6.842 Randomness and Computation March 5, 2012

Lecture 8

Lecturer: Ronitt Rubinfeld *Scribe: Alex Arkhipov*

This lecture, we will use Markov chains and random walks on graphs to solve the undirected *st*-connectivity problem in randomized logspace. We will also develop ideas in linear algebra related to random walks.

1 Last Class

Last time, we started talking about Markov chains and random walks. A **Markov chain** is a memoryless random process in which one transitions from state to state without regard for the prior history. A **random walk** is a special kind of Markov chain associated to a graph in which the state is the current node and you transition by moving along an out-edge of the current node uniformly at random.

We may represent a Markov chain as a **transition matrix** where P where each entry $P_{i,j}$ is the probability of moving from state i to state j. So, if the state at the current step is i, then i^{th} row lists the distribution of states in the next step.

$$\text{Row } i \begin{bmatrix} & & & & \\ P_{i,1} & P_{i,2} & \dots & P_{i,n-1} & P_{i,n-1} \\ & & & & \end{bmatrix}$$

This obervation that the rows represent probability distributions implies that the rows sum to one, and that the entries are non-negative. Such a matrix is called **stochastic**; all transition matrices of Markov chains are stochastic. If the columns also sum to one, we say the Markov chain is **doubly stochastic**. One example of a doubly stochastic Markov chain is a random walk on a d-regular directed (or undirected) graph. This follows because each row distribution is uniform over the the d out-neighbors of that vertex, and so each column also has d entries of $1/d$ of the corresponding d in-neighbors.

$$\text{Vertex } i \begin{bmatrix} & 1/d & & & \\ & 0 & & & \\ 0 & 1/d & 1/d & 1/d & 0 \\ & 0 & & & \\ & 1/d & & & \end{bmatrix}$$

If the vector Π is a probability distribution over states, performing one step of the Markov chain yields the distribution given by ΠP. A **stationary distribution** Π is a fixed point of the Markov chain, meaning it satisfies $\Pi P = \Pi$. Restated, it is a distribution over states so that if one picks a random state as per that distribution and transitions it randomly as per the Markov chain, the resulting distribution over states is the same. Observe that for a doubly stochastic Markov chain, the uniform distribution is stationary.

2 Stationary Distributions of Markov Chains

With stationary distributions defined, a natural question us whether every Markov chain has a stationary distribution, and whether it is unique. It turns out that these both hold for Markov

chains satisfying certain "niceness" conditions.

Throughout this lecture, we assume Markov chains have only a finite number of states and which are time-homogenous, meaning that the transition matrix is the same at each step.

Definition 1. A Markov chain is **irreducible** if there is some value n so that for any states x, y, if you start at x and take n steps, there is a positive probability you end up at y. Equivalently, if the Markov chain has transition matrix P, then all entries of P^n are strictly positive.

We can think of irreducibility of a Markov chain as a property of the graph that marks which states can transition to which state.

Definition 2. To a Markov chain M, associate a directed reachability graph G_M whose nodes are states, with a directed edge for each transition that occurs with positive probability. A Markov chain M is *ergodic* if and only if

1. The graph G_M is **irreducible**: For every large enough n, there is a positive probability path of length n between any two states.
2. The graph G_M **aperiodic**: The GCD of the lengths of positive probability cycles of G_M is 1.

The strongly-connectedness condition means that no matter which state you start in, you eventually reach every other state with positive probability. The aperiodicity condition rules out, say, a random walk on a bipartite graph where you necesarrily alternate between the two halves on odd and even steps.

Together, these conditions suffice to guarantee that there's a unique stationary distribution.

Theorem 3. *Any ergodic Markov chain has a unique stationary distribution.*

3 Mixing in Random Walks

One question about random walks is how long it takes for the walk to diffuse throughout the graph. We define multiple metrics to quantify these times. Although we will give these definition for random walks, they apply in general to Markov chains.

Definition 4. The **hitting time** $h_{s,t}$ from a vertex s to a vertex t on a graph G is the expected number of steps for a random walk on G starting at s to reach t. In the case where $s = t$, this is the **return time.**

One may wonder whether the return time of every node is necessarily even finite. The return time is in fact always finite for ergodic random walks, and can be computed directly from the stationary state in the case of ergodic random walks.

