Lecture 9: Exact Community Detection in the SBM

In which we show that, for a certain range of parameters, the unique optimal solution of the Semidefinite Program associated to a SBM graph is the hidden partition.

1 The Algorithm

In the past lecture, we introduce the following SDP relaxation of the problem of finding a sparsest balanced cut.

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

And we introduced the following algorithm.

- 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.

If the graph is sampled from the $SBM_{n,p,q}$ distribution, and if we define, as usual, $a := pn/2$ and $b := qn/2$, today we will prove that

- There is an absolute constant β such that, if $a - b \geq \beta\sqrt{\log n}\sqrt{a+b}$, the algorithm recovers the exact solution with high probability.

If $\chi \in \{\pm 1\}^n$ is the vector such that $\chi_v = -1$ for $v \in V_1$ and $\chi_v = 1$ for $v \in V_2$, where (V_1, V_2) is the hidden partition, our goal will be to show that, with high probability, the unique optimal solution X^* of the SDP is $X^* = \chi\chi^T$, so the eigenvector of its largest eigenvalue is χ and the reconstruction is exact.

We will need to show that all feasible solutions $X \neq \chi\chi^T$ have an objective function value smaller than the objective function value of $\chi\chi^T$, and so we need tools to upper bound the value of feasible solutions. Such tools will come from the theory of *duality* of Semidefinite Programming. In order to introduce such ideas, we will first revisit duality in *Linear* Programming.

2 LP and SDP duality

Duality provides a method to upper bound the optimal value of maximization problems (and lower bound the optimal value of minimization problems). In the linear case, suppose we start with a maximization LP:

$$\max c^T x \ \text{ s.t. } \ Ax = b, \ x \geq 0,$$

where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ is a matrix of constraint coefficients.

If we wanted to certify that any feasible point for this LP satisfied a certain upper bound on $c^T x$, one natural strategy is to attempt to represent c as a linear combination of the rows a_j of A. Indeed, if we managed to write $c = \sum_{j=1}^m y_j a_j$, we would necessarily get that *any* feasible point had value exactly equal to

$$\sum_{j=1}^m y_j a_j^T x = y^T A x = b^T y.$$

Due to the non-negativity constraint on x, however, we do not need to require that c be a linear combination of the rows of A. Instead, it suffices to look for coefficients y_j such that $c \leq \sum_{j=1}^m y_j a_j = A^T y$. Any such y yields an upper bound on the optimum of $b^T y$. The problem of finding the optimal lower bound by this method is also an LP, given by:

$$\min b^T y \ \text{ s.t. } \ A^T y \geq c.$$

In the language of duality theory, (2) is the *primal* program and (2) is the *dual*.

The informal argument we sketched above can be formalized as the following chain of inequalities to prove that the value of the dual (2) is an upper bound on the value of the primal (2). Indeed, if $x \in \mathbb{R}^n$ is primal feasible and $y \in \mathbb{R}^m$ is dual feasible, then

$$\begin{aligned} c^T x &\leq \left(A^T y\right)^T x, \\ &= y^T A x \\ &= y^T b \\ &= b^T y, \end{aligned}$$

and the claim follows.

A similar construction can be defined for SDPs. To see how this works, let $\bullet$ denote the matrix inner product on $\mathbb{R}^{n\times n}$. That is, $M \bullet M' := \sum_{i,j} M_{ij} M'_{ij}$. An SDP can then be expressed as

$$\max C \bullet X, \quad \text{s.t.} \quad \begin{aligned} &A^{(j)} \bullet X = b_j, \ 1 \leq j \leq m, \\ &X \succeq 0 \end{aligned}$$

where the matrices $A^{(j)} \in \mathbb{R}^{n\times n}$ and the notation $M \succeq M'$ is understood as meaning that $M - M'$ is a PSD matrix.

We can attempt to port over the idea of the LP dual to this new setting as follows:

$$\min b^T y \ \text{ s.t. } \ \sum_{j=1}^m y_j A^{(j)} \succeq C.$$

We now prove that in fact this construction yields weak duality in the sense that the optimum of (2) is an upper bound on the optimum of (2).

For this, again suppose $X \in \mathbb{R}^{n\times n}$ is primal feasible and $y \in \mathbb{R}^m$ is dual feasible (that is, feasible for (2)). We then observe that

$$\begin{aligned} b^T y &= \sum_{j=1}^m y_j b_j \\ &= \sum_{j=1}^m y_j \left(A^{(j)} \bullet X\right) \\ &= \left[\sum_{j=1}^m y_j A^{(j)}\right] \bullet X. \end{aligned}$$

Now the only question is, how do we relate $\left[\sum_{j=1}^m y_j A^{(j)}\right] \bullet X$ to $C \cdot X$ given that $\left[\sum_{j=1}^m y_j A^{(j)}\right] \succeq C$? The following lemma answers this question for us.

