Spectral Graph Theory Lecture 21

Spectral Partitiong in a Stochastic Block Model

Daniel A. Spielman November 16, 2015

Disclaimer

These notes are not necessarily an accurate representation of what happened in class. The notes written before class say what I think I should say. I sometimes edit the notes after class to make them way what I wish I had said.

There may be small mistakes, so I recommend that you check any mathematically precise statement before using it in your own work.

These notes were last revised on December 8, 2015.

21.1 Overview

In this lecture, we will perform a crude analysis of the performance of spectral partitioning algorithms in what are called *stochastic block models* or a *planted partition model*. The name you choose largely depends on your community and application. As we are especially interested today in partitioning, we will call it the planted partition model. In this model, we build a random graph that has a natural partition.

The simplest model of this form is for the graph bisection problem. This is the problem of partitioning the vertices of a graph into two equal-sized sets while minimizing the number of edges bridging the sets. To create an instance of the planted bisection problem, we first choose a paritition of the vertices into equal-sized sets X and Y. When then choose probabilities $p > q$, and place edges between vertices with the following probabilities:

$$\Pr\left[(u,v) \in E\right] = \begin{cases} p & \text{if } u \in X \text{ and } v \in X \\ p & \text{if } u \in Y \text{ and } v \in Y \\ q & \text{otherwise.} \end{cases}$$

The expected number of edges crossing between X and Y will be $q\,|X|\,|Y|$. If p is sufficiently larger than q, then every other bisection will have more crossing edges. If p is too close to q, then the partition given by X and Y will not be the smallest. We will consider the case $p = 1/2$ and $q = p - 24/\sqrt{n}$. If $q = p - \epsilon/\sqrt{n}$ for small ϵ, then one cannot hope to distinguish between X and Y.

In this lecture, we will show that this partition can be recovered from the second eigenvector of the adjacency matrix of the graph. This will be a crude version of an analysis of McSherry [McS01].

McSherry analyzed more general models than this, and his analysis has since been tightened and better results have been obtained in a number of recent papers. To get started, look at [MNS14, DKMZ11, BLM15, Mas14, Vu14]. My goal today is just to give you the idea behind his analysis, as it has had a substantial impact well beyond stochastic block models. McSherry's analysis considers the adjacency matrix of the generated graph as a perturbation of one ideal probability matrix. In the probability matrix the second eigenvector provides a clean partition of the two blocks. McSherry shows that the difference between the generated matrix and the ideal one is small, and so the generated matrix can be viewed as a small perturbation of the idea one. He then uses matrix perturbation theory to show that the second eigenvector of the generated matrix will probably be close to the second eigenvector of the original, and so it reveals the partition.

The idea of using perturbation theory to analyze random objects generated from nice models has been very powerful. For example, it inspired Shkolnisky and Singer to design an exciting algorithm for the image processing problems that occur in cryo-electron microscopy [SS11, SS12].

I finish this introduction with a warning. Stochastic block models have been the focus of a lot of research lately, and there are now very good algorithms for solving problems on graphs generated from these models. But, these are just models and very little real data resembles the product of these models. So, there is no reason to believe that algorithms that are optimized for these models will be useful in practice. Neverthelss, some of them are.

21.2 The Perturbation Approach

As long as we don't tell our algorithm, we can choose $X = \{1, \ldots, n/2\}$ and $Y = \{n/2+1, \ldots, n\}$. Let's do this for simplicity.

Define the matrix

$$\boldsymbol{M} = \begin{bmatrix} p & \cdots & p & q & \cdots & q \\ \vdots & & & & \vdots & \\ p & \cdots & p & q & \cdots & q \\ q & \cdots & q & p & \cdots & p \\ \vdots & & & & \vdots & \\ q & \cdots & q & p & \cdots & p \end{bmatrix} = \begin{bmatrix} p\boldsymbol{J}_{n/2} & q\boldsymbol{J}_{n/2} \\ q\boldsymbol{J}_{n/2} & p\boldsymbol{J}_{n/2} \end{bmatrix},$$

where we write $\boldsymbol{J}_{n/2}$ for the square all-1s matrix of size $n/2$.

The adjacency matrix of the planted partition graph is obtained by setting $\boldsymbol{A}(a,b) = 1$ with probability $\boldsymbol{M}(a,b)$, subject to $\boldsymbol{A}(a,b) = \boldsymbol{A}(b,a)$ and $\boldsymbol{A}(a,a) = 0$. So, this is a random graph, but the probabilities of some edges are different from others.

