Notes on Complexity Theory Last updated: October, 2005

Handout 8

Jonathan Katz

1 Markov Chains and Random Walks on Graphs

Recall from last time that a random walk on a graph gave us an $\mathcal{RL}$ algorithm for the problem of undirected graph connectivity. In this class, we also saw an $\mathcal{RP}$ algorithm for solving 2-SAT (see [2, Chapter 7] for details). We now develop some of the theory behind Markov chains and random walks on (undirected) graphs, toward a proof of the following result that was used to analyze both of the above algorithms:

Theorem 1 *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 the expected time for a random walk, starting at i, to reach j is less than $2m + \ell$.*

Let us begin with a brief introduction to (finite, time-homogeneous) Markov chains and random walks on graphs, along with a proof of a central result in this area. (Some of the definitions and results that follow need to be modified slightly for the case of *infinite* Markov chains.) For finite state space Ω, a sequence of random variables $X_0, \ldots$ on Ω 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:

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

(From now on, we write i instead of x_i for brevity.) In other words, the transition from X_{t-1} to X_t is *memoryless* and depends only on the value of X_{t-1}; furthermore, the probability of a transition from state i to state j is time-independent. The t-step transition probabilities $p^t_{i,j}$ are defined in the natural way:

$$p^t_{i,j} = \begin{cases} p_{i,j} & t = 1 \\ \sum_{k \in \Omega} p_{i,k} \cdot p^{t-1}_{k,j} & t > 1 \end{cases}.$$

Viewing the $\{p_{i,j}\}$ as an $|\Omega| \times |\Omega|$ matrix P (called the *transition matrix*), the values $\{p^t_{i,j}\}$ correspond to the matrix 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 and non-weighted graphs (which is all we will ultimately be interested in for the purposes of these notes), 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 Ω are the vertices of this graph, such a random walk defines the Markov chain given by:

$$p_{i,j} = \begin{cases} 1/\deg(i) & j \text{ is a neighbor of } i \\ 0 & \text{otherwise} \end{cases}.$$

We remark that we allow self-loops in the graph. A self-loop contributes only 1 to the degree of the incident vertex.

Let π be a probability distribution over Ω, viewed as a row vector. We say π is *stationary* if $\pi \cdot P = \pi$; equivalently,

$$\text{for all } j \in \Omega: \quad \pi(j) = \sum_{i \in \Omega} \pi(i) \cdot p_{i,j}\,.$$

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 2 *Let G be an undirected, connected, non-bipartite graph on n vertices. Then:*

1. *There is a unique stationary distribution $\pi = (\pi(1), \ldots, \pi(n))$. Furthermore, all entries in π are non-zero.*
2. *For all vertices i, j, we have $\lim_{t\to\infty} p^t_{i,j} = \pi(j)$. Note that the limit is independent of i. In words, this means that no matter where we start the random walk (i.e., regardless of our starting point i) we end up in state j (for large enough t) with the same probability $\pi(j)$.*
3. *Let $h_{i,i}$ denote the expected number of steps for a random walk beginning at vertex i to return to i. Then $h_{i,i} = 1/\pi(i)$.*

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 self-loops to all vertices. We will prove existence of a stationary distribution π as well as part 2 of the above theorem, and give an intuitive justification for part 3 (uniqueness of the stationary distribution, and the fact that all entries in π are non-zero, are left as exercises). Actually, we will prove a more general result for *ergodic* Markov chains: a finite Markov chain is ergodic if there exists a t_0 such that for all i, j and $t > t_0$ we have $p^t_{i,j} > 0$. It is not too difficult to see that a random walk in an undirected, connected, non-bipartite graph defines an ergodic Markov chain, and so this is indeed a more general result. (In fact, it is not hard to see that the converse — a random walk in an undirected graph G defines an ergodic Markov chain only if G is connected and non-bipartite — holds as well.)