Lemma 1 *Suppose $A,\ B \in \mathbb{R}^{n\times n}$ are PSD. Then $A \bullet B \geq 0$.*

Proof: Recall that we may write $A = \sum_{k=1}^{n} \lambda_k v_k v_k^T$, where each $\lambda_k \geq 0$ and the v_k form an orthonormal basis for $\mathbb{R}^n$. By linearity of the matrix inner product, we then have

$$A \bullet B = \sum_{k=1}^{n} \lambda_k \cdot \Big(v_k v_k^T \bullet B \Big).$$

But now we notice that

$$\begin{aligned} v_k v_k^T \bullet B &= \sum_{i=1}^{n} \sum_{j=1}^{n} v_{k,i} v_{k,j} \cdot B_{ij} \\ &= v_k^T B v_k \geq 0. \end{aligned}$$

Since each $\lambda_k \geq 0$, we conclude that $A \bullet B \geq 0$, as required. $\square$

Applying the conclusion of Lemma 1 to the result of the chain of inequalities (2), we find

$$b^T y \geq \left[\sum_{j=1}^{m} y_j A^{(j)} \right] \bullet X \geq C \bullet X.$$

Taking a minimum over the LHS and a maximum over the RHS yields the comparison of optima that we sought.

3 Duality for the SBM

To see how duality can help us with the SBM, let's first rewrite the SDP relaxation of the minimum balanced cut problem in a way that looks more like the SDP we analyzed above.

Our SDP can be formulated as

$$\begin{aligned} \max A \bullet X \quad \text{s.t.} \quad & E_{ii} \bullet X = 1,\ 1 \leq i \leq n \\ & J \bullet X = 0, \\ & X \succeq 0. \end{aligned}$$

where we used E_{ii} to denote the matrix that is equal to 1 in the position (i,i) and is zero everywhere else.

We note that in the notation of (2), we can let the constraint index run from 0 to n and set $A^{(0)} = J$ and $A^{(j)} = E_{jj}$ for $1 \leq j \leq n$. The constraint vector is then $b = \begin{pmatrix} 0 & 1^n \end{pmatrix} \in \mathbb{R}^{n+1}$.

We are now in a position to write down the dual. For convenience in what follows, we shall view the dual variable as $\begin{pmatrix} y_0 & y \end{pmatrix} \in \mathbb{R}^{n+1}$, so the notation will be slightly different from the dual formulation (2). In the modified notation, we have

$$\min \sum_{i=1}^{n} y_i \ \text{ s.t. } \ \operatorname{diag}\big(y\big) + y_0 J \succeq A,$$

where for any vector $v \in \mathbb{R}^n$, $\operatorname{diag}\big(v\big)$ denotes the corresponding diagonal matrix whose entry at $\big(i,\ i\big)$ is v_i.

4 Exact Reconstruction in the SBM

We are now ready to prove our main result, that we state formally below. As usual, $a = \frac{pn}{2}$ is the average internal degree and $b = \frac{qn}{2}$ is the average external degree.

Theorem 2 *There exists a universal constant $\beta > 0$ such that whenever $a - b > \beta\sqrt{\log n}\sqrt{a+b}$, the solution of the SDP relaxation (1) is given by $\chi\chi^T$. In particular, solving the SDP relaxation yields exact recovery in the SBM in this regime.*

4.1 A candidate dual certificate

The main idea behind our proof is that we already know a primal feasible solution to the SDP relaxation (1). Indeed, we can just set

$$X_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same community,} \\ -1 & \text{otherwise.} \end{cases} \tag{2}$$

Clearly $X_{ii} = 1$ for all $1 \leq i \leq n$ and $J \bullet X = 0$ because the communities have the same size. Meanwhile, $X = \chi\chi^T$, where χ is the indicator of the cut, so it is also PSD. Thus, it is a feasible point for the primal SDP (3) and corresponds to the optimum of the unrelaxed combinatorial problem.

Our aim is to show that actually the feasible solution (2) is the unique optimal solution. The first step toward this goal is to show that it is in fact optimal, and we shall do this by exhibiting a dual solution whose dual objective value is equal to the primal value of the combinatorial solution.

Notice that the value of the combinatorial solution in the primal is given by

$$\begin{aligned} A \bullet X &= \sum_{i=1}^{n} \left[\sum_{j \text{ in same community as } i} A_{ij} - \sum_{j \text{ in other community}} A_{ij} \right] \\ &= \sum_{i=1}^{n} \Big(a_i - b_i \Big), \end{aligned}$$

where a_i is the within-community (or internal-) degree of i and b_i is the cross-community (or external-) degree of i.

