CS369E: Expanders	April 18, 2005

Lecture 3: Random Walks

Lecturer: Prahladh Harsha *Scribe: David Arthur*

3.1 Introduction

Consider an undirected graph G. A random walk of length l starting at the vertex u is a sequence of vertices $u = v_0, v_1, v_2, \cdots, v_l$, where each v_i is chosen to be a random neighbor of v_{i-1} for all $i > 0$. One considers the distribution of v_i for $i \leq l$.

Intuitively, a random walk can be thought of as choosing a globally random vertex on a graph using only local choices. This is something that people actually do in practice. For example, one might shuffle a deck of cards by repeatedly moving the top card to a random position in the deck. We can model all orderings of the deck as the vertices of a graph with edges corresponding to the operation described above. This process of repeatedly moving the top card can then be thought of as a random walk that provides a more convenient way of shuffling a deck than explicitly choosing 1 of 52! possible orderings.

Traditionally, random walks were considered on infinite graphs, and the following result is typical of what was studied.

Theorem 3.1 (Polya, 1921)**.** *Consider a random walk on an infinite D-dimensional grid. If $D = 2$, then with probability 1, the walk returns to the starting point an infinite number of times. If $D > 2$, then with probability 1, the walk returns to the starting point only a finite number of times.*

For the purpose of this lecture, we will consider random walks on finite undirected graphs and even more specifically, d-regular undirected graphs. Refer [Lov] for an excellent survey on Random Walks on Graphs.

Several questions will motivate this lecture. Let $\pi_0 = \pi$ be the starting distribution on the graph G (mostly, we will consider cases when π is concentrated on a single vertex). Let π_i denote the probability distribution of v_i for a random walk beginning at the starting distribution π_0. Since we are interested in the ability of random walks to generate a globally random vertex, it is natural to consider π_i as i gets large.

Question 3.2. *For which π_0, does π_i converge to some stationary distribution as i approaches infinity? What is the stationary distribution that it converge to?*

If we let $A = A(G)$ denote the normalized adjacency matrix of G, then it is easy to check that $\pi_{i+1} = A\pi_i$ for all i. Thus, a given distribution x is a stationary distribution for some starting distribution π_0 only if $x = Ax$. This is equivalent to stating x is an eigenvector of A with corresponding eigenvalue 1. As noted in previously lectures, the uniform distribution $u = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ has this property, but there could be other possible stationary distributions as well. If the graph is disconnected, then there exist multiple (independent) eigenvectors with eigenvalue 1. In fact, one can show the following.

Lemma 3.3. *The multiplicity of the largest eigenvalue (i.e.,. 1) in $A(G)$ is equal to the number of connected components in G.*

In particular, if G is connected, the only possible stationary distribution is u. Thus, the stationary distribution (if it exists) is independent of the starting distribution π_0. This largely answers the second part of Question 1.

Before answering the first part of Question 1, we consider a related question.

Question 3.4. *If π_i converges to a stationary distribution, how fast does it converge?*

We can also recast this in terms of mixing time, described below.

Definition 3.5 (Mixing Time)**.** *The "mixing time" of a graph G with n vertices is the minimum l such that for all starting distributions π*

$$\|A^l \pi - u\|_\infty < \frac{1}{2n}. \tag{1}$$

We will define the $\|\cdot\|_\infty$-norm shortly.

The $\frac{1}{2\pi}$ is largely arbitrary, but this value will prove convenient. If we take the mixing time to be infinity for graphs where no l satisfies (1), answering our remaining questions is equivalent to understanding mixing time.

Finally, we present two concepts related to mixing time, which are interesting in their own right.

