Notes on Complexity Theory Last updated: October, 2011

Lecture 15

Jonathan Katz

1 Randomized Space Complexity

1.1 Undirected Connectivity and Random Walks

1.1.1 Markov Chains

We now develop some machinery that gives a different, and somewhat more general, perspective on random walks. In addition, we get better bounds for the probability that we hit t. (Note that the previous analysis calculated the probability that we end at vertex t. But it would be sufficient to pass through vertex t at any point along the walk.) The drawback is that here we rely on some fundamental results concerning Markov chains that are presented without proof.

We begin with a brief introduction to (finite, time-homogeneous) Markov chains. A sequence of random variables $X_0, \ldots$ over a space Ω of size n is a *Markov chain* if there exist $\{p_{i,j}\}$ such that, for all $t > 0$ and $x_0, \ldots, x_{t-2}, x_i, x_j \in \Omega$ we have:

$$\Pr[X_t = x_j \mid X_0 = x_0, \ldots, X_{t-2} = x_{t-2}, X_{t-1} = x_i] \;=\; \Pr[X_t = x_j \mid X_{t-1} = x_i] \;=\; p_{j,i}.$$

In other words, X_t depends only on X_{t-1} (that is, the transition is *memoryless*) and is furthermore independent of t. We view X_t as the "state" of a system at time t. If we have a probability distribution over the states of the system at time t, represented by a probability vector $\mathbf{p}_t$, then the distribution at time $t+1$ is given by $P \cdot \mathbf{p}_t$ (similar to what we have seen in the previous section). Similarly, the probability distribution at time $t + \ell$ is given by $P^\ell \cdot \mathbf{p}_t$.

A finite Markov chain corresponds in the natural way to a random walk on a (possibly directed and/or weighted) graph. Focusing on undirected graphs (which is all we will ultimately be interested in), a random walk on such a graph proceeds as follows: if we are at a vertex v at time t, we move to a random neighbor of v at time $t+1$. If the graph has n vertices, such a random walk defines the Markov chain given by:

$$p_{j,i} = \begin{cases} k/\deg(i) & \text{there are } k \text{ edges between } j \text{ and } i \\ 0 & \text{otherwise} \end{cases}.$$

We continue to allow (multiple) self-loops; each self-loop contributes 1 to the degree of a vertex.

Let $\mathbf{p}$ be a probability distribution over the states of the system. We say $\mathbf{p}$ is *stationary* if $P \cdot \mathbf{p} = \mathbf{p}$. We have the following fundamental theorem of random walks on undirected graphs (which is a corollary of a more general result for Markov chains):

Theorem 1 *Let G be an undirected, connected, non-bipartite graph on n vertices, and consider the transition matrix corresponding to a random walk on G. Then:*

1. *There is a unique stationary distribution $\mathbf{p} = (p_1, \ldots, p_n)$.*

2. *Let $h_{i,i}$ denote the expected number of steps for a random walk beginning at vertex i to return to i. Then $h_{i,i} = 1/p_i$.*

In particular, the graph need not be regular.

We do not prove Theorem 1 here. (A proof of the first claim, and intuition for the second claim can be found in [1, Lecture 8] or dedicated texts on Markov chains, e.g., [2].) Note that for any undirected graph G, the conditions of the theorem can always be met by (1) restricting attention to a connected component of G, and (2) adding a self-loop to any vertex in the connected component.

What is the stationary distribution for a given graph? Say we have an undirected, connected, non-bipartite graph G with m edges and ℓ self-loops. It can be verified by a simple calculation that setting $p_i = \frac{\deg(i)}{2m+\ell}$ for each vertex i gives a stationary distribution. (For each non-self-loop incident on vertex i, the probability mass exiting i via that edge is $\frac{1}{2m+\ell}$, which is equal to the probability mass entering i via that edge.) It follows that, for any vertex i, we have $h_{i,i} = \frac{2m+\ell}{\deg(i)}$.

There is another way to view the random walk on G: by looking at the graph G' on $2m + \ell$ vertices where each vertex in G' corresponds to an edge plus direction (for non-self-loops) of G, and there is an edge in G' between vertices (i, j) and (j', k') iff $j = j'$. The graph G' is now a directed graph, but Theorem 1 can be shown to apply here as well.[1] Note also that a random walk in G corresponds exactly to a random walk in G'. In G', however, the stationary distribution is the uniform distribution. (This can be verified by calculation, or derived from the stationary distribution on G.) Thus, for any edge (i, j) in G (which is just a vertex in G'), the expected number of steps to return to that edge (with direction) after crossing that edge is $1/(2m + \ell)$.