A candidate for a dual solution with the same objective value is thus given by taking

$$y_i = a_i - b_i,\ 1 \leq i \leq n, \quad \text{and} \quad y_0 = \frac{a+b}{n},$$

where the latter only matters for feasibility and not the objective value. We stress that this vector is actually a random variable, since the a_i and b_i are random quantities.

It is clear by inspection that (4.1) specifies dual variables that achieve the value of the combinatorial solution (2). The only question is whether this specification of the variables yields a dual feasible point. The main thrust of the proof is thus to show that, with high probability, the proposed dual solution is indeed feasible—and therefore optimal, since it achieves a primal feasible value of the objective function.

Feasiblity of the proposed dual certificate is equivalent to the positive definiteness condition

$$M :\ = \operatorname{diag}\big(y\big) + y_0 J - A \succeq 0.$$

We shall actually prove a stronger statement that will be useful for the complementary slackness part of the proof. We intend to show that (a) $M\chi = 0$ with probability 1, where again χ is the indicator of the cut and that (b) $x^T M x > 0$ for all nonzero $x \perp \chi$ with high probability.

The first statement (a) we shall directly verify below. The second (b) will follow from a matrix concentration argument: we shall show that apart from one eigenvalue of 0 corresponding to χ, the eigenvalues of $\mathbb{E}\left[M\right]$ are all $a - b$ and we will then argue that with high probability $\big|\big|M - \mathbb{E}\left[M\right]\big|\big|_{\text{op}} \leq O\big(\sqrt{\log n}\sqrt{a+b}\big)$, which will yield the theorem when $a - b > c\sqrt{\log n}\sqrt{a+b}$ for a suitable absolute constant $c > 0$.

To verify that $M\chi = 0$ with probability 1, we compute

$$\begin{aligned} M\chi &= \operatorname{diag}\big(y\big) \cdot \chi - A\chi + J\chi \\ &= \operatorname{diag}\big(y\big) \cdot \chi - A\chi, \end{aligned}$$

where we have used balance to deduce $J\chi = 0$. We now observe that

$$\big(\operatorname{diag}\big(y\big) \cdot \chi\big)_i = \big(a_i - b_i\big)\chi_i,$$

while

$$\begin{aligned} \big(A\chi\big)_i &= \sum_{j=1}^{n} A_{ij}\chi_j \\ &= \chi_i \sum_{j=1}^{n} A_{ij}\chi_i\chi_j \\ &= \big(a_i - b_i\big)\chi_i, \end{aligned}$$

where we have used the fact that $\chi_i \chi_j = 1$ if i and j are in the same community and -1 otherwise. Thus, $\mathrm{diag}\big(y\big)\chi = A\chi$ and the claim follows.

On the other hand, we observe $\mathbb{E}\left[y_i\right] = \mathbb{E}\left[a_i - b_i\right] = a - b$ for $1 \leq i \leq n$. Thus,

$$
\begin{aligned}
\mathbb{E}\left[M\right] &= \big(a-b\big)\cdot I + \frac{a+b}{n}\cdot J - \mathbb{E}\left[A\right] \\
&= \big(a-b\big)\cdot I + \frac{a+b}{n}\cdot J - \frac{a+b}{n}\cdot J - \frac{a-b}{n}\cdot \chi\chi^T \\
&= \big(a-b\big)\cdot I - \frac{a-b}{n}\cdot \chi\chi^T.
\end{aligned}
$$

It is now clear that χ is in the nullspace of $\mathbb{E}\left[M\right]$ (as it must be since it is in the nullspace of M with probability 1) and that all the other eigenvalues of $\mathbb{E}\left[M\right]$ are $a - b$, as claimed.

We now make the following claim without proof. The proof uses a matrix analog of the Chernoff bound (the "Matrix Bernstein inequality").

Lemma 3 *With high probability over the choice of the graph,*

$$||M - \mathbb{E}\left[M\right]|| \leq O\big(\sqrt{\log n}\sqrt{a+b}\big)$$

The following Lemma is the core of our analysis. (Note that it is not a probabilistic statement.)