Definition 3.6. *(Hitting Time) For a graph G, let $H(u, v)$ denote the expected number of steps a random walk beginning at u must take before reaching v. Define the "hitting time" of G by $H(G) = \max_{u,v} H(u, v)$.*

Definition 3.7. *(Cover Time) For a graph G, let C_u denote the expected number of steps a random walk beginning at u must take before reaching every other vertex at least once. Define the "cover time" of G by $C(G) = \max_u C_u$.*

It easily follows from the definitions that $H(G) \le C(G) \le n \cdot H(G)$. The latter inequality can be tightened (using the coupon-collectors' problem) to show that $C(G) \le O(\log n) \cdot H(G)$.

3.2 Bounding the Mixing Time

As discussed above, the distribution of random walks on disconnected graphs need never converge to u. Bipartite graphs are similarly problematic. Specifically, if a random walk begins at a vertex in one part, it will always be in that part after an even number of steps, and it will always be in the other part after an odd number of steps. Thus, π_l can never converge to u on a bipartite graph.

As with Lemma 3.3, we can characterize this failure case in terms of eigenvalues of A.

Lemma 3.8. *G is bipartite iff -1 is an eigenvalue of $A(G)$.*

For example, if G is bipartite, consider the vector v with a value of 1 at all vertices in one part and a value of -1 at all other vertices. One can check this is an eigenvector of $A(G)$ with eigenvalue -1.

Now, fix a graph G with n vertices and consider the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ of $A(G)$. Without loss of generality we may assume $\lambda_1 = 1$ and $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Furthermore, let $\lambda = |\lambda_2|$. By Lemmas 3.3 and 3.8, we know $\lambda = 1$ iff G is either bipartite or disconnected. Therefore, G has infinite mixing time if $\lambda = 1$. We now show that conversely, if $\lambda < 1$ then G has finite mixing time.

Theorem 3.9. *If G is a connected, d-regular, non-bipartite graph on n vertices, then $\lambda < 1$ and G has mixing time $O\left(\frac{\log n}{1-\lambda}\right)$.*

We review the l^1, l^2 and l^∞ norms before proceeding with the proof.

Definition 3.10. *If $v = (v_1, v_2, \cdots, v_m)$ is an arbitrary vector, define*

$$\begin{aligned} \|v\|_\infty &= \max_i |v_i|, \\ \|v\| = \|v\|_2 &= \sqrt{\sum_i v_i^2}, \text{ and} \\ \|v\|_1 &= \sum_i |v_i|. \end{aligned}$$

Fact 3.11. $\|v\|_\infty \leq \|v\| \leq \|v\|_1 \leq \sqrt{n}\ \|v\|$.

The first two inequalities here can easily be verified and the third follows from the Cauchy-Schwarz inequality. Furthermore, these are all norms, which implies that they satisfy the triangle inequality.

Proof of Theorem 3.9: Note that $A(G)$ is a real, symmetric matrix, which implies that it has n orthonormal eigenvectors $u = v_1, v_2, \cdots, v_n$. Let π denote any (starting) probability distribution on the vertices of G. Then, we can decompose π uniquely as $\sum_{i=1}^n \pi_i$ where π_i is a constant multiple of v_i. Furthermore, as discussed in the previous lecture, the fact that π is a probability distribution guarantees $\pi_1 = u$. Now,

$$\begin{aligned} \|A\pi - u\|^2 &= \|Au + A\pi_2 + A\pi_3 + \cdots + A\pi_n - u\|^2 \\ &= \|\lambda_2\pi_2 + \lambda_3\pi_3 + \cdots + \lambda_n\pi_n\|^2 \text{ since } Au = u \\ &= \lambda_2^2\|\pi_2\|^2 + \lambda_3^2\|\pi_3\|^2 + \cdots + \lambda_n^2\|\pi_n\|^2 \text{ by the Pythagorean theorem} \\ &\leq \lambda^2\left(\|\pi_2\|^2 + \|\pi_3\|^2 + \ldots + \|\pi_n\|^2\right) \\ &= \lambda^2\|\pi_2 + \pi_3 + \cdots + \pi_n\|^2 \text{ again by the Pythagorean theorem} \\ &= \lambda^2\|\pi - u\|^2. \end{aligned}$$

