

[ABLS07] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics*, 9(04):585–603, 2007.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.

[AS15] Emmanuel Abbe and Colin Sandon. Detection in the stochastic block model with multiple clusters: proof of the achievability conjectures, acyclic BP, and the information-computation gap. arXiv 1512.09080, Dec 2015.

[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.

[BM08] Mark Braverman and Elchanan Mossel. Noisy sorting without resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 268–276. ACM, New York, 2008.

[BMNN16] Jess Banks, Cristopher Moore, Joe Neeman, and Praneeth Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, NY, June 23-26 2016*, pages 383–416, 2016.

[Bol01] Béla Bollobás. *Random graphs.* Cambridge university press, 2001.

[BR13] Q. Berthet and P. Rigollet. Complexity theoretic lower bounds for sparse principal component detection. *Journal of Machine Learning Research: Workshop and Conference Proceedings*, 30:1046–1066, 2013.

[CO10] A. Coja-Oghlan. Graph partitioning via adaptive spectral techniques. *Comb. Probab. Comput.*, 19(2):227–284, 2010.

[CX16] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research*, 17(1):882–938, 2016.

[DAM15] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the two-groups stochastic block model. *arXiv:1507.08685*, 2015.

[DG77] Persi Diaconis and Ronald L. Graham. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 262–268, 1977.

[DGGP11] Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres. Finding hidden cliques in linear time with high probability. In *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics*, ANALCO '11, pages 67–75, Philadelphia, PA, USA, 2011. Society for Industrial and Applied Mathematics.

[DKMZ11] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborova. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physics Review E*, 84:066106, 2011.

[Dur10] Rick Durrett. *Probability: theory and examples.* Cambridge university press, 4th edition, 2010.

[FC18] Yingjie Fei and Yudong Chen. Exponential error rates of sdp for block models: Beyond grothendieck's inequality. *IEEE Transactions on Information Theory*, 65(1):551–571, 2018.

[Fel70] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. Wiley, New York, third edition, 1970.

[FK00] U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures & Algorithms*, 16(2):195–208, 2000.

[FK01] Uriel Feige and Joe Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Sciences*, 63(4):639–671, 2001.

[GM75] G. R. Grimmett and C. J. H. McDiarmid. On colouring random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, 77(2):313–324, 1975.

[GV16] Olivier Guédon and Roman Vershynin. Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049, 2016.

[Hås99] Johan Håstad. Clique is hard to approximate withinn 1- ε. *Acta Mathematica*, 182(1):105–142, 1999.

[HWX15] B. Hajek, Y. Wu, and J. Xu. Recovering a hidden community beyond the spectral limit in $O(|E| \log^* |V|)$ time. *Advances in Applied Probability*, Oct 2015. arXiv:1510.02786.

[HWX16] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, May 2016. (arXiv 1412.6156 Nov. 2014).

[HWX17] B. Hajek, Y. Wu, and J. Xu. Information limits for recovering a hidden community. *IEEE Trans. on Information Theory*, 63(8):4729 – 4745, 2017.

[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.

[KS03] Michael Krivelevich and Benny Sudakov. The largest eigenvalue of sparse random graphs. *Combinatorics, Probability and Computing*, 12(1):61–72, 2003.

[KS12] V. Korolev and I. Shevtsova. An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scandinavian Actuarial Journal*, 2012(2):81–105, 2012.

[Kuč95] L. Kučera. Expected complexity of graph partitioning problems. *Discrete Appl. Math.*, 57(2-3):193–212, Feb. 1995.

[MM09] Marc Mezard and Andrea Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[MNS15] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2015.

[Moo17] Cristopher Moore. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467*, 2017.

[MWR18] Cheng Mao, Jonathan Weed, and Philippe Rigollet. Minimax rates and efficient algorithms for noisy sorting. *Proceedings of the 28th International Conference on Algorithmic Learning Theory*, 2018.

[Rie74] Ronald E Rietz. A proof of the grothendieck inequality. *Israel journal of mathematics*, 19(3):271–276, 1974.

[SKZ14] Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the Bethe Hessian. *Advances in Neural Information Processing Systems*, 27, 2014.

[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: Information-theoretical and computational limits. In Yonina Eldar and Miguel Rodrigues, editors, *Information-Theoretic Methods in Data Science.* Cambridge University Press, forthcoming, 2018.