0368-4283: Space-Bounded Computation 20/3/2018 – Lecture 3

Random Walks over Undirected Graphs – II

Amnon Ta-Shma and Dean Doron

1 The Expander Mixing Lemma

We first show that an expander behaves like a random graph in the following sense: The number of edges between every two *large* subsets $S, T \subseteq [n]$ is close to what we would have expected in a random graph of average degree d, i.e., $\frac{d}{n}|S||T|$.

Lemma 1 (Expander Mixing Lemma)**.** *Let $G = (V = [n], E)$ be a d-regular graph and let $S, T \subseteq [n]$. Then,*

$$\left||E(S,T)| - \frac{d|S||T|}{n}\right| \leq \bar{\lambda}(G) \cdot d\sqrt{|S|(1-|S|/n)|T|(1-|T|/n)}$$

where $|E(S,T)|$ is the number of edges between the two sets.

Proof. Let A be the normalized adjacency matrix of G, so we have

$$|E(S,T)| = d \cdot \mathbf{1}_T^\dagger A \mathbf{1}_S.$$

We decompose $\mathbf{1}_S$ and $\mathbf{1}_T$ to a component parallel to $\mathbf{1}$ (the 1-eigenvector of A) and a perpendicular component. Write $\mathbf{1}_S = \frac{|S|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_S^\perp$ where

$$\mathbf{1}_S^\perp[i] = \begin{cases} n - |S| & i \in S \\ -|S| & i \notin S. \end{cases}$$

and notice that $\mathbf{1}_S^\perp \perp \mathbf{1}$. Similarly we write $\mathbf{1}_T = \frac{|T|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_T^\perp$. Then, using the fact that $A\mathbf{1} = \mathbf{1}$:

$$\begin{aligned} E(S,T) &= d \cdot \left(\frac{|T|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_T^\perp\right)^\dagger A \left(\frac{|S|}{n}\mathbf{1} + \frac{1}{n}\mathbf{1}_S^\perp\right) \\ &= d \cdot \frac{|S||T|}{n^2}\mathbf{1}^\dagger\mathbf{1} + \frac{1}{n^2}\left(\mathbf{1}_T^\perp\right)^\dagger A \mathbf{1}_S^\perp. \end{aligned}$$

As both $\mathbf{1}_S$ and $\mathbf{1}_S$ are perpendicular to the 1-eigenvector,

$$\left|\left(\mathbf{1}_T^\perp\right)^\dagger A \mathbf{1}_S^\perp\right| \leq \bar{\lambda}(G) \cdot \left\|\mathbf{1}_T^\perp\right\| \cdot \left\|\mathbf{1}_S^\perp\right\|.$$

A simple calculation shows that $\left\|\mathbf{1}_S^\perp\right\| = \sqrt{n|S|(n-|S|)}$ and likewise for $\left\|\mathbf{1}_T^\perp\right\|$, so overall

$$\begin{aligned} \left||E(S,T)| - \frac{d|S||T|}{n}\right| &\leq \bar{\lambda}(G) \cdot \frac{d}{n^2} \cdot \sqrt{n|S|(n-|S|)}\sqrt{n|T|(n-|T|)} \\ &= \bar{\lambda}(G) \cdot d \cdot \sqrt{|S|(1-|S|/n)}\sqrt{|T|(1-|T|/n)}, \end{aligned}$$

as desired. □

Corollary 2. *With respect to densities (dividing by dn), we can express the above result as*

$$\left| \Pr_{e=(i,j)\in E}[i \in S \wedge j \in T] - \rho(S)\rho(T) \right| \;\leq\; \bar{\lambda}(G) \cdot \sqrt{\rho(S)(1-\rho(S))\rho(T)(1-\rho(T))},$$

where for $A \subseteq B$ we denote $\rho(A) = |A|/|B|$.

1.1 Expanders have no small cuts

An often desirable feature of a graph is that no deletion of few edges can cause the graph to be disconnected. It is indeed the case with expanders. Given an undirected d-regular graph $G = (V, E)$ we define the *edge expansion* of a cut $(S, V \setminus S)$ as

$$h(S) \;=\; \frac{|E(S, V \setminus S)|}{d \cdot \min\{|S|, |V \setminus S|\}},$$

and we let $h(G) = \min_{S \subseteq V} h(S)$.

Exercise 3. *Let $G = (V, E)$ be a d-regular undirected graph over n vertices. Use the expander mixing lemma to prove $h(G) \geq \frac{1-\bar{\lambda}(G)}{2}$.*

We want to prove the stronger theorem:

Theorem 4. *Let $G = (V, E)$ be a d-regular undirected graph over n vertices and let λ_2 be the second eigenvalue of its normalized adjacency matrix A. Then, $h(G) \geq \frac{1-\lambda_2}{2}$.*

That is, for every $S \subseteq [V]$ of cardinality at most $\frac{n}{2}$, $|E(S, V \setminus S)| \geq \frac{d(1-\lambda_2)}{2}|S|$.