Thus, each step of the random walk decreases the l^2-distance of the distribution on the vertices to the uniform distance by a factor of at least λ. Therefore, $\|A^l\pi - u\| \leq \lambda^l\|\pi - u\|$

for all $l \geq 0$. It follows that

$$\begin{aligned}
\|A^l\pi - u\|_\infty &\leq \|A^l\pi - u\| \\
&\leq \lambda^l\|\pi - u\| \\
&< \lambda^l\|\pi\| \text{ since } \pi - u \text{ and } u \text{ are orthogonal} \\
&\leq \lambda^l\|\pi\|_1 \\
&= \lambda^l.
\end{aligned}$$

It follows that $\|A^l\pi - u\|_\infty < \frac{1}{2n}$ when $l = O\left(\frac{\log n}{\log \frac{1}{\lambda}}\right) \approx O\left(\frac{\log n}{1-\lambda}\right)$. To see this last step, note that $\log(1+x) = 1 - \frac{1}{1+x} + O\left(\frac{1}{(1+x)^2}\right)$ by taking the Taylor expansion of both sides. $\square$

3.2.1 Bounding the Spectral Gap $1 - \lambda$

Since Theorem 3.9 depends so heavily on $1 - \lambda$, it is natural to try to bound this quantity for various graphs G. We have already seen that $1 - \lambda = \Omega(1)$ for expanders. We now consider its value for other graphs.

Theorem 3.12. *If G is a connected, d-regular, non-bipartite graph on n vertices, then $1 - \lambda \geq \frac{1}{dn^2}$.*

We will prove the result this theorem only for the case where G has only non-negative eigenvalues.

Proof. As discussed in the previous lecture, we can obtain the following characterization of the spectral gap.

$$\begin{aligned}
\lambda &= \max_{x\perp u, \|x\|=1} \langle Ax, x\rangle \text{ since } G \text{ has only non-negative eigen values} \\
&= \max_{x\perp u, \|x\|=1} \frac{1}{d} \sum_{(u,v)\in E} 2x_u x_v \\
&= \max_{x\perp u, \|x\|=1} \frac{1}{d} \sum_{(u,v)\in E} \left(x_u^2 + x_v^2 - (x_u - x_v)^2\right) \\
&= \max_{x\perp u, \|x\|=1} \frac{1}{d} \sum_{(u,v)\in E} (x_u^2 + x_v^2 - (x_u - x_v)^2 \\
&= \max_{x\perp u, \|x\|=1} \frac{1}{d} \left(d \cdot x_u^2 - \sum_{(u,v)\in E} (x_u - x_v)^2\right)
\end{aligned}$$

Hence, the spectral gap $1 - \lambda$ is given by

$$1 - \lambda = \min_{x\perp u, \|x\|=1} \frac{1}{d} \sum_{(u,v)\in E} (x_u - x_v)^2.$$

Then there exists x with $x \perp u$ and $\|x\| = 1$ so that $1 - \lambda = \frac{1}{d}\sum_{(u,v)\in E}(x_u - x_v)^2$. Since $\|x\| = 1$, there exists v' for which $|x_{v'}| \geq \frac{1}{\sqrt{n}}$. However, since $x \perp u$, we know

$\sum x_v = 0$, and hence there exists v'' for which $x_{v'}$ and $x_{v''}$ have different signs. It follows that $|x_{v'} - x_{v''}| \geq \frac{1}{\sqrt{n}}$.

