Lecture 14

Sparsity; Random Walks

14.1 Sparsity

The *sparsity* is another natural measure of "disconnectedness" of a graph, which turns out to be very closely related to conductance.

Definition 14.1. For a d-regular graph G and $\emptyset \subsetneq S \subsetneq V$, the sparsity of S is defined as

$$\sigma(S) = \frac{\mathbf{E}_{(i,j)\in E}\,|1_S(i) - 1_S(j)|}{\mathbf{E}_{(i,j)\in V^2}\,|1_S(i) - 1_S(j)|} = \frac{\frac{1}{dn}|\partial S|}{\frac{1}{n^2}|S|\cdot|V\backslash S|} = \frac{|V||\partial S|}{d|S||V\backslash S|},$$

and so

$$\Phi(S) \le \sigma(S) \le 2\Phi(S).$$

Note that the definition of the sparsity lets us give a different, almost immediate proof of the "easy" side of Cheeger's inequality:

$$\begin{aligned}
\sigma(G) &= \min_S \sigma(S) \\
&= \min_{x\in\{0,1\}^n, x\neq \mathbf{0},\mathbf{1}} \frac{\frac{1}{dn}\sum_{(i,j)\in E}|x_i - x_j|^2}{\frac{1}{n^2}\sum_{(i,j)\in V^2}|x_i - x_j|^2} \\
&\ge \min_{x\in\mathbb{R}^n, x\neq\mathbf{0}, x\perp\mathbf{1}} \frac{2\sum_{\{i,j\}\in E}(x_i - x_j)^2}{\frac{d}{n}\sum_{(i,j)\in V^2}(x_i - x_j)^2} \\
&= \min_{x\in\mathbb{R}^n, x\neq\mathbf{0}, x\perp\mathbf{1}} \frac{2\,x\overline{L}x}{\frac{1}{n}\sum_{(i,j)\in V^2}(x_i - x_j)^2} \\
&= \min_{x\in\mathbb{R}^n, x\neq\mathbf{0}, x\perp\mathbf{1}} \frac{2\,x\overline{L}x}{\frac{1}{n}\sum_{i\in V}2nx_i^2 - 2\sum_{i,j}x_ix_j} \\
&= \min_{x\in\mathbb{R}^n, x\neq\mathbf{0}, x\perp\mathbf{1}} \frac{2\,x^T\overline{L}x}{2x^Tx} = \lambda_2,
\end{aligned}$$

where here the inequality follows since we are taking the minimum over a larger set (the constraint $x \perp \mathbf{1}$ is without loss of generality since the whole expression is invariant under translation of the vector x by an additive constant multiple of $\mathbf{1}$).

Using the above inequality $\sigma(G) \leq 2\Phi(G)$, we have re-proven the left-hand side of Cheeger's inequality, simply by observing that the second eigenvalue of $\overline{L}$ could be seen as a natural *relaxation* of the sparsity, itself very closely related to the conductance.

The second eigenvalue gives us a good approximation to the conductance in case λ_2 is not too small, say it is a constant. If however λ_2 goes to 0 with n, say $\lambda_2 \sim 1/n$, then the approximation can be very bad, it can be a multiplicative factor $\lambda_2^{-1/2} \sim \sqrt{n}$ off.

Here are two other relaxations that have been considered, and do much better. The first one is due to Leighton and Rao, and can be defined as

$$\mathrm{LR}(G) = \min_{\substack{w \in \mathbb{R}^{n\times n} \\ w_{ij} \geq 0, w_{ii}=0 \\ w_{ij} \leq w_{ik}+w_{kj}}} \frac{\sum_{(i,j)\in E} w_{ij}}{\frac{d}{n}\sum_{(i,j)\in V^2} w_{ij}}. \tag{14.1}$$