Theorem 5. *If a random walk is ergodic, then $h_{s,s} = 1/\Pi_s$, where Π_s is the probability of s in the stationary distribution, which is nonzero.*

Note that nodes with higher weight in the stationary distribution have shorter return times. This theorem is not true for non-ergodic random walks, such as a walk on a line where you always go right until you reach the end.

The hitting time tells us the expected time to reach a given node. What if we instead want to visit every node?

Definition 6. The **cover time** $\mathcal{C}_u(G)$ of a graph starting at a node u is the expected number of steps needed for a random walk starting at u to visit every node. When no node is specified, the cover time $\mathcal{C}(G)$ is the maximum cover time over all start nodes u.

To get a sense of cover times, let's look at some graphs.

Figure 1: Lollipop graph with $n = 10$

Example 7. If the graph is K_n^*, the complete graph with self loops on n vertices, then every step produces a uniformly random vertex. This is equivalent to the coupon collector's problem, in which one picks a random coupon uniformly from n coupons with replacement until one has collected at least one of each coupon. By a standard result, $\theta(n \ln n)$ coupons are needed, so $\mathcal{C}_u(K_n^*) = \theta(n \ln n)$.

Example 8. Consider the line graph L_n on n vertices. Here, a random walk is brownian motion in one dimension, and visiting all the vertices is equivalent to visting the left and right one. Since after n steps, one expects to move distance $\theta(\sqrt{n})$, you might guess that the cover time is $\theta(n^2)$. Indeed, $\mathcal{C}_u(L_n) = \theta(n^2)$. You might like to think about which start positions have the highest cover times.

Example 9. This example illustrates that the cover time may strongly depend on the start node. Define the lollipop graph G to be a path of length $n/2$ attached to a complete graph with self-loops on $n/2$ vertices.

Suppose we start at the far node of the line, which we'll call u. One expects to take $\theta(n^2)$ steps walking to the end of the line to reach the connected cluster, then $\theta(n \ln n)$ steps visiting all the nodes of the cluster. This ignores the fact that one might wander off the cluster to the line with low probability, but this turns out not to waste too many steps, as we usually quickly wander back into the cluster. In fact, $\mathcal{C}_u(G) = \theta(n^2)$.

One the other hand, suppose one starts at a node v in the cluster. From here, it's not easy to get to the line: One only gets to the connecting node every n steps on average, and the probability of taking the line edge is $1/(n+1)$. Even then, a walk on the line is likely to wander back to the cluster before it reaches node u at other end: Most random walks on a line starting from one end return to that end many times before reaching the other end. Overall, $\mathcal{C}_u(G) = \theta(n^3)$ steps are needed on average.

Overall, the cover time of the graph is $\mathcal{C}(G) = \theta(n^3)$

In fact, the cover time for lollipop graph is asymptotically the worst possible.

Theorem 10. *If G is an undirected, nonbipartite graph with n nodes and m edges, then*

$$C(G) \leq 2m(n-1) \leq n^3$$

Corollary 11. *The hitting time for undirected, nonbipartite graph is bounded by $h_{s,t} \leq n^3$.*

Sketch of Proof. We'll give only the idea of the proof. The trick is to look at the line graph G', which is constructed by taking edges of G as nodes and connecting ones that share a common vertex. Because G' turns out to be doubly stochastic, its stationary distribution is uniform. The expected return time of G' from any node is the inverse of its weight in the stationary distribution, and therefore equals m. This can be used to bound the time required to hit each edge of a walk that is defined by a spanning tree of the graph. Once the edges of the spanning tree have all been hit, the random walk has hit all nodes of the graph. □

Note that these results are definitely not true for directed graphs – these can have exponentially large cover times. For example, consider a graph where the nodes are in a line, such that each node i has two outgoing edges – one directed towards node $i+1$ and the other directed towards node 1. So, at each step of a random walk on the graph, there is an equal probability of moving to the next node or returning to the start. Since reaching the end node from the start node requires n successes in a row, the expected cover time is $\theta(2^n)$.