Now, G is connected so there exists some shortest path $v_0(= v'), v_1, \cdots, v_k(= v'')$ from v' to v''. The triangle inequality now implies that

$$\sum_{i=0}^{k-1} |x_{v_i} - x_{v_{i+1}}| \geq |x_{v_0} - x_{v_k}| \geq \frac{1}{\sqrt{n}}.$$

Therefore,

$$\begin{aligned}
1 - \lambda &= \frac{1}{d} \sum_{(u,v)\in E} (x_u - x_v)^2 \\
&\geq \frac{1}{d} \sum_{i=0}^{k-1} (x_{v_i} - x_{v_{i+1}})^2 \\
&\geq \frac{1}{dk} \left(\sum_{i=0}^{k-1} |x_{v_i} - x_{v_{i+1}}| \right)^2 \quad \text{by Fact 3.11} \\
&\geq \frac{1}{dkn} \geq \frac{1}{dn^2}.
\end{aligned}$$

As mentioned earlier, our proof only applies if G has no negative eigenvalues. In the general case, one can apply similar analysis to G^2 to bound $1 - \lambda^2$. This gives us a weaker bound $1 - \lambda^2 \geq \frac{1}{d^2 n^2}$ and hence $1 - \lambda \geq \frac{1}{\mathrm{poly}(n,d)}$. In fact the same bound of $\frac{1}{dn^2}$ can be obtained for the general case using a tighter analysis. $\square$

Note that Theorems 3.9 and 3.12 imply the mixing time of any d-regular, connected, non-bipartite graph on n vertices is $O(dn^2 \log n)$.

3.2.2 A Combinatorial Notion of Mixing

So far we have related the mixing time to the spectral gap $1 - \lambda$. There is a combinatorial parameter of the graph that relates more directly to the mixing time. Suppose G is made up of two cliques joined by just a few edges. This creates a bottleneck that should intuitively limit the mixing time of G. To characterize this, one defines the following. For any set of vertices S, let $\overline{S}$ denote $V(G) - S$, and let $|E(S, \overline{S})|$ denote the number of edges between S and $\overline{S}$. Then, define

$$\begin{aligned}
\Phi(S) &= \frac{|E(S, \overline{S})|}{|S|}, \text{ and} \\
\Phi(G) &= \min_{S:|S|\leq \frac{n}{2}} \Phi(S).
\end{aligned}$$

$\Phi(G)$ is called the edge-expansion of the graph G. $\Phi(G)$ has a direct relation to the mixing time. The edge expansion $\Phi(G)$ is related to the spectral gap as follows:

Theorem 3.13. *Let G be a d-regular graph. Then,*

$$\frac{d(1 - \lambda)}{2} \leq \Phi(G) \leq d\sqrt{2(1 - \lambda)}.$$

3.3 Applications

3.3.1 Undirected s-t Connectivity

Let G be a d-regular, connected, non-bipartite graph with n vertices and mixing time l. Consider a random walk beginning at some vertex s. Then for any vertex t and any integer $l' \geq l$, we know that

$$\text{Prob}\left[\text{Random walk is at vertex } t \text{ after } l' \text{ steps}\right] = \left(A^{l'}\pi\right)_t \geq \frac{1}{n} - \frac{1}{2n} = \frac{1}{2n}.$$

Therefore, a random walk of length $2nl'$ will reach t with constant probability. We know that the mixing time for a d-regular connected bipartite graph is $O(dn^2 \log n)$. This suggests an algorithm for s-t connectivity. Take a random walk of length $\Theta(dn^3 \log n)$ starting at s. If the walk reaches t, then s and t are connected. Otherwise, s and t are disconnected with high probability. Note this runs in polynomial time and uses only $\log n$ space to track the current vertex.

Now, as stated, our argument relies on G being d-regular, connected and and non-bipartite to find a path from s to t. We can remove these assumptions as follows.

1. *(Connected)* Restrict to the connected component of G containing s.
2. *(Bipartite)* Add a self loop at each vertex. This does not affect whether s and t are connected and it causes the graph to be no longer bipartite.
3. *(Regular)* Each vertex of degree $D > 3$ can be replaced by D vertices of degree 3 in a cycle to make the graph 3-regular. Also, non-regular graphs do in fact mix already and the same algorithm works. We have not shown this, however.

We summarize all this as follows.