This can be interpreted as a minimization over all *semi-metrics*: distance measures $d(i,j) = w_{ij}$ on the graph that are always non-negative and satisfy the triangle inequality. An example such metric is $(i,j) \to w_{ij} = |x_i - x_j|$ (for any fixed vector x), but there are others. The advantage of allowing all semi-metrics is that $\mathrm{LR}(G)$ can be computed using a linear program. Moreover, Leighton and Rao showed that

$$O(\log n)\,\mathrm{LR}(G) \geq \sigma(G) \geq \mathrm{LR}(G),$$

thus we get a much tighter approximation to the sparsity than the one given by λ_2 in cases when the sparsity is small. An even tighter relaxation was introduced by Arora, Rao and Vazirani, who considered

$$\mathrm{ARV}(G) = \min_{\substack{w \in \mathbb{R}^{n\times n} \\ w_{ij} \geq 0, w_{ii}=0 \\ w_{ij}^2 \leq w_{ik}^2+w_{kj}^2}} \frac{\sum_{(i,j)\in E} w_{ij}}{\frac{d}{n}\sum_{(i,j)\in V^2} w_{ij}}. \tag{14.2}$$

This is the same as before, except now we require d^2, instead of d, to be a semi-metric. This is a stronger requirement, and it is still a relaxation because $d_{ij} = |x_i - x_j|$ satisfies both conditions. The optimization problem defining $\mathrm{ARV}(G)$ can be solved efficiently using semidefinite programming, and Arora, Rao and Vazirani showed that

$$O(\sqrt{\log n})\,\mathrm{ARV}(G) \geq \sigma(G) \geq \mathrm{ARV}(G).$$

It is open whether one can find a better approximation to $\sigma(G)$ in polynomial time. The best hardness results are Unique Games hardness for constant-factor approximations, and anything in-between is open!

14.2 Random walks on graphs

14.2.1 A motivating example

We start with a motivating application. The problem k-SAT is, given m clauses over n Boolean variables such that each clause is a disjunction of literals (a literal is a variable or its negation), is it possible to find an assignment to the variables that satisfies all clauses? For $k \geq 3$ the problem is NP-hard, but for $k = 2$ there are efficient algorithms.

Here is a simple candidate. Start with a random assignment to the variables. At each step, choose a clause that is not satisfied. Pick one of the two variables it acts on at random, and flip it. Repeat.

Is this going to work? And if so, how long will take? Here is the key idea. Suppose there exists a satisfying assignment, and fix it. Now consider the distance between the current assignment, maintained by the algorithm, and this satisfying assignment. This distance is an integer between 0 and n, and at each step it either increases or decreases by 1. If a clause is violated, at least one of the two variables involved must have a different value in the current assignment as in the satisfying assignment. With probability 1/2 we flip this variable, so that at each step with probability at least 1/2 we decrease the distance by 1. How many steps will it take to find the satisfying assignment? The answer is $O(n^2)$, and we'll see how to show this very easily once we've covered some of the basics of the analysis of random walks on arbitrary graphs. (Note the algorithm we just described is not the best, and it is possible to solve 2SAT in deterministic linear time...)

14.2.2 The random walk matrix

Let G be a undirected weighted graph with adjacency matrix A. Put $d_i = \sum_{j:\{i,j\}\in E} w_{ij}$ (we will always assume all weights to be non-negative). The natural random walk on G is to step from $i \to j$ with probability $\frac{w_{ij}}{d_i}$. Let $p^{(0)} \in \mathbb{R}_+^n$ be a distribution over vertices. One step of the random walk is that

$$p^{(0)} \to p^{(1)}, \quad \text{where} \quad p_j^{(1)} = \sum_{i:\{i,j\}\in E} p_i^{(0)} \frac{w_{ij}}{d_i},$$

which in matrix form can be written as $p^{(1)} = AD^{-1}p^{(0)}$. This version of the random walk has one major drawback, which is that it does not always converge: consider for instance a graph with a single edge, or more generally a bipartite graph; the walk started at a vertex on the left will continue hopping back and forth between left and right without ever converging. To overcome this issue it is customary to consider instead the *lazy* random walk: with probability 1/2, do not move, and with probability 1/2, do as before. The update rule is then

$$p^{(t)} = \left(\frac{\mathbb{I}}{2} + \frac{1}{2}AD^{-1}\right) p^{(t-1)},$$

which in matrix form is $p^{(t)} = Wp^{(t-1)}$ where the random walk matrix W is defined as

$$W = D^{1/2}\left(\mathbb{I} - \frac{1}{2}\left(\mathbb{I} - D^{-1/2}AD^{-1/2}\right)\right)D^{-1/2} = D^{1/2}\Big(\mathbb{I} - \frac{\overline{L}}{2}\Big)D^{-1/2}.$$

