Lecture 17: May 27, 2021

Lecturer: Avrim Blum

1 Introduction to Markov chains

Last time we talked about random walks and cover time: the time to reach all nodes in the graph. By the way, in one of the proofs we showed that if we think of our state as being on the edges, then the stationary distribution of this walk is uniform on all $2m$ edge-directions. Note that this means that in the conventional view (state is on the nodes) the propability of being on a node is proportional to its degree. We can also see this directly by observing what happens if we initialize to this distribution and take one step of the walk.

A random walk on a graph is a special case of a random walk on a Markov chain.

A Markov chain with n states is a random walk process defined by an $n \times n$ matrix P where P_{ij} is the probability of moving to state j given that you currently are in state i. So all entries are non-negative and the row sums are equal to 1. Equivalently, if you describe your current state as a row vector q then your state after one step of the process is qP.

If the underlying graph (considering only those directed edges with non-zero probability) is strongly connected (meaning you can reach any state from any other state) then the chain is *irreducible*. We also say that an irreducible Markov chain is *aperiodic* if, for any starting distribution q, there exists some time T such that qP^T has nonzero probability on every state.

For example, a random walk on a connected bipartite graph would be irreducible but not aperiodic, since if you start at some node on the left, then for any time T, you would either have zero probability of being on the right or zero probability of being on the left depending on whether T is even or odd respectively. However, if you add self-loops (at each step there is now some nonzero chance of staying put) then the chain becomes aperiodic.

A *stationary distribution* π is a left eigenvector of eigenvalue 1. That is, $\pi = \pi P$.

Note, that this is the largest eigenvalue. This is because for any vector v (even if it has negative entries) the sum of the absolute values of the entries cannot increase when multiplying by P.

A Markov chain is *symmetric* if the matrix P is symmetric. For instance, a random walk on an undirected graph where each node has the same degree is symmetric. We will focus here on symmetric Markov chains. Note that for a symmetric Markov chain, all column sums are equal to 1 so the uniform distribution is a stationary distribution.

2 Rapid mixing

In many algorithmic settings, we will define a Markov chain on a "solution space" whose size is exponential in the natural problem parameters. We have no hope of visiting the entire space, but we would at least like to get close to the stationary distribution. This is used in simulated annealing, for example. Note that it is possible to get close to the stationary distribution much more quickly than visiting the entire graph. We will in particular say that a Markov chain is *rapidly mixing* if from any start state we can get close to the stationary distribution in $polylog(n)$ steps. (We'll define our notion of "close" shortly).

The main theorem we'll prove is that if a symmetric Markov chain P has the property that the eigenvalues corresponding to all eigenvectors that are *not* the stationary distribution π have magnitude noticeably less than 1, then the chain is rapidly mixing. In fact, we'll prove something slightly more general:

Theorem 2.1 *Say P is a Markov chain with real eigenvalues and orthogonal eigenvectors. Then, for any starting distribution $q^{(0)}$, the L_2 distance between the distribution after T steps $q^{(T)} = q^{(0)}P^T$ and the stationary distribution π is at most $|\lambda_2|^T$ where λ_2 is the eigenvalue of largest absolute value among eigenvectors orthogonal to π.*

Theorem 2.1 implies that if we have an eigenvalue gap of some value $\varepsilon > 0$, then for any constant c it takes only $T = O(\frac{\log n}{\varepsilon})$ steps to get $\left\|q^{(T)} - \pi\right\|_2 \leq 1/n^c$.

You might ask what happened to irreducibility and aperiodicity in this theorem. The answer is that in those cases, $|\lambda_2| = 1$ so the theorem becomes vacuous. For example, in a complete bipartite graph, a vector with all nodes on the left assigned value $1/n$ and all nodes on the right assigned value $-1/n$ is an eigenvector of eigenvalue -1.

Proof: [Theorem 2.1] Let's say the orthogonal eigenvectors are $v_1, \ldots, v_n$ with $v_1 = \pi$. Since the eigenvectors are a basis, we can write $q^{(0)}$ as:

$$q^{(0)} = c_1\pi + c_2v_2 + \ldots + c_nv_n,$$

for some values $c_1, \ldots, c_n$. After T steps of the walk, we have:

$$q^{(T)} = c_1\pi + c_2\lambda_2^Tv_2 + \ldots + c_n\lambda_n^Tv_n.$$

Notice that for $|\lambda_2| < 1$ (which we may assume since otherwise the theorem is vacuously true) then as $T \to \infty$, this approaches $c_1\pi$ and therefore we must have $c_1 = 1$. So,

$$\begin{aligned}\left\|q^{(T)} - \pi\right\|_2 &= \left\|c_2\lambda_2^Tv_2 + \ldots + c_n\lambda_n^Tv_n\right\|_2 \\ &\leq |\lambda_2|^T \left\|c_2v_2 + \ldots + c_nv_n\right\|_2 \leq |\lambda_2|^T \text{ (by orthogonality)}\end{aligned}$$

■