Theorem 3.14 (Undirected Connectivity is in RL, [AKL$^+$]). *There is a polynomial time, log space Monte Carlo algorithm for s-t connectivity in undirected graphs.*

Recently, Reingold obtained a deterministic algorithm for undirected s-t connectivity (also using expanders)[Rei]. We will cover it in future lectures.

3.3.2 Hitting Time and Cover Time

As in the previous section, let l be the mixing time of a graph G on n vertices. Consider a random walk beginning at an arbitrary vertex s. Then, recall that for $l' \geq l$, it is true for any t that

$$\text{Prob}\left[\text{Random walk is at vertex } t \text{ after } l' \text{ steps}\right] \geq \frac{1}{2n}.$$

It follows that the expected time for a random walk to reach t is at most $2n \cdot l$, so the hitting time of G is at most $2n \cdot l$, which is polynomial in n. Similarly, after $2n^2 \cdot l$ steps, the walk will have reached each vertex with high probability. Thus, the cover time of G is also polynomial in n.

Tighter results are known for both the hitting time and cover time, as summarized in the following theorems.

Theorem 3.15. *Let G be an arbitrary undirected graph on n vertices. Then,*

1. [BW] $H(G) \leq \frac{4}{27}n^3 - \frac{1}{9}n^2 + O(n)$, *and*
2. [Fei1] $C(G) \leq \left(\frac{4}{27} + o(1)\right) n^3$.

Theorem 3.16 ([Fei2])**.** *Let G be a d-regular undirected graph on n vertices. Then, $C(G) \leq 2n^2$.*

The bound given in Theorem 3.16 is also known to be tight.

3.3.3 Universal Traversal Sequences

Universal traversal sequences (UTS) were originally defined by Cook, and later suggested by [AKL+] as a possible means to derandomize Theorem 3.14.

Consider a sequence $\mathcal{S} \in \{1, 2, \cdots, d\}^{l(n)}$ for some function l. Now, consider a d-regular undirected graph G where all the edges adjacent to each vertex have been labeled with distinct integers from 1 to d. These labellings need not be consistent in the sense that one edge might have two different labels assigned to it by two different vertices. For each vertex s, we can now use $\mathcal{S}$ to define a walk on G starting at s. Specifically, if we are at vertex v after i steps, we go to vertex v' where $\overline{vv'}$ is the edge labeled $\mathcal{S}_i$ by v. We say $\mathcal{S}$ is a "universal traversal sequence" if this walk traverses every vertex of the graph for every possible beginning vertex on every labeling of every n vertex d-regular graph. We are interested in constructing UTS of short length (i.e., $l(n) = \text{poly}(n)$).

It is not obvious how to construct a UTS or even whether one exists. However, a sufficiently lengthy random string will be a UTS with high probability.

Theorem 3.17. *Suppose every d-regular, n vertex graph has cover time at most C (note $C \leq 2n^2$ by Theorem 3.16). Then, there exists a UTS for d-regular, n vertex graphs of length at most $4ndC \log n$.*

Proof. Choose $\mathcal{S}$ uniformly at random from $\{1, 2, \cdots, d\}^{4ndC \log n}$.

Let G be a random labeled d-regular graph on n vertices and u be a random starting vertex in G. The expected time for the random walk $\mathcal{S}$ to cover every vertex of G is at most C, so by Markov's inequality, a random walk will cover every vertex within $2C$ steps with probability at least $\frac{1}{2}$. Since $\mathcal{S}$ can be decomposed into $2nd \log n$ disjoint, and hence independent, random sequences of length $2C$, it follows that $\mathcal{S}$ will cover all the vertices of G with probability at least $1 - \frac{1}{2^{2nd \log n}} = 1 - \frac{1}{n^{2nd}}$. Hence,

$$\operatorname*{Prob}_{G,u,\mathcal{S}} [\mathcal{S} \text{ covers all vertices of } G \text{ starting at } u] \geq 1 - \frac{1}{n^{2nd}}$$

.