We will study a very simple algorithm for finding an approximation of the planted bisection: compute $\boldsymbol{v}_2$, the eigenvector of the second-largest eigenvalue of $\boldsymbol{A}$. Then, set $S = \{a : \boldsymbol{v}_2(a) < 0\}$. We guess that S is one of the sets in the bisection. We will show that under reasonable conditions on p and q, S will be mostly right. For example, we might consider $p = 1/2$ and $q = 1/2 - 12/\sqrt{n}$. Intuitively, the reason this works is that $\boldsymbol{A}$ is a slight perturbation of $\boldsymbol{M}$, and so the eigenvectors of $\boldsymbol{A}$ should look like the eigenvectors of $\boldsymbol{M}$.

To see why this would be useful, let's look at the eigenvectors of $\boldsymbol{M}$. Of course, the constant vectors are eigenvectors of $\boldsymbol{M}$. We have

$$\boldsymbol{M}\mathbf{1} = \frac{n}{2}(p+q)\mathbf{1},$$

and so the corresponding eigenvalue is

$$\mu_1 \stackrel{\text{def}}{=} \frac{n}{2}(p+q).$$

The second eigenvector of $\boldsymbol{M}$ has two values: one on X and one on Y. Let's be careful to make this a unit vector. We take

$$\boldsymbol{w}_2(a) = \begin{cases} \frac{1}{\sqrt{n}} & a \in X \\ -\frac{1}{\sqrt{n}} & a \in Y. \end{cases}$$

Then,

$$\boldsymbol{M}\boldsymbol{w}_2 = \frac{n}{2}(p-q)\boldsymbol{w}_2,$$

and the corresponding eigenvalue is

$$\mu_2 \stackrel{\text{def}}{=} \frac{n}{2}(p-q).$$

As $\boldsymbol{M}$ has rank 2, all the other eigenvalues of $\boldsymbol{M}$ are zero.

Let $\boldsymbol{A}$ be the adjacency matrix of the generated graph. I wish to consider

$$\widehat{\boldsymbol{A}} \stackrel{\text{def}}{=} \boldsymbol{A} + p\boldsymbol{I}.$$

Note that the eigenvectors of $\widehat{\boldsymbol{A}}$ and $\boldsymbol{A}$ are the same, so considering $\widehat{\boldsymbol{A}}$ won't change our analysis at all. But, $\widehat{\boldsymbol{A}}$ is closer to $\boldsymbol{M}$. We now consider the difference between $\widehat{\boldsymbol{A}}$ and $\boldsymbol{M}$:

$$\boldsymbol{R} = \widehat{\boldsymbol{A}} - \boldsymbol{M}.$$

For (a, b) in the same component,

$$\Pr\left[\boldsymbol{R}(a,b) = 1-p\right] = p \quad \text{and}$$
$$\Pr\left[\boldsymbol{R}(a,b) = -p\right] = 1-p,$$

and for (a, b) in different components,

$$\Pr\left[\boldsymbol{R}(a,b) = 1-q\right] = q \quad \text{and}$$
$$\Pr\left[\boldsymbol{R}(a,b) = -q\right] = 1-q.$$

One can use matrix concentration bounds to prove that $\|\boldsymbol{R}\|$ is probably small. The bounds like those we studied in Lecture 17 are too general to give a sharp result here. So, instead we appeal to a result of Vu [Vu07, Theorem 1.4], which implies the following.

Theorem 21.2.1. *There exist constants c_1 and c_2 such that with probability approaching 1,*

$$\|\boldsymbol{R}\| \leq 2\sqrt{pn} + c_1(pn)^{1/4}\ln n,$$

provided that

$$p \geq c_2\frac{\ln^4 n}{n}.$$

We apply the following corollary.

Corollary 21.2.2. *There exists a constant c_0 such that with probability approaching* 1*,*

$$\|\boldsymbol{R}\| \leq 3\sqrt{pn},$$

provided that

$$p \geq c_0 \frac{\ln^4 n}{n}.$$

In fact, Alon, Krivelevich and Vu [AKV02] prove that the probability that the norm of $\boldsymbol{R}$ exceeds this value by more than t is exponentially small in t. However, we will not need that fact for this lecture.

If we used concentration bounds such as those from Lecture 17, the 3 in the above bound would be a $\sqrt{\log n}$. Ignoring the details of the asymptotics, let's just assume that $\|\boldsymbol{R}\|$ is small, and investigate the consequences.

