

## Lecture 6

Instructor: Salil Vadhan

Scribes: Junu Lee, Yash Nair

## 0 Agenda

- Random walks on directed graphs
- Pagerank
- Markov Chain Monte Carlo methods

## 1 Random Walks on Digraphs

We first restate the Perron-Frobenius Theorem, generalized for nonnegative matrices (i.e. matrices whose entries are all non-negative).

**Theorem 1** (Perron-Frobenius). *Let  $M$  be a nonnegative real matrix, and let  $G$  be a graph  $G$  with edge set  $E = \{(a, b) \mid M(a, b) < 0\}$ . Then:*

1. *There exists a non-negative eigenvector  $v_1$  with eigenvalue  $\mu_1$  such that  $\mu_1 \in \mathbb{R}$ .*
2.  *$\forall j > 1, \mu_1 \geq |\mu_j|$ .*
3. *If  $G$  is strongly connected, then the eigenvector  $v_1$  of  $\mu_1$  is strictly positive.*
4. *If  $G$  is strongly connected and aperiodic, then  $\mu_1 > |\mu_j| \forall j > 1$ .*
5. *If  $G$  has period  $k$ , then the set of eigenvalues is closed under multiplication by  $e^{2\pi i/k}$  for  $i \in \{1, \dots, k\}$ .*

**Remark** Note that part 5 is analogous to the case of an undirected bipartite graph in which the eigenvalues are closed under multiplication by  $-1$ ; every vertex in an undirected bipartite graph has period 2 and thus the period of the graph is 2, and thus is closed under multiplication by the primitive second root of unity. Thus, this result extends this to the directed case of period  $k$ .

We note that in the generalization,  $M$  is not necessarily symmetric, which means we cannot rule out the existence of complex eigenvalues.

We now revise our previous analysis on random walks for undirected graphs. We will see that in general we are able to say less things in the digraph case, as we are unable to relate convergence parameters like  $\omega_\pi$  to the eigenvalues of  $W$ , the random walk matrix of the graph.

Recall that for undirected graphs, we define the random walk matrix as  $W = MD^{-1}$ . In the directed case we instead define  $W = MD_{out}^{-1}$ . For initial distribution  $p_0 \in [0, 1]^V$ , the distribution over the states at time step  $t$  is  $p_t = W^t p_0$ .

In the undirected case, we were able to diagonalize  $W$  and analyze long-term behavior by looking at its spectrum. Unfortunately we are unable to do this in the directed case. Rather, we use the Jordan Normal Form to write  $W = SLS^{-1}$  for a lower triangular matrix  $L$  whose diagonal consists of the eigenvalues  $\mu_1, \dots, \mu_n$  and an invertible matrix  $S$ . Letting  $v_1$  be the non-negative eigenvector of  $W$  guaranteed by our directed version of Perron-Frobenius, and defining  $\pi = \frac{v_1}{\|v_1\|_1}$ , where  $\|x\|_1 = \sum_a |x(a)|$ , we see that  $\mu_1$  must be 1, because  $W\pi$  must be a valid probability distribution.

At this point, we assume that the graph associated with  $W$  is strongly connected and aperiodic, i.e. ergodic. In this case, we can use the Jordan normal form to write  $W$  as  $SLS^{-1}$ , where  $L$  has the properties

described above with the addition property that the first column is  $(1, 0, \dots, 0)^\top$ . This is since our restatement of the Perron-Frobenius theorem ensures that  $\mu_1$  is the only eigenvalue equalling 1. Then as  $t \rightarrow \infty$ , we have that

$$L^t \rightarrow N := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

since  $\mu_1 = 1$  will dominate and will be the only eigenvalue remaining as  $t$  gets large. As  $W^t = SLS^{-1} = SL^tS^{-1}$  for all  $t$ , it follows that  $\lim_{t \rightarrow \infty} W^t = SNS^{-1}$ .

From here, observe that  $\pi = Se_1$  (it is via a change-of-basis transformation of the first basis vector since our diagonalization has  $\mu_1 = 1$  in the top left entry). Then, as a result of the above limit,

$$W^t \rightarrow (c_1\pi \quad c_2\pi \quad \cdots \quad c_n\pi).$$

By the stochasticity of the matrix, we must have that all the  $c_i$  are 1.