4 Space-Efficient Single-Pair Connectivity Checking

We can use random walks to give a simple algorithm to check in randomized logspace whether two vertices lie on the same connected component. For concreteness, define:

Definition 12. The **undirected connectivity problem (UST-conn)** is to decide whether two given vertices s and t of an undirected graph G lie on the same connected component.

Definition 13. The complexity class **RL** is class of problems that can be solved by polynomial-time and logarithmic-space Turing machines M with one-sided error, meaning

- If $x \in L$, then M accepts with probability at least $2/3$.
- If $x \notin L$, then M rejects.

Theorem 14. *The undirected connectivity problem (UST-conn) is in RL.*

Proof. We use the following simple algorithm: Let the input graph be G, start node s, and end node t. Begin at the start node s, and take a random walk in G for $4n^3$ steps. If you reach t among those steps, accept; otherwise, reject.

First, we check the one-sided error correctness condition.

If s and t lie on the same connected component, then by applying Theorem 10 to that subgraph G' of that component gives $h_{s,t} \leq C(G') \leq n^3$. Thus, the expected number of steps to reach t is n^3, so by Markov's inequality, the probability of failing to reach t within $4n^3$ steps is at most $1/4$.

If s and t lie on different connected components, no path from s can reach t, so the algorithm always rejects.

Now, we only need to check that this algorithm runs in polynomial time and logarithmic space. The polynomial running time is clear. For logarithmic space, it is easy to see that each of the steps needed can be done in logarithmic space:

- Use $\lg n$ bits to store the location of the current node
- Use $O(\lg n)$ bits to choose a random neighbor.
- Use $\lg 4n^3 = \theta(\lg n)$ bits to keep the count of the current step.

□

A further result (Reingold 2008) shows that UST-conn is in L (logspace) by derandomizing the random walk. It's open where $RL \subseteq L$, but it's known that RL is contained in $L^{3/2}$, the class of problems solvable in $\log(n)^{3/2}$ space and polynomial time.

5 Linear Algebra Review

We'll write down some definitions from linear algebra that will be useful for proving things about random walks. We'll use the convention where vectors are written as row vectors $v = (v_1, \ldots, v_n)$, and matrices act on the right, as in $w = vM$. This might look opposite to what you're used to, but the convention doesn't affect the results.

Definition 15. A vector v is an **eigenvector** of A with **eigenvalue** λ if

$$vA = \lambda A$$

For example, if P is the transition matrix of a d-regular graph, then the fact that the graph is doubly stochastic implies that

$$\left(\frac{1}{n},\ldots,\frac{1}{n}\right)P=\left(\frac{1}{n},\ldots,\frac{1}{n}\right),$$

so the uniform distribution vector $\left(\frac{1}{n},\ldots,\frac{1}{n}\right)$ is an eigenvector of P with eigenvalue 1.

Fact 16. *If a matrix P has eigenvalues $v^{(1)},\ldots,v^{(n)}$ with corresponding eigenvalues $\lambda^{(1)},\ldots,\lambda^{(n)}$, then the following matrices have the same eigenvectors and eigenvalues related as follows:*

- *αP has eigenvectors $\left\{v^{(i)}\right\}$ with corresponding eigenvalues$\left\{\alpha\lambda^{(i)}\right\}$, since*

$$v^{(i)}\left(\alpha P\right)=\left(\alpha\lambda^{(i)}\right)v^{(i)}$$

- *$P+I$ has eigenvectors $\left\{v^{(i)}\right\}$ with corresponding eigenvalues$\left\{\lambda^{(i)}+1\right\}$, since*

$$v^{(i)}\left(P+I\right)=\lambda^{(i)}v^{(i)}+v^{(i)}=\left(\lambda^{(i)}+1\right)v^{(i)}$$

- *P^k has eigenvectors $\left\{v^{(i)}\right\}$ with corresponding eigenvalues$\left\{\left(\lambda^{(i)}\right)^k\right\}$, since*

$$\begin{aligned}v^{(i)}P^k &= \lambda^{(i)}v^{(i)}P^{k-1}\\ &= \left(\lambda^{(i)}\right)^2v^{(i)}P^{k-2}\\ &\cdots\\ &= \left(\lambda^{(i)}\right)^kv^{(i)}\end{aligned}$$

Here's an example application: If P is the transition matrix for a Markov chain, the $\left(P+I\right)/2$ corresponds to a "slowed down" Markov chain where at each step we have a $1/2$ chance of staying still and a $1/2$ chance of transitioning. Then, the slowed down Markov chain has eigenvalues $\left\{\left(\lambda^{(i)}+1\right)/2\right\}$.

