# Lecture 8: Grothendieck Inequality and Approximate Community Detection in the SBM

*In which we show that an SDP relaxation can be used for approximate community detection in the stochastic block model. Along the way, we introduce matrix norms that show concentration of random matrices around their mean even in the case of very sparse matrices.*

## 1 The Algorithm

Our algorithm will be based on semi-definite programming. Intuitively, the problem of reconstructing the partition is essentially the same as the min-bisection problem, which is the problem to find a balanced cut crossed by the fewest number of edges. There are two intuitive explanations for this. One is that the balanced cut with the fewest *expected* edges is exactly our hidden cut, so we may hope that it will be so even in the empirical graph and not just in the expected graph. The other explanation for looking at the balanced cut with the minimum number of edges is that it is the *maximum likelihood* cut, that is, the cut that has the highest conditional probability of being the hidden cut given the graph that we see. This is because if we let $G = (V, E)$ be the random variables of the stochastic block model graph and $S$ be the (correlated) random variable such that $(S, V - S)$ is the hidden cut, then

$$\mathbb{P}[S|G] = \frac{\mathbb{P}[G|S] \cdot \mathbb{P}[S]}{\mathbb{P}[G]}$$

Where $\mathbb{P}[G]$ is a fixed quantity given the graph,

$$\mathbb{P}[S] = \frac{1}{\binom{n}{n/2}}$$

is the same for all cuts and

$$\mathbb{P}[G|S] = p^{|E|-|E(S,V-S)|} \cdot (1-p)^{2\binom{n/2}{2}-(|E|-|E(S,V-S)|)} \cdot q^{|E(S,V-S)|} \cdot (1-q)^{\frac{n^2}{4}|E(S,V-S)|}$$

where $E(S, V-S)$ is the set of edges of $G$ with one endpoint in $S$ and one in $V-S$. One can see that the above quantity is maximized (over the choice of $S$) when $E(S, V-S)$ is minimized.

Unfortunately, the min-bisection problem is **NP**-hard, so we will use semi-definite programming. The min-bisection problem can be stated as the following quadratic program with real-valued variables:

$$\begin{aligned} \text{minimize} \quad & \sum_{(u,v)\in E} \frac{1}{4}(x_u - x_v)^2 \\ \text{subject to} \quad & x_v^2 = 1, \forall v \in V \\ & \sum_{v\in V} x_v = 0. \end{aligned}$$

Which is equivalent (in the sense that the set of optimal solutions is the same) to the quadratic program

$$\begin{aligned} \text{maximize} \quad & \sum_{u,v\in V} A_{u,v} x_u x_v \\ \text{subject to} \quad & x_v^2 = 1, \forall v \in V \\ & \sum_{v\in V} x_v = 0. \end{aligned}$$

Its semi-definite programming relaxation is

$$\begin{aligned} \text{maximize} \quad & \sum_{u,v\in V} A_{u,v} \cdot \langle x_u, x_v \rangle \\ \text{subject to} \quad & \|\mathbf{x}_v\|^2 = 1, \forall v \in V \\ & \|\sum_{v\in V} \mathbf{x}_v\|^2 = 0. \end{aligned} \tag{1}$$

Our algorithm will be as follows.

- Solve the semi-definite programming above.
- Let $\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*$ be the optimal solution and $X = (X_{ij})$ such that $X_{ij} = \langle \mathbf{x}_i^*, \mathbf{x}_j^* \rangle$.
- Find $\mathbf{z} = (z_1, \ldots, z_n)$, which is the eigenvector corresponding to the largest eigenvalue of $X$.
- Let $S = \{i : z_i > 0\}$, $V - S = \{i : z_i \leq 0\}$.
- Output $(S, V - S)$ as our partition.

Defining, as usual, $a := pn/2$ and $b := qn/2$, we will prove that:

- For every $\epsilon$, there is a constant $\beta_\epsilon$ such that, if $a - b \geq \beta_\epsilon \sqrt{a+b}$, the algorithm recovers a solution that misclassifies at most an $\epsilon$ fraction of vertices.
- There is an absolute constant $\beta$ such that, if $a - b \geq \beta \sqrt{\log n}\sqrt{a+b}$, the algorithm recovers the exact solution.

Both statements are tight, up to the exact value of the constants $\beta_\epsilon$ and $\beta$. In this lecture we will prove the first statement and in the next lecture we will prove the second.

The proof of the first statement relies on an understanding of the concentration of the adjacency matrix of a SBM graphs under appropriate norms, so we will start by introducing some matrix norms and studying concentration.