21.3 Perturbation Theory for Eigenvectors

Let $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ be the eigenvalues of $\widehat{\boldsymbol{A}}$, and let $\mu_1 > \mu_2 > 0 = \mu_3 = \cdots = \mu_n$ be the eigenvalues of $\boldsymbol{M}$. Using the Courant-Fischer theorem, one can prove that

$$|\alpha_i - \mu_i| \leq \|R\|.$$

I was going to make this an exercise, but I forgot. In particular, if

$$\|\boldsymbol{R}\| < \frac{n}{4}(p-q),$$

then

$$\frac{n}{4}(p-q) < \alpha_2 < \frac{3n}{4}(p-q)$$

and, assuming $q > p/3$, we have

$$\alpha_1 > \mu_1 - \|R\| > \frac{n}{2}(p+q) - \frac{n}{4}(p-q) \geq \frac{3n}{4}(p-q).$$

So, we can view α_2 as a perturbation of μ_2. The natural question is whether we can view $\boldsymbol{v}_2$ as a perturbation of $\boldsymbol{w}_2$.

The Davis-Kahan theorem [DK70] says that $\boldsymbol{v}_2$ will be close to $\boldsymbol{w}_2$, in angle, if the norm of $\boldsymbol{R}$ is significantly less than the distance between μ_2 and the other eigenvalues of $\boldsymbol{M}$. That is, the eigenvector does not move too much if its corresponding eigenvalue is isolated.

Theorem 21.3.1. *Let $\widehat{\boldsymbol{A}}$ and $\boldsymbol{M}$ be symmetric matrices. Let $\boldsymbol{R} = \boldsymbol{M} - \widehat{\boldsymbol{A}}$. Let $\alpha_1 \geq \cdots \geq \alpha_n$ be the eigenvalues of $\widehat{\boldsymbol{A}}$ with corresponding eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ and let Let $\mu_1 \geq \cdots \geq \mu_n$ be the eigenvalues of $\boldsymbol{M}$ with corresponding eigenvectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$. Let θ_i be the angle between $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$. Then,*

$$\sin 2\theta_i \leq \frac{2\|R\|}{\min_{j \neq i} |\mu_i - \mu_j|}.$$

Note that this theorem is really bounding the angle between the lines through the eigenvectors, rather than the particular eigenvectors. For this reason, the angle can never be more than $\pi/2$. You can find a proof of this theorem in [Dem97, Theorem 5.4]. I will prove and use a slightly weaker statement in which we replace 2θ with θ.

21.4 Partitioning

Consider

$$\boldsymbol{\delta} = \boldsymbol{v}_2 - \boldsymbol{w}_2.$$

For every vertex i that is misclassified by $\boldsymbol{v}_2$, we have $|\boldsymbol{\delta}(i)| \geq \frac{1}{\sqrt{n}}$. So, if $\boldsymbol{v}_2$ misclassifies k vertices, then

$$\|\boldsymbol{\delta}\| \geq \sqrt{\frac{k}{n}}.$$

As $\boldsymbol{w}_2$ and $\boldsymbol{v}_2$ are unit vectors, we may apply the crude inequality

$$\|\boldsymbol{\delta}\| \leq \sqrt{2}\sin\theta_2$$

(the $\sqrt{2}$ disappears as θ_2 gets small).

To combine this with the perturbation bound, we assume $q > p/3$, and find

$$\min_{j\neq 2} |\mu_2 - \mu_j| = \frac{n}{2}(p-q).$$

Assuming that $\|R\| \leq 3\sqrt{pn}$, we find

$$\sin\theta_2 \leq \frac{2\cdot 3\sqrt{pn}}{\frac{n}{2}(p-q)} = \frac{12\sqrt{p}}{\sqrt{n}(p-q)}.$$

So, the number k of misclassified vertices satisfies

$$\sqrt{\frac{k}{n}} \leq \frac{\sqrt{2}12\sqrt{p}}{\sqrt{n}(p-q)},$$

which implies

$$k \leq \frac{288p}{(p-q)^2}.$$

So, if p and q are both constants, we expect to misclassify at most a constant number of vertices. If $p = 1/2$, and $q = p - 24/\sqrt{n}$, then we get

$$\frac{288p}{(p-q)^2} = \frac{n}{4},$$

so we expect to misclassify at most a constant fraction of the vertices.

21.5 Proof of the Davis-Kahan Theorem

For simplicity, we will prove a statement that is weaker by a factor of 2.