Definition 14.2. Let $G = (V, E)$ be an n-vertex weighted, undirected graph with weights $w_{ij} \geq 0$. Define the *lazy random walk* on G: if $p \in \mathbb{R}^n_+$ is a distribution on the vertices $V = \{1, ..., n\}$, one step of the random walk brings p to Wp where

$$W = \frac{1}{2}\mathbb{I} + \frac{1}{2}AD^{-1} = D^{1/2}\Big(\mathbb{I} - \frac{\overline{L}}{2}\Big)D^{-1/2},$$

and $\overline{L}$ is the normalized Laplacian associated with G.

Note that W is not symmetric, so it is not diagonalizable. Nevertheless you can check that if v is an eigenvector for $\overline{L}$ with eigenvalue λ, then $w = D^{1/2}v$ is a right eigenvector for W with eigenvalue $(1 - \lambda/2)$; thus W has n eigenvalues $w_i = 1 - \lambda_i/2$ that are directly related to those of the normalized Laplacian. This will let us transfer our understanding of the λ_i to derive convergence properties of the random walk.

14.2.3 Mixing Time

Let $p^{(0)}$ be any distribution on $V = \{1, \ldots, n\}$, and for every $t \geq 1$ define $p^{(t)} = Wp^{(t-1)}$ inductively as the distribution on vertices after t steps of the lazy random walk have been performed. The main question is: does $\lim_{t\to\infty} p^{(t)}$ exist, and if so, how fast is convergence?

Definition 14.3. A distribution π is a *stationary distribution* if $W\pi = \pi$.

By definition, any stationary distribution π is such that π is an eigenvector of W with eigenvalue 1, thus $D^{1/2}\pi$ is a right eigenvector of $\overline{L}$ with eigenvalue 0. But we know that if G is connected $\overline{L}$ always has a single non-degenerate eigenvalue equal to 0, and moreover the associated eigenvector $v_1 \propto (\sqrt{d_1}, \ldots, \sqrt{d_n})$ can be taken to have all its components non-negative. Thus there exists a unique stationary distribution $\pi \propto D^{1/2}(\sqrt{d_1}, \ldots, \sqrt{d_n}) \propto (d_1, \ldots d_n)$; normalizing the vector π so that it indeed corresponds to a distribution we obtain that $\pi = \frac{1}{d(V)}(d_1, \ldots d_n)$.

Definition 14.4. Given $\varepsilon > 0$, the *mixing time* is defined as

$$\tau_\varepsilon \;=\; \min\big\{t:\, \|W^t p - \pi\|_1 \leq \varepsilon\, \forall p \in \mathbb{R}^n_+ \text{ s.t.} \sum_i p_i = 1\big\},$$

where π is the stationary distribution, and $\|\cdot\|_1$ is the statistical distance. By convention we also let $\tau = \tau_{1/4}$.

Theorem 14.5. *For any connected, weighed, undirected graph G the lazy random walk mixes in time $\tau_\varepsilon = O(\frac{\log(n/\varepsilon)}{\lambda_2})$, where λ_2 is the second-smallest eigenvalue of $\overline{L}$.*

Proof. Take p any distribution on V. Then, $p = \sum_{i=1}^N \alpha_i D^{1/2} v_i$, where v_i are the eigenvectors of $\overline{L}$ and α_i some arbitrary coefficients. Here, $v_1 = d(V)^{1/2} D^{-1/2}\pi$, so $\alpha_1 = (D^{-1/2}p)^T v_1 = d(V)^{-1/2}$ since p is a normalized distribution. Then,

$$\begin{aligned} W^t p &= \sum \alpha_i W^t D^{1/2} v_i \\ &= \sum_i \alpha_i w_i^t D^{1/2} v_i \\ &= \pi + \sum_{i\geq 2} \alpha_i w_i^t D^{1/2} v_i. \end{aligned}$$