# 2 The infinity-to-one matrix norm

In the past few lectures, we've been heavily relying on the spectral norm,

$$\|M\| = \max_{\|y\|=1} \|M\boldsymbol{y}\| = \max_{\|\boldsymbol{x}\|=1, \|\boldsymbol{y}\|=1} \boldsymbol{x}^T M \boldsymbol{y}$$

which is efficiently computable and turns out to be pretty handy in a lot of cases.

Unfortunately, high-degree vertices have a disproportionately large influence on the spectral norm's value, limiting its usefulness in very sparse graphs where such outliers exist.

In this lecture, we'll attack this problem by introducing a different norm, the *infinity-to-one* norm, defined as follows:

$$\|M\|_{\infty\to 1} := \max_{||\mathbf{y}||_\infty = 1} ||M\mathbf{y}||_1$$

The infinity-norm is dual to the 1-norm, in the sense that, for every vector $\mathbf{z}$ we have

$$||\mathbf{z}||_1 = \max_{||\mathbf{x}||_\infty = 1} \mathbf{x}^T \mathbf{z}$$

and we can use this fact to reformulate the infinity-to-one norm as

$$||M||_{\infty\to 1} = \max_{||\mathbf{x}||_\infty = 1, ||\mathbf{y}||_\infty = 1} \mathbf{x}^T M \mathbf{y}$$

From the above characterization, and from the fact that a vector of infinity-norm 1 has 2-norm at most $\sqrt{n}$ we get

$$\|M\|_{\infty\to 1} \le n\|M\| \ .$$

So spectral norm always gives us a bound for the infinity-to-one norm, and concentration bounds for the spectral norm imply concentration bounds for the infinity-to-one norm. For the latter norm, however, we can directly prove concentration bounds that are stronger, particularly in the case of sparse graphs.

It is not difficult to prove the characterization

$$\|M\|_{\infty\to 1} = \max_{\boldsymbol{x},\boldsymbol{y}\in\{-1,1\}^n} \boldsymbol{x}^T M \boldsymbol{y} \ .$$

and we will do so in Lemma 3 below. This is very useful, because it reduces the calculation of the norm to the examination of a finite number cases, enabling probabilistic bounds that rely on Chernoff bounds and union bounds.

**Theorem 1** *Let $P$ be a symmetric matrix with entries in $[0,1]$. Pick a random graph $G$ such that $\{i,j\}$ is in $G$ w.p. $P_{ij}$. Then whp:*

$$\|A-P\|_{\infty\to 1} \le O\left(\sqrt{n\sum_{ij} P_{ij}}\right) = O(n\sqrt{d}).$$

Where $d = \frac{1}{n}\sum_{i,j} P_{i,j}$ is the expected average degree of the random graph.

PROOF: Fix $\boldsymbol{x},\boldsymbol{y} \in \{+1,-1\}^n$. Examine the expression

$$\boldsymbol{P}\left[\boldsymbol{x}^T(A-P)\boldsymbol{y} \ge t\right] = \boldsymbol{P}\left[\sum_{ij} x_y y_j (A_{ij}-P_{ij}) \ge t\right]$$

We want to make sure that this probability exponentially decreases w.r.t. $n$.

Recall that Bernstein's inequality states that given $N$ independent random variables, absolutely bounded by $M$, with expectation 0, we have $P[\sum_i X_i > t] \le \exp\left(-\frac{t^2/2}{\sum_i EX_i^2+3Mt}\right)$.

$A_{ij}-P_{ij}$ is either $-P_{ij}$ or $1-P_{ij}$ and therefore is bounded by $[-1,+1]$. Since $\boldsymbol{x},\boldsymbol{y} \in \{+1,-1\}^n$, we can have the $M$ from Bernstein's inequality take on value 1. Combining this with the fact that

$$\mathbb{E}(x_i y_i (A_{ij}-P_{ij}))^2 \le P_{ij},$$

Bernstein's inequality gives us

$$\boldsymbol{P}\left[\sum_{ij} x_y y_j (A_{ij}-P_{ij}) \ge t\right] \le \exp\left(-\frac{t^2/2}{\sum_i p_{ij}+3t}\right) \le 2^{-3n}$$

provided that $t > \sqrt{n\sum_{i,j} p_{ij}}$, as desired. $\Box$

So this norm allows us to easily sidestep the issue of high-degree vertices. The problem is that it's NP-hard to compute.