Proof of Theorem 21.3.1. By considering the matrices $\boldsymbol{M} - \mu_i I$ and $\widehat{\boldsymbol{A}} - \mu_i I$ instead of $\boldsymbol{M}$ and $\widehat{\boldsymbol{A}}$, we can assume that $\mu_i = 0$. As the theorem is vacuous if μ_i has multiplicity more than 1, we may also assume that μ_i has multiplicity 1 as an eigenvalue, and that $\boldsymbol{w}_i$ is a unit vector in the nullspace of $\boldsymbol{M}$.

Our assumption that $\mu_i = 0$ also leads to $|\lambda_i| \leq \|R\|$.

Expand $\boldsymbol{v}_i$ in the eigenbasis of $\boldsymbol{M}$, as

$$\boldsymbol{v}_i = \sum_j c_j \boldsymbol{w}_j, \qquad \text{where } c_j = \boldsymbol{w}_j^T \boldsymbol{v}_i.$$

Setting

$$\delta = \min_{j \neq i} |\mu_j|,$$

we may compute

$$\begin{aligned} \|M\boldsymbol{v}_i\|^2 &= \sum_j c_j^2 \mu_j^2 \\ &\geq \sum_{j \neq i} c_j^2 \delta^2 \\ &= \delta^2 \sum_{j \neq i} c_j^2 \\ &= \delta^2 (1 - c_i^2) \\ &= \delta^2 \sin^2 \theta_i. \end{aligned}$$

On the other hand,

$$\|\boldsymbol{M}\boldsymbol{v}_i\| \leq \|\boldsymbol{A}\boldsymbol{v}_i\| + \|\boldsymbol{R}\boldsymbol{v}_i\| = \lambda_i + \|\boldsymbol{R}\boldsymbol{v}_i\| \leq 2\|\boldsymbol{R}\|.$$

So,

$$\sin \theta_i \leq \frac{2\|\boldsymbol{R}\|}{\delta}.$$

□

It may seem surprising that the amount by which eigenvectors move depends upon how close their respective eigenvalues are to the other eigenvalues. However, this dependence is necessary. To see why, first consider a matrix with a repeated eigenvalue, such as

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Now, let $\boldsymbol{v}$ be any unit vector, and consider

$$B = A + \epsilon \boldsymbol{v}\boldsymbol{v}^T.$$

The matrix B will have $\boldsymbol{v}$ as an eigenvector of eigenvalue $1+\epsilon$ as well as an eigenvalue of 1. So, by making an arbitrarily small perturbation, we were able to select which eigenvalue of B was largest.

To make this effect clearer, let $\boldsymbol{w}$ be any other unit vector, and consider the matrix

$$C = A + \epsilon \boldsymbol{w}\boldsymbol{w}^T.$$

So, $\boldsymbol{w}$ is the eigenvector of C of eigenvalue $(1+\epsilon)$, and the other eigenvalue is 1. On the other hand,

$$\|C - B\| \leq \left\|\epsilon \boldsymbol{w}\boldsymbol{w}^T\right\| + \left\|\epsilon \boldsymbol{w}\boldsymbol{w}^T\right\| = 2\epsilon.$$

So, while B and C differ very little, their dominant eigenvectors can be completely different. This is because the eigenvalues were close together.

21.6 Further Reading

If you would like to know more about bounding norms and eigenvalues of random matrices, I recommend [Ver10] and [Tro12].

References

[AKV02] Noga Alon, Michael Krivelevich, and Van H. Vu. On the concentration of eigenvalues of random symmetric matrices. *Israel Journal of Mathematics*, 131(1):259–267, 2002.

[BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. *arXiv preprint arXiv:1501.06087*, 2015.

[Dem97] James W Demmel. *Applied numerical linear algebra*. Siam, 1997.

[DK70] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.

[Mas14] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM, 2014.

[McS01] F. McSherry. Spectral partitioning of random graphs. In *FOCS '01: Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, page 529, 2001.

[MNS14] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Proceedings of The 27th Conference on Learning Theory*, pages 356–370, 2014.

[SS11] Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM journal on imaging sciences*, 4(2):543–572, 2011.

[SS12] Yoel Shkolnisky and Amit Singer. Viewing direction estimation in cryo-em using synchronization. *SIAM journal on imaging sciences*, 5(3):1088–1110, 2012.

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

[Ver10] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[Vu07] Van Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.

[Vu14] Van Vu. A simple svd algorithm for finding hidden partitions. *arXiv preprint arXiv:1404.3918*, 2014.