Lemma 4 *Suppose that $a - b > ||M - \mathbb{E}\, M||$. Then the solution $X = \chi\chi^T$ is the **unique** optimum for the minimum bisection SDP.*

PROOF: We begin by showing that $y_0, \ldots, y_n$ defined above is feasible for the dual, which implies that $\chi\chi^T$ is *an* optimal solution. Let $\mathbf{x}$ be any vector, and write $\mathbf{x} = \alpha\chi + \mathbf{y}$, where $\mathbf{y}$ is orthogonal to χ. Then

$$
\begin{aligned}
\mathbf{x}^T M \mathbf{x} &= (\alpha\chi + \mathbf{y})^T M (\alpha\chi + \mathbf{y}) \\
&= \alpha^2 \chi^T M \chi + 2\alpha \mathbf{y} M \chi + \mathbf{y}^T M \mathbf{y} \\
&= \mathbf{y}^T M \mathbf{y}
\end{aligned}
$$

where the last line uses the fact, which we established before, that $M\chi = 0$ for every graph. Since $\mathbf{y}$ is orthogonal to χ, we have

$$\mathbf{y}^T(\mathbb{E}\, M)\mathbf{y} = (a-b)\cdot ||\mathbf{y}||^2$$

and, by definition of spectral norm and the assumption of the lemma, we have that, if $\mathbf{y} \neq 0$,

$$\mathbf{y}^T M \mathbf{y} \geq \mathbf{y}^T (\mathbb{E}\, M)\mathbf{y} - ||M - \mathbb{E}\, M|| \cdot ||\mathbf{y}||^2 = (a - b - ||M - \mathbb{E}\, M||) \cdot ||\mathbf{y}||^2 > 0$$

We conclude that

$$\mathbf{x}^T M \mathbf{x} \geq 0$$

for every $\mathbf{x}$, meaning that $M \succeq \mathbf{0}$, and that

$$\mathbf{x}^T M \mathbf{x} > 0$$

for every $\mathbf{x}$ that is not a multiple of χ.

From the fact that M is PSD we have

$$\begin{aligned}
\sum_v (a_v - b_v) &= \mathrm{Cost}(\{\mathbf{x}_v\}_{v \in V}) = A \cdot X \\
&\leq [\mathrm{diag}(a_1 - b_1, \ldots, a_n - b_n) + \frac{a+b}{n} \cdot J] \cdot X \\
&\leq \mathrm{diag}(a_1 - b_1, \ldots, a_n - b_n) \cdot X + \frac{a+b}{n}[J \cdot X] \\
&= \sum_v y_v \cdot X_{v,v} + \frac{a+b}{n} \cdot \sum_{u,v} \langle \mathbf{x}_u, \mathbf{x}_v \rangle \\
&= \sum_v y_v + \frac{a+b}{n} \| \sum_v \mathbf{x}_v \|^2 \\
&= \sum_v (a_v - b_v)
\end{aligned}$$

Which implies that $\chi\chi^T$ is an optimal solution. Let now X be any other optimal solution.

Thus, all of the above inequalities are actually equalities, and we have:

$$A \cdot X = [\mathrm{diag}(a_1 - b_1, \ldots, a_n - b_n) + \frac{a+b}{n} \cdot J] \cdot X$$

$$\text{which implies } [\mathrm{diag}(a_1 - b_1, \ldots, a_n - b_n) + \frac{a+b}{n} \cdot J - A] \cdot X = M \cdot X = 0$$

To show uniqueness of our solution $\chi\chi^T$, it suffices to show that the $X = \chi\chi^T$ is the only solution that satisfies $M \cdot X = 0$.

We showed that the assumption of the lemma imply that for all $\mathbf{x} \perp \chi$:

$$\mathbf{x}^T M \mathbf{x} \geq (a-b) - O(\sqrt{\log n}\sqrt{a+b}) > 0$$

We can also write a PSD matrix as positive combinations of certain rank 1 matrices $\mathbf{z}_i \mathbf{z}_i^T$, so:

$$X = \sum_i \lambda_i \mathbf{z}_i \mathbf{z}_i^T \text{ where } \lambda_i > 0$$

$$M \cdot X = M \cdot (\sum_i \lambda_i \mathbf{z}_i \mathbf{z}_i^T) M = \sum_i \lambda_i (\mathbf{z}_i^T M \mathbf{z}_i)$$

The quantity $\mathbf{z}_i^T M \mathbf{z}_i$ will always be strictly positive, unless either $\lambda_i = 0$ or $\mathbf{z}_i$ is parallel to χ. Therefore, if $M \cdot X = 0$, we must have $X = \chi\chi^T$, which proves uniqueness of our solution. □

Putting everything together, we see that there is a constant β such that, if $a - b > \beta \cdot \sqrt{\log n} \cdot \sqrt{a+b}$, then with high probability the unique optimum of the SDP is $\chi\chi^T$ and the algorithm of the previous lecture finds the hidden partition.