# 3 Grothendieck's Inequality

However, it turns out that we can *approximate* the infinity-to-one norm to within a constant factor using SDP:

**Theorem 2** *(Grothendieck's Inequality) There exists some $c$ (turns out to be around 1.7, but we won't worry about its exact value here) such that for all $M$,*

$$\max_{\substack{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \\ \boldsymbol{y}_1,\ldots,\boldsymbol{y}_n \\ \|\boldsymbol{x}_i\|=1 \\ \|\boldsymbol{y}_y\|=1}} \sum_{i,j} M_{ij}\left\langle \boldsymbol{x}_i, \boldsymbol{y}_j \right\rangle \leq c \max_{\boldsymbol{x},\boldsymbol{y}\in\{-1,1\}^n} \sum_{i,j} M_{ij}x_iy_j = \|M\|_{\infty\to 1}$$

So instead of dealing the combinatorially huge problem of optimizing over all $\pm 1$ vectors, we can just solve an SDP instead to get a good approximation. For convenience, let's denote the quantity on the left of the above expression as $\|M\|_{SDP}$.

Let's start with a warmup lemma. The proof techniques in this lemma - specifically, the trick of replacing a vector from a continuous distribution with a *random* vector from some discrete distribution, and then taking the expectation to relate the two quantities, will come in handy later on.

**Lemma 3** $\max_{\boldsymbol{x},\boldsymbol{y}\in\{-1,1\}^n} \boldsymbol{x}^T M\boldsymbol{y} = \max_{\boldsymbol{x},\boldsymbol{y}\in[-1,1]^n} \boldsymbol{x}^T M\boldsymbol{y}$. *In other words, maximizing over the discrete space of $\pm 1$ random vectors and maximizing over the continuous space of vectors in that range gives the same result.*

PROOF: It is obvious that the expression on the left is at least the right (since it's just a relaxation). In order to show that the right is at least the left: given some continuous vectors $\boldsymbol{x}_i, \boldsymbol{y}_i$, we can find discrete $-1$,1 vectors $\boldsymbol{b}_i, \boldsymbol{c}_i$ such that their expectations are equal to $\boldsymbol{x}_i, \boldsymbol{y}_i$. We can do that by having $\boldsymbol{b}_i$ take on value $+1$ w.p. $1/2 + x_i/2$ and $-1$ w.p. $1/2 - x_i/2$, likewise for $\boldsymbol{c}_i$.

That means that:

$$\begin{aligned}
\mathbb{E}\sum_{i,j} M_{i,j}\boldsymbol{b}_i\boldsymbol{c}_j &= \sum_{i,j} M_{i,j}(\mathbb{E}\boldsymbol{b}_i)(\mathbb{E}\boldsymbol{c}_j) \\
&= \sum_{i,j} M_{ij}\boldsymbol{x}_i\boldsymbol{y}_j
\end{aligned}$$

which gives us the desired result. $\square$

**Fact 4** $\|M\|_{SDP}$ *is a norm.*

PROOF: **Multiplicative scaling:** obvious.

**Nonnegativity (except iff $M = 0$):** It is obvious that the SDP norm is zero if $M$ is zero.

Now suppose $M$ is nonzero.

Notice that we can replace the constraints in the SDP norm requiring that $\|\boldsymbol{x}_i\| = 1$, $\|\boldsymbol{y}_i\| = 1$ with $\|\boldsymbol{x}_i\| \leq 1$, $\|\boldsymbol{y}_i\| \leq 1$. Why? We'll use the same trick as we did in the proof of Lemma 3:

Obviously maximizing over $\|\boldsymbol{x}_i\| \leq 1$, $\|\boldsymbol{y}_i\| \leq 1$ will give us at least as good a result as maximizing over $\|\boldsymbol{x}_i\| = 1$, $\|\boldsymbol{y}_i\| = 1$, since it's a relaxation, so it suffices to show that if we can obtain some value for $\sum_{i,j} M_{ij} \left\langle \boldsymbol{x}_i, \boldsymbol{y}_j \right\rangle$ using $\|\boldsymbol{x}_i\| \leq 1$, $\|\boldsymbol{y}_i\| \leq 1$, we can do at least as well using $\|\boldsymbol{x}_i\| = 1$, $\|\boldsymbol{y}_i\| = 1$.