Before proving the theorem, we introduce the notion of *coupling* and prove a lemma which will allow us to bound the statistical difference between two random variables. For distributions μ, ν over the same space Ω, a distribution ω on $\Omega \times \Omega$ is a *coupling* if the marginal distributions of ω give μ and ν, respectively; i.e.,

$$\text{for all } x \in \Omega: \quad \mu(x) = \sum_{y \in \Omega} \omega(x, y)$$

$$\text{for all } y \in \Omega: \quad \nu(y) = \sum_{x \in \Omega} \omega(x, y).$$

We have the following lemma:

Lemma 3 *Let ω be a coupling of distributions μ, ν. Then:*

$$\mathsf{SD}(\mu, \nu) \leq \Pr_{(X,Y) \leftarrow \omega}[X \neq Y],$$

where SD is the statistical difference between μ and ν, defined as:

$$\mathsf{SD}(\mu, \nu) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

Proof Let us first re-write the expression for SD:

$$\begin{aligned}
\mathsf{SD}(\mu,\nu) &= \frac{1}{2}\cdot\sum_{x\in\Omega}|\mu(x)-\nu(x)| \\
&= \frac{1}{2}\cdot\left(\sum_{x|\nu(x)<\mu(x)}\big(\mu(x)-\nu(x)\big)+\sum_{x|\mu(x)\le\nu(x)}\big(\nu(x)-\mu(x)\big)\right) \\
&= \frac{1}{2}\cdot\left(\sum_{x|\nu(x)<\mu(x)}\mu(x)-\sum_{x|\nu(x)<\mu(x)}\nu(x)+\sum_{x|\mu(x)\le\nu(x)}\nu(x)-\sum_{x|\mu(x)\le\nu(x)}\mu(x)\right) \\
&= \frac{1}{2}\cdot\left(\left(1-\sum_{x|\mu(x)\le\nu(x)}\mu(x)\right)-\sum_{x|\nu(x)<\mu(x)}\nu(x)+\left(1-\sum_{x|\nu(x)<\mu(x)}\nu(x)\right)-\sum_{x|\mu(x)\le\nu(x)}\mu(x)\right) \\
&= 1-\sum_{x|\mu(x)\le\nu(x)}\mu(x)-\sum_{x|\nu(x)<\mu(x)}\nu(x).
\end{aligned}$$

Now, since ω is a valid coupling we know that $\omega(x,x)\le\min\{\mu(x),\nu(x)\}$ for all $x\in\Omega$. So:

$$\begin{aligned}
\Pr[X\neq Y]=1-\sum_{z\in\Omega}\omega(z,z) &\ge 1-\sum_{z\in\Omega}\min\{\mu(z),\nu(z)\} \\
&= 1-\sum_{z|\mu(z)\le\nu(z)}\mu(z)-\sum_{z|\nu(z)<\mu(z)}\nu(z) \\
&= \mathsf{SD}(\mu,\nu),
\end{aligned}$$

as desired. ■

We now prove the relevant parts of Theorem 2:

Proof Given a Markov chain defined by transition matrix P, create two copies of it denoted $X_0,\ldots$ and $Y_0,\ldots$ (where X_0,Y_0 will be chosen according to some distribution to be specified). Create a coupling by defining random variables $W_0,\ldots$ in the following way: $W_0=(X_0,Y_0)$ where X_0,Y_0 are distributed independently according to X_0,Y_0. If $W_t=(x_t,y_t)$ then we choose $W_{t+1}=(X_{t+1},Y_{t+1})$ as follows: (1) choose X_{t+1} based on x_t according to P, then (2) if $y_t=x_t$, set $Y_{t+1}=X_{t+1}$; otherwise, choose Y_{t+1} based on y_t according to P. (In words: we let X_i and Y_i evolve independently until they meet; once they meet, they evolve in tandem.) Clearly W_t is a coupling of X_t and Y_t for any t.