We will want to talk about distance between vectors, for instance to talk about the convergence of distributions. To do so, we define a norm, as well as the related notion of orthonormality.

Definition 17. The $\mathcal{L}_2$ norm of $v=(v_1,\ldots,v_n)$ is

$$\mathcal{L}_2\left(v\right)=\sqrt{v\cdot v}=\sum_{i=1}^{n}v_i^2$$

The $\mathcal{L}_2$ norm is also written as $\|\cdot\|_2$.

Definition 18. A collection of vectors $v^{(1)},\ldots,v^{(n)}$ is **orthonormal** if they are mutually orthogonal and each have $\mathcal{L}_2$ norm equal to 1, or equivalently if

$$v^{(i)}\cdot v^{(i)}=\delta_{ij}=\begin{cases}0, & \text{if } i\neq j\\ 1, & \text{if } i=j\end{cases}$$

When a matrix is symmetric and real, its eigenvectors have a special form.

Theorem 19. *If P is a real, symmetric $n\times n$ matrix, then it admits a collection of eigenvectors $v^{(1)},\ldots,v^{(n)}$ that are orthonormal.*

The choice of orthogonal eigenbasis is not unique in general. When an eigenvalue is repeated, there are multiple choices of orthogonal eigenvectors spanning that eigenspace.

For an orthogonal eigenbasis, we use eigenvectors that are of unit $\mathcal{L}_2$ length. In contrast, distribution vectors have unit length in the $\mathcal{L}_1$ norm because the total probability is 1. For instance, the uniform distribution eigenvector $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ yields $\mathcal{L}_1$ norm 1. To use this as an eigenvector in an orthonormal eigenbasis, we'd need to scale it to $\left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$.

Here's a further fact about stochastic matriches that we'll state without proof:

Fact 20. *If P is stochastic, then its eigenvectors are at most 1 in absolute value, $\left|\lambda^{(i)}\right| \leq 1$.*

The stationary distributions are the eigenvectors with eigenvalue 1. If P is ergodic, the by Theorem 3, there is a unique stationary distribution. So, if we write its eigenvalues in decreasing order of absolute value, we have

$$1 = \lambda^{(1)} < \left|\lambda^{(2)}\right| \leq \left|\lambda^{(3)}\right| \leq \cdots \leq \left|\lambda^{(n)}\right| \leq 0$$

Writing distributions in the orthonormal eigenbasis of the transition matrix will be very useful for understanding what happens to the distribution when the transition matrix is applied repeatedly.

Theorem 21. *Suppose that P is the transition matrix of a random walk on undirected, non-bipartite, d-regular connected graph. Then, for any start distribution Π_0, the distributions $\Pi_k = \Pi_0 P^k$ tend to the stationary distribution $\overline{\Pi}$*

$$\lim_{k\to\infty} \Pi_k = \overline{\Pi}$$

Proof. Since P is real and symmetric, it has an orthonormal collection of eigenvectors $v^{(1)}, \ldots, v^{(n)}$. Write the initial distribution Π_0 in this eigenbasis.

$$\Pi_0 = \alpha^{(1)} v^{(1)} + \alpha^{(n)} v^{(2)} + \cdots + \alpha^{(n)} v^{(n)}$$

In the eigenbasis, the effect of applying the transition matrix is simple. The distribution after one step is

$$\begin{aligned} \Pi_1 = \Pi_0 P &= \left(\alpha^{(1)} v^{(1)} + \alpha^{(n)} v^{(2)} + \cdots + \alpha^{(n)} v^{(n)}\right) P \\ &= \alpha^{(1)} v^{(1)} P + \alpha^{(2)} v^{(2)} P + \cdots + \alpha^{(n)} v^{(n)P} \\ &= \alpha^{(1)} \lambda^{(1)} v^{(1)} + \alpha^{(2)} \lambda^{(2)} v^{(2)} + \cdots + \alpha^{(n)} \lambda^{(n)} v^{(n)} \end{aligned}$$