Let's suppose we have some vectors $\boldsymbol{x}_i, \boldsymbol{y}_i$ with length at most 1. Now let's replace them with *random* vectors $\boldsymbol{r}_i, \boldsymbol{s}_i$ of length exactly 1 whose expectation are $\boldsymbol{x}_i, \boldsymbol{y}_i$ respectively (just scale the $\boldsymbol{x}_i, \boldsymbol{y}_i$ up, and have $\boldsymbol{r}_i, \boldsymbol{s}_i$ be either the scaled value or its negative with probability calibrated appropriately so their expectations work out to be $\boldsymbol{x}_i, \boldsymbol{y}_i$). Then we can just say:

$$\begin{aligned} \mathbb{E} \sum_{i,j} M_{ij} \left\langle \boldsymbol{r}_i, \boldsymbol{s}_j \right\rangle &= \sum_{i,j} M_{ij} \left\langle \mathbb{E}\boldsymbol{r}_i, \mathbb{E}\boldsymbol{s}_j \right\rangle \\ &= \sum_{i,j} M_{ij} \left\langle \boldsymbol{x}_i, \boldsymbol{y}_j \right\rangle . \end{aligned}$$

So $\boldsymbol{r}_i, \boldsymbol{s}_j$ must take on some values (of length exactly 1) that make $\sum_{i,j} M_{ij} \left\langle \boldsymbol{r}_i, \boldsymbol{s}_j \right\rangle \geq \sum_{i,j} M_{ij} \left\langle \boldsymbol{x}_i, \boldsymbol{y}_j \right\rangle$, as desired.

Since we assumed $M$ is nonzero, let $a$, $b$ be such that $M_{a,b} \neq 0$. If we set $\boldsymbol{x}_a = \pm \boldsymbol{y}_b$ to be some arbitrary vector of unit length - using a plus if $M_{a,b}$ is positive and a minus if it's negative - and all other $\boldsymbol{x}_i, \boldsymbol{y}_j$ to zero (which we can do without affecting the value of the max by the fact we just proved), we can immediately see that

$$\sum_{i,j} M_{ij} \left\langle \boldsymbol{x}_i, \boldsymbol{y}_j \right\rangle = M_{ab} \langle \boldsymbol{x}_a, \boldsymbol{y}_b \rangle > 0,$$

giving a positive lower bound for the maximizer, as desired.

**Triangle inequality:** Just look at the $x$ and $y$ that maximize $\sum_{i,j} M_{ij} \left\langle \boldsymbol{x}_i, \boldsymbol{y}_i \right\rangle$ for $M = A$ and $M = B$, and observe that

$$\max_{\substack{\boldsymbol{x}_1,...,\boldsymbol{x}_n \\ \boldsymbol{y}_1,...,\boldsymbol{y}_n \\ \|\boldsymbol{x}_i\|=1 \\ \|\boldsymbol{y}_y\|=1}} \sum_{i,j} (A+B)_{ij} \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle \le \max_{\substack{\boldsymbol{x}_1,...,\boldsymbol{x}_n \\ \boldsymbol{y}_1,...,\boldsymbol{y}_n \\ \|\boldsymbol{x}_i\|=1 \\ \|\boldsymbol{y}_y\|=1}} \sum_{i,j} A_{ij} \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle + \max_{\substack{\boldsymbol{x}_1,...,\boldsymbol{x}_n \\ \boldsymbol{y}_1,...,\boldsymbol{y}_n \\ \|\boldsymbol{x}_i\|=1 \\ \|\boldsymbol{y}_y\|=1}} \sum_{i,j} B_{ij} \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle$$

since we can always match the quantity on the left hand side by choosing the same $\boldsymbol{x_i}$ and $\boldsymbol{y_j}$ for both terms on the right-hand-side. □

## 3.1 Proof of Grothendieck's inequality

Now let's prove Grothendieck's inequality:

PROOF: Observe that, by Lemma 3, maximizing over the choice of $\boldsymbol{x}, \boldsymbol{y} \in \{-1, 1\}^n$ in the zero-to-one norm is equivalent to maximizing over the choice of $\boldsymbol{x}, \boldsymbol{y} \in [-1, 1]^n$, so we can rewrite our proof goal as

$$\max_{\substack{\boldsymbol{x}_1,...,\boldsymbol{x}_n \\ \boldsymbol{y}_1,...,\boldsymbol{y}_n \\ \|\boldsymbol{x}_i\|=1 \\ \|\boldsymbol{y}_y\|=1}} \sum_{i,j} M_{ij} \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle \le c \max_{\boldsymbol{x},\boldsymbol{y} \in [-1,1]^n} \sum_{i,j} M_{ij} x_i y_j$$

Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{y}_1, ..., \boldsymbol{y}_n$ be of length 1 and be the optimal choices used in $\|M\|_{SDP}$, i.e. the maximizers of $\sum_{i,j} M_{ij} \langle \boldsymbol{x}_i, \boldsymbol{y}_i \rangle$. It suffices to prove that there exist vectors $\boldsymbol{b}, \boldsymbol{c} \in [-B, B]^n$ (for some fixed constant $B$, since we can just scale the result by changing $c$) to plug into the right-hand side of the above expression such that $c \sum_{i,j} M_{ij} b_i c_j \ge \sum_{i,j} M_{ij} \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle$.