Now, since the Markov chain is ergodic (and finite), we know there exist t^*,ε such that $p_{i,j}^{t^*}\ge\varepsilon>0$ for all i,j. Therefore, for any choice of initial states x_0,y_0 we have

$$\Pr[X_{t^*}\neq Y_{t^*}\mid X_0=x_0,Y_0=y_0]\le 1-\varepsilon.$$

Now, if $X_{t^*}=Y_{t^*}$ then $X_t=Y_t$ for all $t\ge t^*$. On the other hand, say $X_{t^*}=x_0'$ and $Y_{t^*}=y_0'$ with $x_0'\neq y_0'$. Using memorylessness:

$$\Pr[X_{2t^*}\neq Y_{2t^*}\mid X_{t^*}=x_0',Y_{t^*}=y_0']=\Pr[X_{t^*}\neq Y_{t^*}\mid X_0=x_0',Y_0=y_0']\le 1-\varepsilon.$$

Putting these observations shows that

$$\Pr[X_{2t^*}\neq Y_{2t^*}\mid X_0=x_0,Y_0=y_0]\le(1-\varepsilon)^2,$$

and so on for any multiple kt^*. Since for $t' > t$ we have $X_t = Y_t \Rightarrow X_{t'} = Y_{t'}$, we see that for any choice of initial states we have

$$\lim_{t\to\infty} \Pr[X_t \neq Y_t] = 0.$$

By Lemma 3, this implies that the statistical difference between X_t and Y_t goes to 0 as t approaches infinity.

Fix some $y \in \Omega$ and for any x define $a_t(x) = p^t_{x,y}$. We first show: (1) $a_t(x)$ converges for all x, and (2) $a_t(x)$ converges to the *same value* for all x. In fact, setting $X_0 = x_1$ and $Y_0 = x_2$ and using the fact that the statistical difference between X_t and Y_t goes to 0 as t approaches infinity, condition (2) follows immediately once we show (1). But (1) follows from the observation that we can set $X_0 = x$ and $Y_0 = \pi_x P$ (where π_x is the distribution in which x takes probability 1 and all other states have probability 0), so that the distribution of Y_t is exactly the same as the distribution of X_{t+1}. It follows from the fact that $\mathsf{SD}(X_t, Y_t)$ approaches 0 that $\mathsf{SD}(X_t, X_{t+1})$ approaches 0; since $a_t(x)$ is bounded (it is a probability, so $0 \leq a_t(x) \leq 1$), it follows that the sequence converges.

Let $\pi(y) \stackrel{\text{def}}{=} \lim_{t\to\infty} p^t_{x,y}$ (by what we have just proved, the choice of x does not matter). We want to show that π is stationary. For any $y \in \Omega$ we have:

$$\begin{aligned}
\sum_{x\in\Omega} \pi(x) \cdot p_{x,y} &= \sum_{x\in\Omega} \left(\lim_{t\to\infty} p^t_{x_0,x}\right) \cdot p_{x,y} \\
&= \lim_{t\to\infty} \sum_{x\in\Omega} p^t_{x_0,x} \cdot p_{x,y} \\
&= \lim_{t\to\infty} p^{t+1}_{x_0,y} = \pi(y)\,,
\end{aligned}$$

as claimed (in the above, $x_0 \in \Omega \setminus \{x\}$ is an arbitrary state). ■

1.1 Hitting Times

Given a Markov chain/random walk on a graph, we are often interested in the expected number of steps to go from vertex i to vertex j; this is known as the *hitting time* and is denoted $h_{i,j}$. In particular, $h_{i,i}$ represents the expected time to walk from vertex i back to itself. We now heuristically argue — without giving a formal proof — that if the Markov chain has unique stationary distribution π, then $h_{i,i} = 1/\pi(i)$ (this is part 3 of Theorem 2). To see this, consider a random walk $X_0, \ldots$ starting from an arbitrary vertex, and look at the long-term behavior of the walk. We know that for all t large enough we have $\Pr[X_t = i] = \pi(i)$. Let $\delta_i(t)$ be an indicator random variable which is 1 iff $X_t = i$. Then for t large enough we have

$$\mathbf{Exp}\left[\frac{\delta_i(t) + \delta_i(t+1) + \cdots + \delta_i(t+n-1)}{n}\right] = \frac{1}{n} \cdot \sum_{k=0}^{n-1} \mathsf{Exp}[\delta_i(t+k)] = \pi(i);$$

in other words, the expected frequency with which we are at vertex i is $\pi(i)$. It follows that, for t large enough, the expected time between visits to i is $1/\pi(i)$. Since the random walk is memoryless, this actually holds at all times (and not just for t large enough).