So, each component coefficient in the eigenbasis gets multiplied by the corresponding eigenvector. After k steps, the result is

$$\Pi_k = \Pi_0 P^k = \sum_{i=1}^{n} \alpha^{(n)} \left(\lambda^{(n)}\right)^k v^{(n)}$$

Since the Markov chain corresponding to P is ergodic, there is a unique stationary state $\overline{\Pi}$. This corresponds, up to a constant, to the eigenvector $v^{(1)}$ whose eigenvalue is $\lambda^{(1)} = 1$. Since $\lambda^{(1)} = 1$ and the rest of the eigenvalues have$\left|\lambda^{(i)}\right| < 1$, the coefficient of $v^{(i)}$ stays the same , while all other coefficients have $\left|\lambda^{(i)}\right| < 1$ and therefore decay exponentially to 0.

$$\lim_{k\to\infty} \left(\lambda^{(n)}\right)^k = \begin{cases} 1, & \text{if } n = 1 \\ 0, & \text{otherwise} \end{cases}$$

Therefore,

$$\lim_{k\to\infty} \Pi_k = \alpha^{(1)} v^{(1)}$$

Recall that $v^{(1)}$ is a multiple of $\overline{\Pi}$, so $\lim_{k\to\infty} \Pi_k = c\overline{\Pi}$ for some constant c. Since each Π_k has total probability 1, so does their limit, and therefore $c = 1$. So,

$$\lim_{k\to\infty} \Pi_k = \overline{\Pi}$$

□

We will now analyze the error terms to show that this convergence is exponentially fast

Theorem 22. *Suppose that P is the transition matrix of a random walk on a undirected, non-bipartite, d-regular connected graph. Then, for any start distribution Π_0,*

$$\left\|\Pi_0 P^k - \overline{\Pi}\right\|_2 \leq |\lambda_2|^k$$

Proof. As in Theorem 21, we have

$$\begin{aligned}
\Pi_k &= \sum_{i=1}^{n} \alpha^{(n)} \left(\lambda^{(n)}\right)^k v^{(n)} \\
&= \alpha^{(1)} v^{(1)} + \sum_{i=2}^{n} \alpha^{(n)} \left(\lambda^{(n)}\right)^k v^{(n)} \\
&= \overline{\Pi} + \sum_{i=2}^{n} \alpha^{(n)} \left(\lambda^{(n)}\right)^k v^{(n)}
\end{aligned}$$

So,

$$\begin{aligned}
\left\|\Pi_0 P^k - \overline{\Pi}\right\|_2 &= \left\| \sum_{i=2}^{n} \alpha^{(n)} \left(\lambda^{(n)}\right)^k v^{(n)} \right\|_2 \\
&= \sqrt{\sum_{i=2}^{n} \left(\alpha^{(n)}\right)^2 \left(\lambda^{(n)}\right)^{2k}}
\end{aligned}$$

by orthonormality. Since $\lambda^{(2)}$ is the greatest of the eigenvalues $\lambda^{(2)}, \ldots, \lambda^{(n)}$ in absolute value, we may use it to bound

$$\begin{aligned}
\left\|\Pi_0 P^k - \overline{\Pi}\right\|_2 &= \sqrt{\sum_{i=2}^{n} \left(\alpha^{(n)}\right)^2 \left(\lambda^{(n)}\right)^{2k}} \\
&\leq \sqrt{\sum_{i=2}^{n} \left(\alpha^{(n)}\right)^2 \left(\lambda^{(2)}\right)^{2k}} \\
&= \left|\lambda^{(2)}\right|^k \sqrt{\sum_{i=2}^{n} \left(\alpha^{(n)}\right)^2} \\
&= \left|\lambda^{(2)}\right|^k \sqrt{\|\Pi_0\|^2 - \left(\alpha^{(1)}\right)^2}
\end{aligned}$$

Since the entries of Π_0 are non-negative and sum to 1, the sum of their squares is at most 1 (with equality achieve when all but one entry is 0), so $\|\Pi_0\| \leq 1$. So,

$$\left\|\Pi_0 P^k - \overline{\Pi}\right\|_2 \leq \left|\lambda^{(2)}\right|^k$$

as desired. $\square$