Pick $\boldsymbol{g} = (g_1, ..., g_m)$, with each coordinate being drawn from the normal distribution with mean 0 and variance 1, and let $b_i = \langle \boldsymbol{g}, \boldsymbol{x}_i \rangle$ and $c_i = \langle \boldsymbol{g}, \boldsymbol{y}_j \rangle$. Then

$$\begin{aligned} \mathbb{E} b_i c_j &= \mathbb{E}_{\boldsymbol{g}} \boldsymbol{x}_i^T \boldsymbol{g} \boldsymbol{g}^T \boldsymbol{y}_j \\ &= \boldsymbol{x}_i^T (\mathbb{E}_{\boldsymbol{g}} \boldsymbol{g} \boldsymbol{g}^T) \boldsymbol{y}_j \end{aligned}$$

But $\mathbb{E}_{\boldsymbol{g}} \boldsymbol{g} \boldsymbol{g}^T$ is just the identity, since the diagonal elements are just the expectation of the square of the Gaussian (which is 1, its variance) and the off-diagonals are the expectation of the product of two independent Gaussians, which is zero (since the expectation of each individual Gaussian is zero).

So

$$\mathbb{E} b_i c_j = \boldsymbol{x}_i^T \boldsymbol{y}_j = \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle \ .$$

At this point we've got something that looks a lot like what we want: if the expectation of $b_i c_j$ is equal to $\langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle$, then maximizing over them is definitely going to give us a quantity greater than or equal to $\langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle$. Unfortunately, there's an issue here: since Gaussian random variables are unbounded, $b_i$ and $c_i$ are unbounded; on the other hand, what we're trying to do is maximize over vectors whose elements are bounded by $[B, B]$.

Our approach will be to "clip" the $b_i$ and $c_i$ to a finite range, and then bound the error introduced by the clipping process. Formally, fix a constant $B$ and pick a Gaussian random vector $\boldsymbol{g}$. Now define a 'clipped' inner product:

$$b_i := \begin{cases} -B & \text{if } \langle \boldsymbol{x}_i, \boldsymbol{g} \rangle < -B \\ \langle \boldsymbol{x}_i, \boldsymbol{g} \rangle & \text{if } -B \leq \langle \boldsymbol{x}_i, \boldsymbol{g} \rangle \leq B \\ B & \text{if } \langle \boldsymbol{x}_i, \boldsymbol{g} \rangle > B \end{cases}$$

and likewise for $\boldsymbol{c}_j$. For convenience, let's define the *truncation error* as follows, to represent how far the clipped value differs from the actual value.

$$t(z) := \begin{cases} z + B & \text{if } z < -B \\ 0 & \text{if } -B \leq z \leq B \\ z - B & \text{if } z > B \end{cases}$$

So we can rewrite $b$ as:

$$b_i := \langle \boldsymbol{x}_i, \boldsymbol{g} \rangle - t(\langle \boldsymbol{x}_i, \boldsymbol{g} \rangle)$$

and similarly we can define:

$$c_j := \langle \boldsymbol{y}_j, \boldsymbol{g} \rangle - t(\langle \boldsymbol{y}_j, \boldsymbol{g} \rangle) \ .$$

So now we have

$$\begin{aligned} \mathbb{E} \sum_{ij} M_{ij} b_i c_j &= \sum_{ij} M_{ij} \mathbb{E} \left[ \left( \langle \boldsymbol{x}_i, \boldsymbol{g} \rangle - t(\langle \boldsymbol{x}_i, \boldsymbol{g} \rangle) \right) \left( \langle \boldsymbol{y}_j, \boldsymbol{g} \rangle - t(\langle \boldsymbol{y}_j, \boldsymbol{g} \rangle) \right) \right] \\ &= \sum_{ij} M_{ij} \mathbb{E} \left[ \langle \boldsymbol{g}, \boldsymbol{x}_i \rangle \langle \boldsymbol{g}, \boldsymbol{y}_j \rangle \right] \\ &\quad - \sum_{ij} M_{ij} \mathbb{E} \left[ \langle \boldsymbol{g}, \boldsymbol{x}_i \rangle t(\langle \boldsymbol{g}, \boldsymbol{y}_j \rangle) \right] \\ &\quad - \sum_{ij} M_{ij} \mathbb{E} \left[ t(\langle \boldsymbol{g}, \boldsymbol{x}_i \rangle) \langle \boldsymbol{g}, \boldsymbol{y}_j \rangle \right] \\ &\quad + \sum_{ij} M_{ij} \mathbb{E} \left[ t(\langle \boldsymbol{g}, x_i \rangle) t(\langle \boldsymbol{g}, y_j \rangle) \right] \end{aligned}$$

