Emanuele Ceccarelli - Spectral Graph Theory I, Part II

Outline:

- Review
- Useful operators
- Minimizing energy
- Maximizing energy

Review:

Before starting with today's material, I want to quickly review some topics from Christopher's lecture.

One of the main points regards the Dirichlet Form, or local variance, $\mathcal{E}$. We will also call it energy.

$$\mathcal{E}(f) = \frac{1}{2} \mathop{E}_{u \sim v} \left[(f(u) - f(v))^2 \right]$$

Last time, we analyzed this with an example indicator function, which will turn out to be useful today.

Let $f = 1_S$ be defined as $\begin{cases} f(v) = 1 & \text{if } v \in S \\ f(v) = 0 & \text{otherwise} \end{cases}$

Then, $\mathcal{E}(f) = \Pr_{u \sim v} [u \in S, v \notin S]$

Another useful concept I want to review is that of the stationary distribution π, which can be used to draw vertices with probability proportional to their degrees.

$$\text{so, } \pi(v) = \frac{\deg(v)}{2|E|}$$

Another thing to recall is that vertex functions like the one we saw before can be written as vectors:

$$\begin{bmatrix} f(v_1) \\ f(v_2) \\ \vdots \\ \\ \end{bmatrix}$$

Putting this notion together with π allows us to talk about inner products

$\langle f, g \rangle$ = weighted dot product of the two vectors ($\pi \to$ weights)

Now, we'll develop some tools that will be useful in the process of minimizing and maximizing energy.

Operators

When computing the local variance, given a vertex u and a function f, we're often concerned with computing f at the neighbors of u.

So, we define an operator K, such that

$$(Kf)(u) := \mathbb{E}_{v \sim u}[f(v)] \quad \text{(normalized adjacency operator)}$$

K is a linear operator: $K(f+g) = Kf + Kg$. Because of this, just like we saw f as a vector, we can see K as a matrix. So, we can see Kf as:

$$\begin{bmatrix} & & \\ & K & \\ & & \end{bmatrix} \begin{bmatrix} \\ f \\ \\ \end{bmatrix} = \begin{bmatrix} \\ Kf \\ \\ \end{bmatrix}$$

From the definition of K, we need to take a weighted average of the various $f(v)$ for $v \sim u$. Thus, $K_{uv} = \dfrac{\#\text{ edges between } u \text{ and } v}{\deg(u)}$

K is then the normalized adjacency matrix of the graph, hence the name of K.

(Optional) K and random walks

Suppose that we have a vector u which is all 0's and a 1 in correspondence of a specific vertex. This represents "being at that vertex". Now, find Ku. This is the distribution after a step of the random walk.
The same works if u is a random variable, or the function of vertices

So, $(K^2 f)(u) = \mathbb{E}_{u \to\to w}[f(w)]$

And, also $(K^t f)(u) = \mathbb{E}_{u \xrightarrow[t \text{ steps}]{} w}[f(w)]$

The normalized Laplacian operator

We want, first of all, to expres $\mathcal{E}(f)$ using K.

Notice that $\langle f, Kg\rangle = \mathbb{E}_{u \sim v}[f(u) \cdot g(v)]$, so $\langle f, Kg\rangle = \langle Kf, g\rangle$

Now: $\mathcal{E}(f) = \frac{1}{2} \mathbb{E}_{u \sim v}[(f(u) - f(v))^2]$

$$= \langle f, f\rangle - \mathbb{E}_{u \sim v}[f(u)f(v)]$$

$$= \langle f, f\rangle - \langle f, Kf\rangle$$

$$= \langle f, f - Kf\rangle$$

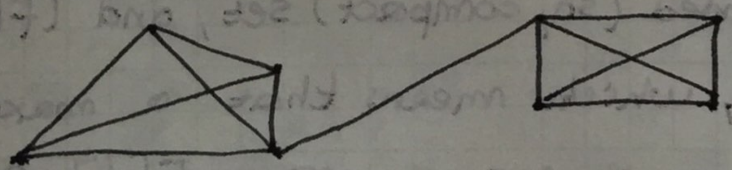
$$= \langle f, (I-K)f\rangle$$

We define $L = I - K$, so that $\mathcal{E}(f) = \langle f, Lf\rangle$, and we call L the normalized laplacian operator (I is the identity matrix)

We now move to what today's lecture's actual purpose is: minimizing and maximizing $\mathcal{E}$