Let N denote the number of ways of choosing a labeled d-regular graph G and a starting vertex u. For each vertex and each label, we can choose an adjacent vertex, and we can also choose one distinguished starting vertex. Thus, $N \leq n \cdot n^{nd}$.

Now, the probability that there exists one configuration (i.e., a labeled graph G and a starting vertex u) where $\mathcal{S}$ does not cover every vertex is at most $N \cdot \frac{1}{n^{2nd}} \leq \frac{n}{n^{nd}} = o(1)$. Thus, with high probability, $\mathcal{S}$ is a UTS. The result follows. $\square$

To use a UTS to derandomize Theorem 3.14, this result is insufficient. For that purpose, we would need to construct a UTS deterministically using logarithmic space. In general, such a construction has not been found. However, we know how to construct UTS if we relax either the restriction that the length of the UTS must be polynomial or the restriction that the UTS holds good for all d-regular graphs on n vertices. The following is a flavor of such results under such relaxations.

Theorem 3.18.
- [Ist] *A UTS (of polynomial length) can be constructed deterministically in $O(\log n)$ space for cycles.*
- [HW] *A UTS (of polynomial length) can be constructed deterministically in $O(\log n)$ space for d-regular expanders which are "consistently labelled"*[1]

3. [Rei] *A UTS (of polynomial length) can be constructed deterministically in $O(\log n)$ space for d-regular undirected graphs which are consistently labelled.*
4. [Nis] *A UTS, of length $O(n^{\log n})$, can be constructed deterministically in $O(\log^2 n)$ space for general d-regular graphs with general labellings.*

3.3.4 Random Walks on Expanders

Finally, we apply Theorem 3.9 to the case of expanders. We know the mixing time is $O(\frac{\log n}{1-\lambda})$ but λ is bounded by a constant for expanders. Thus, the mixing time on an expander is just $O(\log n)$, which is the best possible (up to constant factors) since the diameter of an expander is $O(\log n)$. It follows that expanders are rapidly mixing, which will allow for some applications to derandomization in the next lecture.

References

[AKL+] Romas Aleliunas, Richard M. Karp, Richard J. Lipton, László Lovász, Charles Rackoff: "Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems". FOCS 1979: 218-223

[BW] Graham Brightwell, Peter Winkler: "Maximum Hitting Time for Random Walks on Graphs". Random Struct. Algorithms 1(3): 263-276 (1990)

[Fei1] Uriel Feige: "A Tight Upper Bound on the Cover Time for Random Walks on Graphs". Random Struct. Algorithms 6(1): 51-54 (1995)

[Fei2] Uriel Feige: "Collecting Coupons on Trees, and the Cover Time of Random Walks". Computational Complexity 6(4): 341-356 (1997)

[HW] Shlomo Hoory, Avi Wigderson: "Universal Traversal Sequences for Expander Graphs". Inf. Process. Lett. 46(2): 67-69 (1993)

[1] A labeling is said to be be consistent if for any two edges $e_1 = (u_1, v)$ and $e_2 = (u_2, v)$ incident on the same vertex v, it is the case that the label of e_1 wrt u_1 is different from that of e_2 wrt u_2.

[Ist] Sorin Istrail: "Polynomial Universal Traversing Sequences for Cycles Are Constructible (Extended Abstract)". STOC 1988: 491-503

[Lov] Lázló Lovász: "Random Walks on Graphs: A Survey", Combinatorics, Paul Erdös is Eighty, Vol. 2 Jänos Bolyai Mathematical Society, Budapest, 1996, 353–398

[Nis] Noam Nisan: "Pseudorandom generators for space-bounded computation". Combinatorica 12(4): 449-461 (1992)

[Rei] Omer Reingold: "Undirected ST-Connectivity in Log-Space", to appear in STOC 2005 (Also see ECCC 2005).