Clearly $\sum_{ij} M_{ij} \mathbb{E} \langle \boldsymbol{g}, \boldsymbol{x}_i \rangle \langle \boldsymbol{g}, \boldsymbol{y}_j \rangle$ is just the SDP norm , since$\langle \boldsymbol{g}, \boldsymbol{x}_i \rangle$ and $\langle \boldsymbol{g}, \boldsymbol{y}_i \rangle$ were the 'original' $b_i$ and $c_i$.

What we will show now is that the remaining three terms are bounded by constant factors of SDP norm, so the entire sum of all four terms getting a constant factor approximation of it. The analysis is the same for all three terms so, for brevity, we'll just look at the last one:

$$\sum_{ij} M_{ij}\mathbb{E}t(\langle \boldsymbol{g}, x_i\rangle)t(\langle \boldsymbol{g}, y_j\rangle) \ .$$

For convenience, let's define $f_i, h_j : \mathbb{R}^m \to \mathbb{R}$ as:

$$f_i(\boldsymbol{g}) := t(\langle \boldsymbol{x}_i, \boldsymbol{g}\rangle)$$

$$h_j(\boldsymbol{g}) = t(\langle \boldsymbol{y}_j, \boldsymbol{g}\rangle) \ .$$

Now, it's convenient to think of the $f_i$ and $h_i$ as vectors of infinite dimension indexed by input $g$ (we'll bring the dimensionality down to a finite value at the end of the proof). Let's define the following inner product and norm in this space:

$$\langle f_1, f_2\rangle = \mathbb{E}_g f_1(\boldsymbol{g}) f_2(\boldsymbol{g})$$

$$\|f\|^2 = \mathbb{E}_g f^2(\boldsymbol{g})$$

Now, since the Gaussian distribution is rotation-independent (we can just rotate a Gaussian random variable around without changing its distribution), the squared norms of the $f_i$ and the $h_j$ are all the same (since all the $\boldsymbol{x}_i$, $\boldsymbol{y}_i$ have length 1, and dotting with them can be thought of a rotation). That means that the above norm takes on the same value for all $f$, $h$, so all we need to do is figure out a constant bound on it.

Fortunately, this value is pretty easy to bound. Notice that the function $t$ is zero if the dot product's absolute value is smaller than $B$. If we have $w \sim N(0,1)$,

$$|t(w)|^2 \leq |w|^2 1_{\{w\geq B, w\leq -B\}} \leq e^{-\Omega(B^2)} \leq 1/(10B)$$

for sufficiently large $B$, which is a constant.

So if the squared norms are bounded above by $1/(10B)$, which means we can substitute this and the norm bound into the above expression to get:

Now, armed with this, we can conclude (with similar results for the other two error terms)

$$\sum_{ij} M_{ij}\mathbb{E}t(\langle g, \boldsymbol{x}_i\rangle)t(\langle g, \boldsymbol{y}_j\rangle) = \sum_{ij} M_{ij}\langle f_i, h_j\rangle \leq \|M\|_{SDP}(1/100B^2) \ .$$

which tells us that $\mathbb{E}\sum_{ij} M_{ij} b_i c_j$ , is within a constant bound of the SDP norm $\sum_{ij} M_{ij}\mathbb{E}\langle g, \boldsymbol{x}_i\rangle\langle g, \boldsymbol{y}_j\rangle$ as desired.

There's only one slightly fishy bit in the proof we used above, though, and that's the treatment of functions of infinite-dimensional vectors indexed by a Gaussian vector. Let's conclude by constructing a (finite-dimensional) solution to the SDP from the functions:

**Claim:** If $X_{ij} = \langle f_i, f_j \rangle$, then $X_{ij} \succeq 0$.

**Proof:**

$$\begin{aligned} a^T X a &= \sum_{iji} a_i \langle f_i, f_j \rangle a_j \\ &= \sum_{ij} \langle a_i f_i, a_j f_j \rangle \\ &= \langle \sum_i a_i f_i, \sum_j a_j f_j \rangle \\ &= \| \sum_i a_i f_i \|^2 \\ &\geq 0 \end{aligned}$$

as desired. □

So the matrix $X_{ij}$ comprises a nice finite-dimensional solution to the SDP, and we're done with the proof.