Let $h_{i,j}$ denote the expected number of steps to go from vertex i to vertex j. With the above in hand we can prove the following:

Theorem 2 *Consider a random walk on an undirected, connected, non-bipartite graph G with ℓ self-loops and m (other) edges. If there is an edge in G from vertex i to vertex j then $h_{i,j} + h_{j,i} \leq 2m + \ell$ and, in particular, $h_{i,j} < 2m + \ell$.*

Proof We prove the theorem in two ways. Looking at the random walk in G, we have seen already that $h_{i,i} = \frac{2m+\ell}{\deg(i)}$. If $i = j$ in the theorem then there is a self-loop from i to itself; because G is connected we must have $\deg(i) \geq 2$ and so the theorem holds. For $i \neq j$, we have:

$$\frac{2m + \ell}{\deg(j)} = h_{j,j} = \frac{1}{\deg(j)} \cdot \sum_{k \in N(j)} (1 + h_{k,j}),$$

where $N(j)$ are the neighbors of j (the above assumes j has no self-loops or multiple edges, but the analysis extends to those cases as well). Thus if there is an edge connecting (distinct) vertices i, j (so $i \in N(j)$), then $h_{i,j} < 2m + \ell$. (That $h_{i,j} + h_{j,i} \leq 2m + \ell$ is left as an exercise, but see next.)

Alternately, we may consider the random walk on the graph G' defined earlier. When we take a step from vertex i to vertex j in our random walk on G, we view this as being at vertex (i, j) in the graph G'. We have seen that the stationary distribution in G' is uniform over the $2m + \ell$ edges (with direction), which means that the expected time to re-visit the edge (i, j) is $2m + \ell$. But re-visiting edge (i, j) corresponds to a one-step transition from i to j, re-visiting i, and then following edge (i, j) again. In other words, beginning at j, the expected number of steps to visit i and then follow edge (i, j) is $2m + \ell$. This gives the desired upper bound on $h_{j,i} + h_{i,j}$. ■

[1] Advanced note: G' is connected since G is, and is *ergodic* since G is. Ergodicity is all that is needed for Theorem 1.

With can now analyze the random-walk algorithm for UCONN. Given undirected graph G with n vertices and $|E|$ edges, and vertices s, t in G, consider the connected component of G containing s. (Technically, we can imagine adding a self-loop at t to ensure that G is non-bipartite. However, it is clear that this has no effect on the algorithm.) If t is in the same connected component as s then there is a path $(s = v_0, v_1, \ldots, v_\ell = t)$ with $\ell < n$; the expected number of steps to go from v_i to v_{i+1} is less than $2|E| + 1$. Thus the expected number of steps to go from $s = v_0$ to $t = v_\ell$ is $O(n|E|)$. Taking a random walk for twice as many steps, we will hit t at some point with probability at least 1/2.

1.1.2 A Randomized Algorithm for 2SAT

Another easy application of random walks is the following $\mathcal{RP}$ algorithm for 2SAT: Begin by choosing a random assignment for the n variables. Then, while there exists an unsatisfied clause C, choose one of the variables in C at random and flip its value. Repeat for at most $\Theta(n^2)$ steps, and output 1 if a satisfying assignment is ever found.

Let us show that this algorithm finds a satisfying assignment with high probability when one exists. Fix some satisfying assignment $\vec{x}$, and let the state of the algorithm be the number of positions in which the current assignment matches $\vec{x}$. (So the state i ranges from 0 to n.) When the algorithm chooses an unsatisfied clause, the value of at least one of the variables in that clause must differ from the corresponding value of that variable in $\vec{x}$; thus, the state increases with probability at least 1/2. The worst case is when the state increases with probability exactly 1/2 (except when $i = 0$, of course). (We can mentally add a self-loop to state n so the graph is non-bipartite.) We thus have a random walk on a line, in the worst case starting at $i = 0$. The expected number of steps to move from state 0 to state n is $h_{0,1} + \cdots + h_{n-1,n} \leq n \cdot (2n + 1) = O(n^2)$.

References

[1] J. Katz. Lecture notes for *CMSC 652 — Complexity Theory*. Fall 2005.

[2] M. Mitzenmacher and E. Upfal. *Probablity and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.