Minimizing $\mathcal{E}$; the sparsest cut problem

At the end of last lecture, Chris mentioned an interesting fact about minimizing $\mathcal{E}(f)$ (i.e. finding f that minimizes $\mathcal{E}$): It turns out that $\mathcal{E}(f) = 0$ is always achievable, but may not be what we're looking for. Finding f such that $\mathcal{E}(f) = 0$ does nothing that is not already easy to do. Consider a graph like this:



We can find f such that $\mathcal{E}(f) = 0$, but that is just a constant

On the other hand, if we could minimize nonzero $\mathcal{E}$, we could get something very useful for Divide and conquer algorithms

We can phrase this problem as follows (let $f = 1_S$, $S \subseteq V$)

$$\langle f, Lf \rangle = \mathcal{E}(f) = \Pr_{u \sim v} [u \in S, v \notin S]$$

$$vol(S) = \langle f, f \rangle = \mathop{E}_{u \sim \pi} [f(u)^2] = \Pr_{u \sim \pi} [u \in S]$$

And their ratio, then, is $\Pr_{u \sim v} [v \notin S \mid u \in S]$

We define this quantity as conductance, $\Phi(S)$

Sparsest cut problem: given G, find $S \subseteq V$ that minimizes $\Phi(S)$, keeping $vol(S) \leq 1/2$

This probelm is NP-hard and its approximability is quite open, and this will be the topic of future lectures.

Maximizing $\mathcal{E}$

We start by thinking, what kind of graph, and with which function, would have a high value of $\mathcal{E}(f)$.

[Note that we want to keep $\|f\|_2^2 = 1$, or we could just scale up the function and obtain arbitrarily large values of $\mathcal{E}(f)$.]

Bipartite graphs. Let $f = \begin{cases} 1 & \text{if } u \in V_1 \\ -1 & \text{if } u \in \overline{V_1} \end{cases}$, V_1 is one of the two sides of G.

In this case, whatever edge we pick, we're guaranteed to have $(f(u) - f(v))^2 = 4$, so $\mathcal{E}(f) = 2$. It turns out that this is the best we can achieve for any graph, which is a result that can be proved with some calculations.

Consider now the probelm of finding a maximizer function $\phi : V \to \mathbb{R}$ which maximizes $\mathcal{E}(f)$ with $\|f\|_2^2 \leq 1$. Note that $\|f\|_2^2 = 1$ is a closed and bounded (so, compact) set, and $(f(u) - f(v))^2$ is a continuous function, which means that a maximizer exists.

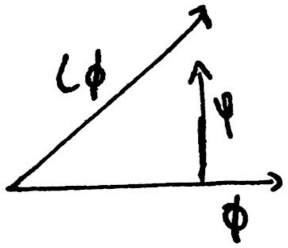
To make calculations easier, we assume $E[\phi] = 0$ and $Var[\phi] = 1$ (subtract mean to obtain if not the case; $Var[\phi] = \|\phi\|_2^2 - 0 = 1$

Most important claim of today:

$L\phi = \lambda\phi$ for some $\lambda \in \mathbb{R}$ (ϕ is an eigenvector of L)

The proof is by contradiction and uses the fact that ϕ can be represented as an n-dimensional vector, where n is the number of vertices in the graph.

Suppose that $L\phi$ and ϕ are not parallel:



Take φ as shown. Consider a vector $f = \phi + \varepsilon\varphi$ for some small $\varepsilon \in \mathbb{R}$

We have $\langle f, f\rangle = 1 + \varepsilon^2$ (it's a ~~rectan~~ right-angle triangle)

Also,

$$\begin{aligned}
\langle f, Lf\rangle &= \langle \phi + \varepsilon\varphi, L(\phi + \varepsilon\varphi)\rangle \\
&= \langle \phi + \varepsilon\varphi, L\phi + L\varepsilon\varphi\rangle \\
&= \langle \phi, L\phi + L\varepsilon\varphi\rangle + \langle \varepsilon\varphi, L\phi + L\varepsilon\varphi\rangle \\
&= \langle \phi, L\phi\rangle + 2\varepsilon\underbrace{\langle \varphi, L\phi\rangle}_{\substack{\neq 0 \\ (\text{not parallel})}} \pm O(\varepsilon^2)
\end{aligned}$$

So, for ε sufficiently small and of the same sign as $\langle \varphi, L\phi\rangle$, $\dfrac{\langle f, Lf\rangle}{\langle f, f\rangle} > \langle \phi, L\phi\rangle$

So, ϕ is not a maximizer

This is it; one more interesting thing to note is that once we find ϕ we can find another ϕ' in the same way, with the constraint $\phi' \perp \phi$. This way we can find n ~~eigen~~ eigen vectors and eigenvalues of L and use them as a basis to express function