Also, noticing that we have four terms in the expansion of $\mathbb{E} \sum_{ij} M_{ij} b_i c_j$, each one of which is a feasible value for the SDP, we can figure out how much we deviate from the actual optimum, i.e. $\|M\|_{SDP}$. Since each term can't exceed the value of the SDP optimum at all, $\mathbb{E} \sum_{ij} M_{ij} b_i c_j$ is separated from $\sum_{ij} M_{ij} \mathbb{E}\left[ \langle \boldsymbol{g}, \boldsymbol{x}_i \rangle \langle \boldsymbol{g}, \boldsymbol{y}_j \rangle \right]$ by a factor of four, giving us a bound on the constant $c$ in the inequality. □

To recap, notice that there were two key "tricks" here:

1) Assuming that we were rounding an optimal solution to our SDP (i.e. starting with $\boldsymbol{x}_i, \boldsymbol{y}_j$ as optimizers). We don't get any bounds otherwise!

2) Treating the rounding error itself as a feasible solution of the SDP.

This proof was communicated to us by James Lee.

# 4 Analysis of the Algorithm

Consider the feasible solution $\chi = (\chi_v)_{v \in V}$ for SDP1 defined in the first section, where

$$\chi_v = \begin{cases} +1 & \text{if } v \in V_1, \\ -1 & \text{if } v \in V_2. \end{cases}$$

The expected cost of such solution over the randomness of the graph will be

$$\mathop{\mathbb{E}}_{\text{choice of graph}}(\text{cost of } \chi) = \frac{n(n-1)}{2}p - \frac{n^2}{2}q = n(a-b) - a.$$

Since each edge is chosen independently, with high probability our cost will be at least $n(a-b)-a-O(\sqrt{n(a+b)}) \geq n(a-b)-o(n)$, which implies that the optimal solution of SDP1 will be at least $n(a-b)-o(n)$. Let $\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*$ be the optimal solution of the SDP, then we have

$$\begin{aligned} n(a-b) - o(n) &\leq \mathit{cost}(\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*) \\ &= \sum_{u,v} A_{uv} \langle \mathbf{x}_u^*, \mathbf{x}_v^* \rangle \\ &= \sum_{u,v} \left( A_{uv} - \frac{a+b}{n} \right) \langle \mathbf{x}_u^*, \mathbf{x}_v^* \rangle \end{aligned} \tag{2}$$

In the last equality we used the fact that $\sum_{u,v} \langle \mathbf{x}_u^*, \mathbf{x}_v^* \rangle = \| \sum_u \mathbf{x}_u^* \|^2 = 0$
Recall that the SDP norm of a matrix $M$ is defined to be

$$\|M\|_{SDP} := \max_{\substack{||\mathbf{x}_1||=\cdots=||\mathbf{x}_n||=1 \\ ||\mathbf{y}_1||=\cdots=||\mathbf{y}_n||=1}} \sum_{u,v} M_{uv} \langle \mathbf{x}_u, \mathbf{y}_v \rangle.$$

Let $R = \left( \begin{array}{c|c} \mathbf{p} & \mathbf{q} \\ \hline \mathbf{q} & \mathbf{p} \end{array} \right)$, then by Grothendieck inequality we have

$$\|A - R\|_{SDP} \leq O(1) \cdot \|A - R\|_{\infty \to 1}.$$

We proved that $\|A-R\|_{\infty\to 1} \leq O(n\sqrt{a+b})$ with high probability, so we know that the SDP norm $\|A-R\|_{SDP} \leq O(n\sqrt{a+b})$ with high probability as well. By definition, this means

$$\sum_{u,v} (A_{uv} - R_{uv}) \langle \mathbf{x}_u^*, \mathbf{x}_y^* \rangle \leq \|A - R\|_{SDP} \leq O(n\sqrt{a+b}). \tag{3}$$

Substracting 3 from 2, we obtain

$$\sum_{u,v} \left( A_{uv} - \frac{a+b}{n} - A_{uv} + R_{uv} \right) \langle \mathbf{x}_u^*, \mathbf{x}_y^* \rangle \geq n(a-b) - O(n\sqrt{a+b}). \tag{4}$$