This theorem is one side of "Cheeger's Inequality". The other, harder, side is $h(G) \leq \sqrt{2(1-\lambda_2)}$ and we will not prove it. Morally, Cheeger's Inequality tells us that algebraic expansion and edge expansion are equivalent up to some loss in parameters.

Proof. We need to prove that $\lambda_2 \geq 1 - 2h(S)$ for every S with $|S| \leq \frac{n}{2}$. Equivalently, we can find a $v \perp \mathbf{1}$ for which $\frac{v^\dagger A v}{v^\dagger v} \geq 1 - 2h(S)$. Define a vector v such that:

$$v_i \;=\; \begin{cases} -n + |S| & i \in S \\ |S| & i \notin S. \end{cases}$$

First, notice that $v \perp \mathbf{1}$, as $\sum_i v_i = |S|(-n + |S|) + |S|(n - |S|) = 0$. Also, we have

$$v^\dagger v \;=\; |S|(-n + |S|)^2 + (n - |S|)|S|^2 \;=\; n|S|(n - |S|).$$

In our case,

$$\sum_{i,j} A[i,j](v_i - v_j)^2 \;=\; \frac{1}{d} \sum_{(i,j)\in E(S,\overline{S})} (|S| - (|S| - n))^2 \;=\; \frac{n^2}{d} 2|E(S, V \setminus S)|,$$

so $v^\dagger A v \;=\; v^\dagger v - \frac{n^2}{d}|E(S, V \setminus S)|$, and

$$\frac{v^\dagger A v}{v^\dagger v} \;=\; 1 - \frac{n^2|E(S, V \setminus S)|}{d \cdot v^\dagger v} \;=\; 1 - \frac{n|E(S, V \setminus S)|}{d \cdot |S|(n - |S|)} \;\geq\; 1 - \frac{2|E(S, V \setminus S)|}{d \cdot |S|} \;=\; 1 - 2h(S).$$

□

2 Density Samplers

Suppose we have some set $A \subseteq \{0,1\}^n$ and we want to (probabilistically) approximate its density $\rho(A) = |A|/2^n$. Suppose we are given oracle access to an indicator $I[x]$ which is 1 if and only if $x \in A$ and we want to approximate $\rho(A)$ by drawing $X_1, \ldots, X_k$ uniformly at random and outputting $\frac{1}{k}\sum_{i=1}^{k} I[X_i]$. Clearly, that naive approach requires kn random bits. How large should k be in order to get an additive error of ε with high probability (say at least $1-\delta$)? As $X_1, \ldots, X_k$ are independent, we can use Chernoff and get

$$\Pr\left[\left|\frac{1}{k}\sum_{i=1}^{k} I[X_i] - \rho(A)\right| > \varepsilon\right] \;\leq\; e^{-2k\varepsilon^2},$$

so $k = \frac{1}{2\varepsilon^2}\ln\frac{1}{\delta}$ is sufficient (and in a sense also necessary) and we use $O(\frac{n}{\varepsilon^2}\log\frac{1}{\delta})$ random bits.

The question we ask is whether we can re-use randomness and approximate the density well without using too many random bits.

2.1 Via random walks on expander

Let us first address the easier problem: We only want to hit a set $A \subseteq \{0,1\}^n$. A simple calculation shows that $k = \frac{1}{\rho(A)}\ln\frac{1}{\delta}$ is sufficient, and in the naive approach we therefore use $\frac{n}{\varepsilon}\ln\frac{1}{\delta}$ random bits.

Consider the following derandomization of that naive approach:

- Let $G = (V = \{0,1\}^n, E)$ be an expander with constant degree D and a constant $\bar{\lambda}(G) = \bar{\lambda} < 1$.
- Choose X_1 uniformly at random and take a random walk on G of length $T-1$ to obtain $X_2, \ldots, X_T$. Accept if and only if $I[X_i] = 1$ for some i.

Clearly, the probability that the algorithm errs is

$$\Pr\left[\bigwedge_{i=1}^{T}(X_i \notin A)\right]$$

and we want to estimate it. Note that if we choose G to be the complete clique over 2^n vertices than this is exactly the naive, random sampling, algorithm.

For the analysis, we need to prove that a random walk over an expander visits every large enough set with high probability (or, equivalently, is contained in a given set with very low probability).

Theorem 5. *Using the above notations,*

$$\Pr\left[\bigwedge_{i=1}^{T}(X_i \notin A)\right] \;\leq\; (1-\gamma\rho(A))^T,$$

where $\gamma = 1-\bar{\lambda}$ is the spectral gap of the expander.

In our case, we want $(1-\gamma\rho(A))^T < \delta$, so it is sufficient to take $T = \frac{1}{\gamma\rho(A)} \ln \frac{1}{\delta}$ (which is essentially the same as the truly random case), but we only use $n + \log D \cdot (T-1) = n + O(T)$ random coins, instead of the trivial nT.

Proof. The proof has two main components. First, we need to translate the condition $\bigwedge_{i=1}^{T}(X_i \in \bar{A})$ to an algebraic terminology, and then we analyze it.