Finally, we come to the main result of interest for these notes (this is a slightly stronger version of Theorem 1):

Theorem 4 *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 Define $\deg(i)$ as the number of edges incident to vertex i, counting self-loops only once. We can prove the theorem in two ways: looking at either the vertices or the edges.

First approach. Consider the Markov chain in which the vertices are the states, and transition probabilities are defined in the natural way. It can be checked by a simple calculation that $\pi(v) = \frac{\deg(v)}{2m+\ell}$ is a stationary distribution. It follows that, for any vertex v, we have $h_{v,v} = \frac{2m+\ell}{\deg(v)}$. If there is a self-loop at v then, since the graph is connected, $\deg(v) \geq 2$ and so $h_{v,v} \leq \frac{2m+\ell}{2}$ and $h_{v,v} + h_{v,v} \leq 2m + \ell$. This proves the theorem for the case $i = j$. More generally, we have:

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

where $N(v)$ are the neighbors of v (the above assumes that v has no self-loops, but the analysis is the same either way). It follows that if there is an edge connecting (distinct) vertices u_0, v, then $h_{u_0,v} < 2m + \ell$. (That $h_{u_0,v} + h_{v,u_0} \leq 2m + \ell$ is left as an exercise, but see the next part.)

Second approach. Consider a Markov chain in which we have $2m + \ell$ states E' corresponding to the *edges* in our graph, taking direction into account (except for self-loops where the direction is irrelevant). When we take a step from vertex i to vertex j in our random walk, we view this as being in the "state" (i, j). Thus, we have a transition matrix defined by:

$$p_{(i,j),(j',k')} = \begin{cases} 1/\deg(j) & j = j' \\ 0 & \text{otherwise} \end{cases}.$$

(Note that, viewing this [in the natural way] as a new graph in which vertices correspond to edges [with direction] in our original graph, the resulting graph is ergodic). One can check that the uniform distribution over E' is stationary. It follows that the expected time to re-visit the edge (j, i) is $|E'| = 2m + \ell$. But re-visiting edge (j, i) corresponds to a one-step transition from j to i, re-visiting j, and then following edge (j, i) again. By the memorylessness of the random walk, this means that, beginning at i, the expected number of steps to visit j and then follow edge (j, i) is $|E'|$. But this gives the desired upper bound on the expected value of $h_{i,j} + h_{j,i}$. ■

2 Counting and $\#\mathcal{P}$

2.1 Introduction

Let R be a polynomial-time[1] (and polynomially-bounded) relation on pairs of elements. R of course defines the $\mathcal{NP}$ language $L_R = \{x \mid \exists y : (x, y) \in R\}$. But in addition to asking about *existence* of witnesses, we can also ask about the *number* of witnesses. With that in mind, define:

$$f_R(x) \stackrel{\text{def}}{=} |\{y \mid (x, y) \in R\}|.$$

[1]From now one, when we say that R is a polynomial-time relation we mean also that it is polynomially-bounded. Equivalently, membership of (x, y) in R can be decided in time polynomial in $|x|$ alone.

We want to measure the complexity of computing f_R. But f_R is a function and we like to speak in terms of languages. Given a relation R, define the language

$$\#R \stackrel{\text{def}}{=} \{(x,k) \mid f_R(x) \geq k\}.$$

We justify soon that the following definition of the *counting class* $\#\mathcal{P}$ captures what we want:

Definition 1 We say $L \in \#\mathcal{P}$ if there exists a polynomial-time relation R such that $L = \#R$. $\diamondsuit$

It should be clear that the class $\mathcal{NP}$ is Karp-reducible to $\#\mathcal{P}$, and it is not too hard to see that $\#\mathcal{P} \subseteq \mathsf{PSPACE}$.

Before continuing, one remark is due. For any language $L \in \mathcal{NP}$, there are multiple possible relations R such that $L = L_R$. So, technically, the counting problem corresponding to a language L is not well-defined (instead, as we have done, we need to specify the counting problem corresponding to an $\mathcal{NP}$ relation R). Nevertheless, sometimes we abuse terminology and say "$\#L$" when what we really mean is "$\#R$ for the natural relation R corresponding to L."