Now, in the undirected case we were able to define a measure

$$\omega_\pi = \max_{\text{probability distributions } p} \frac{\|Wp - \pi\|_{\pi^{-1}}}{\|p - \pi\|_{\pi^{-1}}},$$

where

$$\|x\|_{\pi^{-1}} := \sqrt{\sum_{a=1}^n \frac{x(a)^2}{\pi(a)}}.$$

This measure still makes sense in the directed setting, but we are unable to relate it to the eigenvalues. Additionally, another key difficulty in this case is that we cannot immediately calculate  $\pi$ , the stationary distribution.

The  $\omega_\pi$  measure, however, does behave in the same way as in the undirected case. In particular, the discussion directly preceding Theorem 2 previous scribe notes holds in the directed case. Theorem 2 also holds in the directed case, except that we are not guaranteed the same bounds on  $\pi_{\min}$ ; in particular,  $\pi_{\min}$  can be exponentially small in directed graphs of bounded degree. Thus, the bound on mixing time from last time becomes as follows:

**Theorem 2.** *The "mixing time" to get within an  $\ell_1$  distance  $\epsilon$  of the stationary distribution is*

$$t = O\left(\frac{\log(1/\pi_{\min})}{1 - \omega_\pi}\right).$$

It can be shown that  $\omega_\pi$  can be 1 even in ergodic graphs  $G$  (a fact which we will show in Pset 2). Additionally, for the lazy random walk,  $\tilde{W} = \frac{1}{2}W + \frac{1}{2}I$ , we have that  $\tilde{\omega}_\pi$ , the corresponding measure for  $\tilde{W}$  is strictly less than 1 when  $G$  is strongly connected.

Additionally as shown in Pset 1, we have that for undirected, connected, non-bipartite graphs,  $\omega_\pi \leq 1 - 1/\text{poly}(n \cdot d_{\max})$  and thus were able to deduce that the mixing time is upper-bounded by  $\text{poly}(n \cdot d_{\max})$ . Unfortunately, this is not true in general for ergodic directed graphs: mixing time can be  $2^{\Omega(n)}$  even for bounded degree, unweighted directed graphs.

## 1.1 Non-Strongly Connected Graphs

In the case where  $G$  is not strongly connected, we can instead use the graph of strongly connected components in  $G$ , which itself is a directed acyclic graph  $D$ . Then every strongly connected component in  $D$  that is not a sink (or leaf, if we interpret  $D$  as a tree) is a transient state of the random walk. We can see then that any node in any of these non-sink components will have asymptotically zero mass, so all the probability mass concentrates in our sink components. It follows from this that our stationary distribution of our random walk over  $G$  is a convex combination of the stationary distributions within each sink of  $D$ .

**Exercise 3.** In terms of probabilities, what do the coefficients for the convex combination represent?

**Solution.** Let  $\pi_i$  be the stationary distribution within sink  $S_i$  of  $D$ . Now, treating  $D$  as a graph itself (so each strongly connected component is now a node), we define its stationary distribution as  $\pi_D$ . It is evident that  $\pi_D(X) > 0$  if and only if  $X = S_i$  for some  $i$ , as all non-sinks are part of the reducible component of  $D$ . If we interpret  $\pi_D(S_i)$  as the probability that we end up in  $S_i$ , then it makes sense that the probability we end up at  $v \in S_i$  is

$$P(v) = P(v \mid \text{in } S_i)P(S_i) = \pi_i(v)\pi_D(S_i).$$

Thus, the stationary distribution within sink  $S_i$  is weighted by  $\pi_D(S_i)$ .

## 2 Pagerank

We treat the World Wide Web as a directed graph of web pages, where we have an edge from page  $b$  to page  $a$  if the former contains links to the latter. In our adjacency matrix  $M$ ,  $b$  links to  $a$  if and only if  $M_{a,b} = 1$ ; otherwise,  $M_{a,b} = 0$ . Then we can define our random walk matrix as  $W = MD_{out}^{-1}$ , and furthermore we define a new stochastic matrix

$$\widetilde{W} = (1 - \alpha)W + \alpha J,$$

where  $\alpha \in (0, 1)$  and  $J$  is a matrix with entries all  $\frac{1}{n}$ . Then  $\widetilde{W}$  is both strongly connected and aperiodic, since it is now possible to move between any two sites in one step with probability at least  $\frac{\alpha}{n}$ .