Observe that

$$R = \left( \begin{array}{c|c} \mathbf{p} & \mathbf{q} \\ \hline \mathbf{q} & \mathbf{p} \end{array} \right) = \frac{p+q}{2} J + \frac{p-q}{2} C = \frac{a+b}{n} J + \frac{a-b}{n} C, \tag{5}$$

where $J$ is the all-one matrix and $C = \left(\begin{array}{c|c} \mathbf{1} & \mathbf{-1} \\ \hline \mathbf{-1} & \mathbf{1} \end{array}\right)$. Plugging 5 into 4, we get

$$\sum_{u,v} \frac{a-b}{n} C_{uv} \langle \mathbf{x}_u^*, \mathbf{x}_y^* \rangle \geq n(a-b) - O(n\sqrt{a+b}),$$

which can be simplified to

$$\sum_{u,v} C_{uv} \langle \mathbf{x}_u^*, \mathbf{x}_y^* \rangle \geq n^2 \left(1 - \frac{O(\sqrt{a+b})}{a-b}\right).$$

For simplicity, in the following analysis the term $\frac{O(\sqrt{a+b})}{a-b}$ will be called $1/c$. Notice that $C$ is a matrix with 1 for nodes from the same side of the cut and -1 for nodes from different sides of the cut, and $\langle \mathbf{x}_u^*, \mathbf{x}_y^* \rangle$ is an inner product of two unit vectors. If $1/c$ is very close to zero, then the sum will be very close to $n^2$. This means that $C_{uv} \langle \mathbf{x}_u^*, \mathbf{x}_y^* \rangle$ should be 1 for almost every pair of $(u, v)$, which shows that $X = \langle \mathbf{x}_u^*, \mathbf{x}_y^* \rangle$ is actually very close to $C$. Now, we will make this argument robust. To achieve this, we introduce the Frobenius norm of a matrix.

**Definition 5 (Frobenius norm)** *Let $M = (M_{ij})$ be a matrix. The Frobenius norm of $M$ is*

$$\|M\|_F := \sqrt{\sum_{i,j} M_{ij}^2}.$$

The following fact is a good exercise.

**Fact 6** *Let $M = (M_{ij})$ be a matrix. Then*

$$\|M\| \leq \|M\|_F,$$

*where $\|\cdot\|$ denotes the spectral norm.*

To see how close are $C$ and $X$, we calculate the Frobenius norm of $C - X$, which will be

$$\begin{aligned} \|C - X\|_F^2 &= \sum_{u,v} C_{uv}^2 + \sum_{u,v} X_{uv}^2 - 2\sum_{u,v} C_{uv} X_{uv} \\ &\leq 2n^2 - 2n^2 \left(1 - \frac{1}{c}\right) = \frac{2}{c} n^2. \end{aligned}$$

This gives us a bound on the spectral norm of $C - X$, namely

$$\|C - X\| \leq \|C - X\|_F \leq \sqrt{\frac{2}{c}} n.$$

Let $\mathbf{z} = (z_1, \ldots, z_n)$ be the unit eigenvector of $X$ corresponding to its largest eigenvalue, then by Davis-Kahan theorem we have[1]

$$\left\| \mathbf{z} - \frac{1}{\sqrt{n}} \chi \right\| \leq \sqrt{2} \cdot \frac{\sqrt{\frac{2}{c}} n}{n - \sqrt{\frac{2}{c}} n}.$$

For any $\epsilon > 0$, if $c$ is a large enough constant then we will have $\left\| \mathbf{z} - \frac{1}{\sqrt{n}} \chi \right\| \leq \sqrt{\epsilon}$. Now we have the following standard argument:

$$\begin{aligned} \epsilon n &\geq \|\sqrt{n}\mathbf{z} - \chi\|^2 \\ &= \sum_i (\sqrt{n} z_i - \chi_i)^2 \\ &\geq \#\{i : \operatorname{sign}(\sqrt{n} z_i) \neq \chi_i\}. \end{aligned}$$

The last inequality is because every $i$ with $\operatorname{sign}(\sqrt{n} z_i) \neq \chi_i$ will contribute at least 1 in the sum $\sum_i (\sqrt{n} z_i - \chi_i)^2$. This shows that our algorithm will misclassify at most $\epsilon n$ vertices.

[1] When we apply Davis-Kahan theorem, what we get is actually an upper bound on $\min\{\|\mathbf{z} - \chi/\sqrt{n}\|, \|-\mathbf{z} - \chi/\sqrt{n}\|\}$. We have assumed here that the bound holds for $\|\mathbf{z} - \chi/\sqrt{n}\|$, but the exact same proof will also work in the other case.