Therefore,

$$\begin{aligned} \|W^t p - \pi\|_1 &= \Big\| \sum_{i\geq 2} D^{1/2} w_i^t \alpha_i v_i \Big\|_1 \\ &\leq \sqrt{n} \Big\| \sum_{i\geq 2} \alpha_i w_i^t D^{1/2} v_i \Big\|_2 \\ &\leq \sqrt{n}\sqrt{d_{max}} \Big(\sum_{i\geq 2} \alpha_i^2 w_i^{2t} \Big)^{1/2} \\ &\leq \sqrt{n}\sqrt{n}\, \omega_2^t, \end{aligned}$$

where for the last step we bounded the maximum degree by n and used that $\sum_{i\geq 2} \alpha_i^2 \leq p^T D^{-1} p \leq \sum_i p_i^2 \leq 1$ since p is a distribution. The bound is $\leq \varepsilon$ for $t = \frac{\log(n/\varepsilon)}{\log(w_2)} = \frac{\log(n/\varepsilon)}{\log(1-\lambda_2/2)} = O(\frac{\log(n/\varepsilon)}{\lambda_2})$, as desired. □

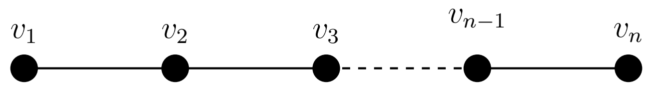
Example 14.6. The following simple test cases show that, aside from the $\log n$ factor we lost by switching from the Euclidean norm to the statistical distance, the bound provided by Theorem 14.5 can be very accurate.

- For the n-vertex clique K_n, we saw that the conductance $\phi(K_n) \approx 1/2$ and thus by Cheeger λ_2 is at least a constant. If we put all the probability at time 0 on a single vertex, it will take $\sim \log 1/\varepsilon$ steps before the lazy random walk spreads at least $1-\varepsilon$ probability to the other vertices, so $\tau_\varepsilon = \Omega(\log 1/\varepsilon)$.

- For the n-vertex path P_n, we saw $\lambda_2 \sim 1/n^2$, and you can check that $\tau_\varepsilon \geq n^2 \log(\frac{1}{\varepsilon})$: starting on the leftmost vertex, it takes $\sim n^2$ steps before we hit the rightmost vertex, and the $\log(1/\varepsilon)$ term is overhead before the walk becomes sufficiently close to uniform. Note this proves the bound on the convergence time for the randomized algorithm for 2-SAT we saw earlier.

- For the dumbbell graph $K_{n/3}$ — $P_{n/3}$ — $K_{n/3}$ (two $(n/3)$-cliques linked by a path of length $n/3$), the conductance is at most $\sim 1/n^2$ (cut in the middle), so by Cheeger $\lambda_2 = O(1/n^2)$, and it is possible to check $\lambda_2 = \Theta(n^2)$. Consider a random walk starting on the left-most vertex. Then, one step leads to a uniform distribution on the left $K_{n/3}$

clique. From there the probability to select the crossing edge is $1/n$, at which point it takes $\sim n^2$ steps to cross over to the other side. So the mixing time is of order n^3.



How bad can it get? From the definition we see that the conductance $\phi(G)$ is always at least $1/n^2$, so by Cheeger's inequality as long as G is connected we have $\lambda_2 = \Omega(n^{-4})$.

Claim 14.7. *Let G be a connected graph, λ_2 the second smallest eigenvalue of the normalized Laplacian and r the diameter of G. Then $\lambda_2 \geq \frac{2}{r(n-1)^2}$.*

Proof. For any two vertices u, v let $\overline{E}_{u,v}$ be the normalized Laplacian associated to the graph having a single edge $u \to v$, and $\overline{P}_{u,v}$ the normalized Laplacian for a path of length at most r from u to v. Then $\overline{L}_{u,v} \leq r\overline{P}_{u,v}$, as can be seen from the associated quadratic forms: $(x_u - x_v)^2 \leq r((x_u - x_{u_1})^2 + \cdots + (x_{u_{r-1}} - x_v)^2)$. We can cover G by paths of length at most r between any two of its vertices, thus $\overline{L}_{K_n} \leq r\binom{n}{2}\overline{L}_G$, where K_n is the clique on n vertices. Since the normalized Laplacian for K_n has second smallest eigenvalue $\frac{n}{n-1}$, we get the claimed bound on λ_2.

□