**Definition 4.** The Pagerank vector is defined to be the unique stationary distribution  $\pi$  of  $\widetilde{W}$ .

The utility of the Pagerank vector  $\pi$  comes from its uniqueness property when our graph is directed and has no zero out-degrees—the intuition for this follows from the fact that combining  $W$  and  $J$  in this fashion makes the resulting Markov chain ergodic. In the specific case of an undirected webpage graph (i.e.  $M$  symmetric), we can show this directly. From our analysis in the previous section, we deduce that the mixing time  $t_{mix}$  of  $\widetilde{W}$  is on the order  $O\left(\frac{\log(n/\epsilon)}{\alpha}\right)$ , which is pretty small for  $\alpha$  that are not unreasonably tiny. Alternatively, there are a couple of approximations one can make to find  $\pi$ :

$$\begin{aligned} \widetilde{W}\pi = \pi &\implies \alpha\mathbf{u} = (I - (1 - \alpha)W)\pi \\ &\implies \pi = \alpha(I - (1 - \alpha)W)^{-1}\mathbf{u}, \end{aligned} \tag{1}$$

for which we argue that  $(I - (1 - \alpha)W)$  is, in fact, invertible. Note that the  $\mathbf{u}$  vector here is the vector with every entry  $\frac{1}{n}$ , which we have get  $J\pi$ . The other:

$$\pi \approx_\epsilon \widetilde{W}^t \mathbf{u} \implies \pi \approx_\epsilon \sum_{j=0}^t \alpha(1 - \alpha)^j W^j \mathbf{u}, \tag{2}$$

a Taylor expansion to approximate  $\pi$ .

At the end of the day, this is only a theoretical explanation of the underlying idea behind Google's page ranking system. They most likely use more tools to be able to approximate  $\pi$  faster, and also create better rankings through different choices of the matrix  $J$  (which leads to the process known as “personalized Pagerank”).

## 3 Markov Chain Monte Carlo

Our goal for this section is to develop a method to efficiently sample from a desired probability distribution on a huge, complicated graph. There are several applications in multiple fields for such methodology:

- Theoretical Computer Science: approximate counting.
- Statistics and Machine Learning: Bayesian inference and posterior sampling.
- Statistical Physics: estimating physical properties.

We begin with an analysis of an example related the first point above.

**Example.** *Given an undirected graph  $G$ , how do we count the number of spanning forests (i.e. acyclic subgraphs) of  $G$  (denoted  $Z_G$ )?*

If we want to find the exact number of spanning forests, we will find that this problem is actually NP-hard (and even #P-complete). However, there does exist a polynomial-time algorithm if we solely want to find the exact number of spanning trees.

To demonstrate the utility of efficient sampling, we note the following fact:

**Proposition 5.** *We can sample a near-uniform spanning forest in time  $\text{poly}(n, 1/\varepsilon) \iff$  we can approximately count the number of spanning forests to within a factor of  $(1 \pm \varepsilon)$  in time  $\text{poly}(n, 1/\varepsilon)$ .*

We can see the intuition for the left-to-right implication by considering this step-by-step analysis:

1. Pick an edge  $e$ .
2. By repeated sampling, estimate the probability

$$\widehat{P}_e \approx \mathbb{P}[\text{random spanning forest contains edge } e]$$

3.
  - If  $\widehat{P}_e < 1/2$ , recursively estimate the number of spanning forests in  $G - \{e\}$ , denoted as  $\widehat{Z}_{G-\{e\}}$ . We can then output  $\frac{\widehat{Z}_{G-\{e\}}}{1-\widehat{P}_e}$  as an estimate for  $Z_G$ .
  - Else, merge endpoints of edge  $e$  to get a new graph  $G'$ , then recursively estimate  $\widehat{Z}_{G'}$ . Then we output  $\frac{\widehat{Z}_{G'}}{\widehat{P}_e}$  as our estimate for  $Z_G$ .

The if-statement in step 3 based on  $\widehat{P}_e$  is important since it will keep our standard error for our estimate low. In the beginning of next lecture, we will tie up how we can use the Markov Chain Monte Carlo process to sample near-uniform random forests so that we can show the left-to-right implication of Proposition 6.