Spectral Graph Theory Lecture 21

Spectral Partitioning in the Planted Partition Model

Daniel A. Spielman November 11, 2009

21.1 Introduction

In this lecture, we will perform a crude analysis of the performance of spectral partitioning algorithms in 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 V^1 and V^2. When then choose probabilities $p > q$, and place edges between vertices with the following probabilities:

$$\Pr\left[(u,v) \in E\right] = \begin{cases} p & \text{if } u \in V^1 \text{ and } v \in V^1 \\ p & \text{if } u \in V^2 \text{ and } v \in V^2 \\ q & \text{otherwise.} \end{cases}$$

The expected number of edges crossing between V^1 and V^2 will be $q\left|V^1\right|\left|V^2\right|$. If p is sufficiently larger than q, then every other bisection will have more crossing edges.

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]

There have been many analyses of graph partitioning algorithms under planted partition models such as this. The model is motivated by the idea that vertices (or general items) belong to certain categories, and that vertices in the same categories are more likely to be connected. Such models also arise in the analysis of clustering algorithms. However, it is not clear that these models represent practice very well.

21.2 The Perturbation Approach

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

Define the matrix

$$M = \begin{bmatrix} p & \cdots & p & q & \cdots & q \\ \vdots & & & \vdots & & \\ p & \cdots & p & q & \cdots & q \\ q & \cdots & q & p & \cdots & p \\ \vdots & & & \vdots & & \\ q & \cdots & q & p & \cdots & p \end{bmatrix} = \begin{bmatrix} pJ_{n/2} & qJ_{n/2} \\ qJ_{n/2} & pJ_{n/2} \end{bmatrix},$$

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

The adjacency matrix of the planted partition graph is obtained by setting $A(i,j) = 1$ with probability $M(i,j)$, subject to $A(i,j) = A(j,i)$. 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 A. Then, set $S = \{i : \boldsymbol{v}_2(i) < 0\}$. 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. Intuitively, the reason is that A is a slight perturbation of M, and so the eigenvectors of A should look like the eigenvectors of M.

For that to make sense, I should have said what the eigenvectors M look like. Of course, the constant vectors are eigenvectors of M. We have

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

The second eigenvector of M has two values: one on V^1 and one on V^2. Let's be careful to make this a unit vector. We take

$$\boldsymbol{w}_2(i) = \begin{cases} \frac{1}{\sqrt{n}} & i \in V^1 \\ -\frac{1}{\sqrt{n}} & i \in V^2. \end{cases}$$

Then,

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

So, the second-largest eigenvalue of M is $(n/2)(p-q)$. As M has rank 2, all the other eigenvalues of M are zero.

Now, let $R = A - M$. For (u,v) in the same component,

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

and for (u,v) in different components,

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

As in the last lecture, we can bound the probability that the norm of R is large. While I'm not yet sure if we can bound it using the same technique, we can 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*,*

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

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*,*

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

provided that

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

In fact, Krivelevich and Vu [**?**, Theorem ??] prove that the probability that the norm of R exceeds this value by more than t is exponentially small in t. However, we will not need that fact for this lecture.

Ignoring the details of the asymptotics, let's just assume that $\|R\|$ 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 A, and let $\mu_1 > \mu_2 > \mu_3 = \cdots = \mu_n$ be the eigenvalues of M. We know from a problem set that

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

In particular, if

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

then

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

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

$$\alpha_1 > \frac{3n}{4}(p-q).$$

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

Here is the theory that says we can.

Theorem 21.3.1. *Let A and M be symmetric matrices. Let $R = M - A$. Let $\alpha_1 \geq \cdots \geq \alpha_n$ be the eigenvalues of 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 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,*

$$\sin \theta_i \leq \frac{2 \|R\|}{\min_{j \neq i} |\alpha_i - \alpha_j|},$$

and

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

We remark that this bound may be tightened slightly, essentially eliminating the 2. We defer the proof of this theorem for a few minutes, and first see what it implies.

21.4 Partitioning

Consider

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

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

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

As $\boldsymbol{w}$ and $\boldsymbol{v}$ 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

$$\sin \theta_2 \leq \frac{3\sqrt{pn}}{\frac{n}{2}(p-q)} = \frac{6\sqrt{p}}{\sqrt{n}(p-q)}.$$

So, the number k of misclassified vertices satisfies

$$\sqrt{\frac{k}{n}} \leq \frac{6\sqrt{p}}{\sqrt{n}(p-q)},$$

which implies

$$k \leq \frac{36p}{(p-q)^2}.$$

So, if p and q are both constants, we expect to misclassify at most a constant number of vertices. If $p = 1/2$, and $q = p - 12/\sqrt{n}$, then we get

$$\frac{36p}{(p-q)^2} = \frac{n}{8},$$

so we expect to mis-classify at most a constant fraction of the vertices.

21.5 Proof of Eigenvector Perturbation

Proof of Theorem 21.3.1. By considering the matrices $M - \lambda_i I$ and $A - \mu_i I$ instead of M and 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 M.

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

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

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

So,

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

□

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 Improving the Partition

If I get a chance, I'll describe how one improves such a partition in practice, and how McSherry did it in theory. I'll begin by observing that the analysis we performed is very pessimistic. It relies on an upper bound on $\|\boldsymbol{w}_2 - \boldsymbol{v}_2\|$. But, $\boldsymbol{v}_2$ was produced by a random process. So, it seems unlikely that all of its weight would be concentrated on a few vertices.

References

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

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