The translation to algebraic terminology. Let M be the transition matrix of G and denote $|V| = 2^n = N$. Pick $X_1 \in V$ uniformly at random. That is, the initial distribution over the vertices is $u = \frac{1}{N}\mathbf{1}$. Define an $N \times N$ diagonal projection matrix B with $B[x,x] = 1$ if $x \in \bar{A}$ and 0 otherwise. In this terminology, $|\langle \mathbf{1}, Bu\rangle|$ is the probability a random element belongs to $\bar{A}$ (and so is β). $|\langle \mathbf{1}, BMBu\rangle|$ is the probability in a random walk of length two, both samples belong to $\bar{A}$. Similarly, $|\langle \mathbf{1}, (BM)^k Bu\rangle|$ is the probability that in a random walk of length $k+1$ the walk is confined to the set $\bar{A}$, i.e., all samples belong to $\bar{A}$.

Reducing the analysis to understanding a single step : As B is a projection, $B^2 = B$, and so $|\langle \mathbf{1}, (BM)^k Bu\rangle| = |\langle \mathbf{1}, (BMB)^k Bu\rangle|$. But, $|\langle \mathbf{1}, (BMB)^T Bu\rangle| \leq \left\|(BMB)^T\right\| \leq \|BMB\|^T$. The claim will follow by proving $\|BMB\| \leq 1 - \gamma\rho(A)$.

Claim 6 ([2])**.** *Let G be an undirected regular graph on n vertices and let M be its transition matrix. Then, $M = (1-\bar{\lambda})J + \bar{\lambda}E$ for some E with $\|E\| \leq 1$ and J that is the normalized all-ones matrix. I.e., M is a convex combination of J (that corresponds to a completely random walk) and E (that is some arbitrary error matrix).*

Proof. The first eigenvector of M is $\mathbf{1}$ (possibly normalized) with eigenvalue 1. $\mathbf{1}$ is also an eigenvector of J with eigenvalue 1. We conclude that $\mathbf{1}$ is a common eigenvector of M, J and E and with eigenvalue 1 for all of them (check!). What about vectors in the orthogonal complement? Let $W^{\perp}$ denote all vectors perpendicular to $\mathbf{1}$, i.e., all x such that $\langle x, \mathbf{1}\rangle = 0$. Then $Jx = 0$. Also, $W^{\perp}$ is invariant under M (why?) and thus $W^{\perp}$ is invariant also under E. Hence, to bound the norm of E, it is enough to limit our attention to $W^{\perp}$. For $v \in W^{\perp}$, $\|Ev\| = \frac{1}{\lambda}\|Mv\| \leq \frac{\bar{\lambda}}{\lambda}\|v\| = \|v\|$. Thus, $\|E\| \leq 1$. □

Now, let us express BMB in this decomposition. We get

$$BMB = B((1-\lambda)J + \lambda E)B \;=\; (1-\lambda)BJB + \lambda BEB.$$

The BJB part is the part corresponding to a true random walk step, the other part is "junk", and indeed we easily see that $\|BEB\| \leq \|B\|\,\|E\|\,\|B\| \leq 1$. Thus, we are now reduced to analyzing BJB, i.e., one true random walk step. For any $x \neq 0$, $x = \sum_i x_i e_i$. Then, $(BJBx)[i] = \frac{1}{N}\sum_{i\in\bar{A}} x_i$ if $i \in \bar{A}$ and 0 otherwise (check!). Thus, by Cauchy-Schwarz,

$$\|BJBx\| \;=\; \sqrt{|\bar{A}|\left(\frac{1}{N}\sum_{i\in\bar{A}} x_i\right)^2} \;=\; \sqrt{\frac{|\bar{A}|}{N^2}}\sum_{i\in\bar{A}} x_i \;\leq\; \sqrt{\frac{|\bar{A}|}{N^2}}\sqrt{|\bar{A}|}\,\|x\| = \frac{|\bar{A}|}{N} = \rho(\bar{A}).$$

Together,

$$\|BMB\| \;\leq\; \gamma\rho(\bar{A}) + 1 - \gamma = 1 - \gamma\rho(A).$$

□

The two-sided case (i.e., approximating the density) is along the same ideas, but a bit more complicated. The analysis may use the useful *expander Chernoff bound*.

Theorem 7 ([1])**.** *Let G be an undirected D-regular graph with $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$ and spectral gap $\gamma = 1 - \bar{\lambda}(G)$ and let $f_i : V \to [0,1]$ for $i \in [T]$. Take a random walk $v_1, \ldots, v_T$ and let X_i be the random variable $f_i(v_i)$. Denote $\mu_i = \mathbb{E}[X_i]$ and $\bar{\mu} = \frac{1}{T}\sum_i \mu_i$. Then,*

$$\Pr\left[\left|\frac{1}{T}\sum_i X_i - \bar{\mu}\right| \geq \varepsilon\right] \;\leq\; 2e^{-\frac{1}{4}\gamma\varepsilon^2 T}.$$

References

[1] Alexander D Healy. Randomness-efficient sampling within nc. *Computational Complexity*, 17(1):3–37, 2008.

[2] Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 436–447. Springer, 2005.