Why is Definition 2.1 given in terms of "$\geq$" rather than equality? One justification for the definition (as promised...) follows.

Proposition 5 *Lee R be a polynomial-time relation. Then deciding membership in $\#R$ and computing f_R are Cook reducible to each other.*

Proof Clearly, if we can compute f_R efficiently we can then efficiently decide membership in $\#R$. For the other direction, given an efficient procedure to decide membership in $\#R$ we can use binary search to compute f_R. ■

As further justification for Definition 2.1, note that for any relation R:

$$x \in L_R \Leftrightarrow (x,1) \in \#R.$$

This matches our intuition that *counting* solutions is at least as hard as determining *existence* of solutions. (Note that it would no longer be true if we defined $\#R$ using equality.)

2.2 $\#\mathcal{P}$-Completeness

We define $\#\mathcal{P}$-completeness in the natural way:

Definition 2 L is *$\#\mathcal{P}$-complete* if: (1) $L \in \#\mathcal{P}$, and (2) every $L' \in \#\mathcal{P}$ is Karp-reducible to L. $\diamondsuit$

(Note: $\#\mathcal{P}$-completeness is also sometimes defined via Cook reductions.) As one natural way to find $\#\mathcal{P}$-complete problems, we will look for an $\mathcal{NP}$-complete language L_R defined via a poly-time relation R and having the following additional property: For any poly-time relation Q (and associated language $L_Q \in \mathcal{NP}$), there exists a *parsimonious* Karp reduction f from L_Q to L_R, by which we mean that there exists a polynomial-time computable function f such that

$$f_Q(x) = f_R(f(x)).$$

Note that such an f immediately implies a Karp reduction from $\#Q$ to $\#R$ since:

$$(x,k) \in \#Q \;\Leftrightarrow\; (f(x),k) \in \#R.$$

One can verify (by examining the proofs) that the Karp reductions we have already shown for the "trivial" $\mathcal{NP}$-complete language (i.e., bounded halting), the circuit satisfiability problem, and SAT are all parsimonious. The implication is that, e.g., $\#SAT$ is $\#\mathcal{P}$-complete.

The above approach is not the only way to obtain $\#\mathcal{P}$-complete problems. In fact:

Proposition 6 *There exists a poly-time relation R such that $\#R$ is $\#\mathcal{P}$-complete, but $L_R \in \mathcal{P}$.*

Proof The counterexample is rather silly, but illustrative. Let R be any poly-time relation such that $\#R$ is $\#\mathcal{P}$-complete (say, $\#SAT$). Define the following poly-time relation R':

$$(x, y') \in R' \Leftrightarrow (y' = 0) \text{ or } (y' = 1y \text{ and } (x, y) \in R).$$

Note that $L_{R'} = \{0,1\}^* \in \mathcal{P}$; however, $\#R'$ is still clearly $\#\mathcal{P}$-complete (*why*?). ■

The above problem is not very natural. However, there are examples of natural problems which give the same result. We discuss one prominent example next.

2.3 Computing the Permanent is $\#\mathcal{P}$-Complete

The permanent of a square matrix $A = \{a_{i,j}\}$ is defined as:

$$\sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)} ,$$

where S_n is the set of permutations on n elements. You might recognize that this formula is very similar to the formula defining the *determinant* of a matrix; the difference is that in the case of the determinant there is an extra factor of $(-1)^{\text{sign}(\sigma)}$ in the sum, where $\text{sign}(\sigma)$ is the sign of σ. Nevertheless, although the determinant can be computed in polynomial time, computing the permanent (even of a 0-1 matrix) is $\#\mathcal{P}$-complete. We will say more about this problem next lecture.

Bibliographic Notes

Section 1 was written using [3, Lecture 2] and [2, Chapter 7] as references. Sections 2.1 and 2.2 are largely based on [1, Lect. 10].

References

[1] O. Goldreich. Introduction to Complexity Theory (July 31, 1999).

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

[3] E. Vigoda. Lecture notes for CS37101-1: Markov Chain Monte Carlo Methods (Autumn 2003). Available at http://www.cc.gatech.edu/~vigoda/MCMC_Course/index.html.