CS 798 - Algorithmic Spectral Graph Theory, Fall 2015, Waterloo

Lecture 7: Random walks

Random walks on graphs is a simple stochastic process with many algorithmic applications.

Today we analyze the mixing time using spectral techniques and discuss some applications in random sampling of combinatorial objects.

Overview

Given a graph, a random walk starts from a vertex, at each time step the walk moves to a uniformly random neighbor of the current vertex, and repeat.

We are interested in understanding the long-term behaviour of the walk. For example, if we take the walk for many time steps, will we end up at every vertex with equal probability?

Some of the basic mathematical questions that we will study in this course are:

(1) What is the limiting distribution of the random walk? (stationary distribution)

(2) How long does it take before the walk approaches the limiting distribution? (mixing time)

(3) Starting from a vertex s, what is the expected number of steps to first reach t? (hitting time)

(4) How long does it take to reach every vertex at least once? (cover time)

There are two main approaches to questions (1) and (2), one is probabilistic and uses the idea of "coupling" two random processes, another is spectral and uses the eigenvalues of the adjacency matrix.

We will study the spectral approach in this course, but not the probabilistic approach unfortunately.

Questions (3) and (4) are best answered by viewing the graph as an electrical network, and we will study this view point later in this course.

Matrix formulation

We first consider random walks in undirected graphs.

Let p_0 be the initial probability distribution. If we start the random walk in the i-th vertex, then p_0 is simply the i-th standard unit vector (with 1 in the i-th position and zero otherwise).

Let p_t be the probability distribution after t steps of random walk.

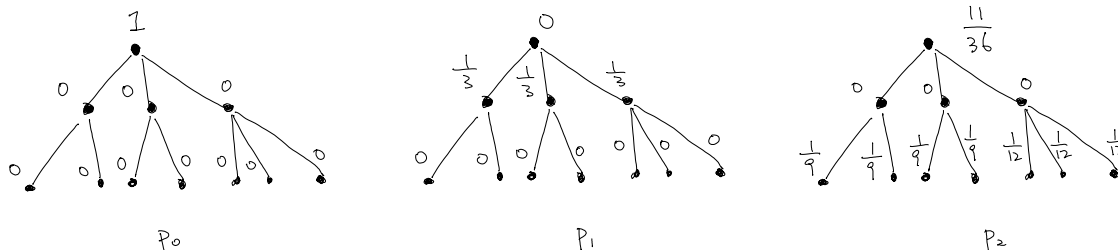For example, if we start the random walk in the top vertex of the following figure, then the probability distribution evolves as shown

1   0   11

then the probability distribution evolves as shown

$$P_0 \qquad\qquad P_1 \qquad\qquad P_2$$

Let $N(v)$ be the neighbor set of a vertex v.

Consider the random walk where in each step it moves to a uniformly random neighbor of the present node.

Then $\quad p_t(u) = \sum_{v \in N(u)} p_{t-1}(v) \cdot \frac{1}{d(v)} \quad$ for all $t \geq 1$ where $d(v)$ denotes the degree of v.

Let A be the adjacency matrix of G and D be the diagonal matrix where the i-th diagonal entry is $d(i)$.

Then the above equations can be written compactly as $\quad p_{t+1} = AD^{-1} p_t$, and thus $p_t = (AD^{-1})^t p_0$.

Stationary Distribution

A probability distribution π is a <u>stationary distribution</u> if $\quad \pi = (AD^{-1})\pi$. So, once we reach a stationary distribution, it stays as a stationary distribution forever. (Note that it is common to write the probability distribution as a row vector, e.g. $\pi = \pi(D^{-1}A)$, but we write it as a column.)

In undirected graphs, there is a natural stationary distribution based on the degree of vertices.

Let $\quad \pi(v) = \dfrac{d(v)}{\sum_{u \in V} d(u)} = \dfrac{d(v)}{2m}$. Let $\vec{d}$ be the vector with the i-th entry being $d(i)$. Then $\pi = \dfrac{\vec{d}}{2m}$.

It is easy to check that $\quad \pi = (AD^{-1})\pi$, as $(AD^{-1})\pi = (AD^{-1})\dfrac{\vec{d}}{2m} = \dfrac{1}{2m}(A\vec{1}) = \dfrac{\vec{d}}{2m} = \pi$.

So, if at some point we reach a stationary distribution π, then it will be the limiting distribution, i.e. $\lim_{t \to \infty} p_t = \pi$.

Is there always a limiting distribution?

Not necessarily. For example if the graph is disconnected, then the distribution after many steps depends on the initial distribution (e.g. which component does the starting vertex belongs to). Actually, even if the graph is connected we may not reach a stationary distribution. Consider a connected bipartite graph. If the starting vertex is on the left side, then the walk will be on the left side on even time steps, while it is opposite if the starting vertex is on the right side.

It turns out that these are the only obstacles, that every connected non-bipartite graph has a unique

stationary distribution, and it will be reached regardless of the initial distribution.

More generally, every connected graph (even bipartite graphs) has a unique stationary distribution
if we consider the _lazy_ random walks where we stay at the same vertex with probability $1/2$.

That is, $p_t(v) = \frac{1}{2} \cdot p_{t-1}(v) + \frac{1}{2} \sum_{u \in N(v)} p_{t-1}(u) \cdot \frac{1}{d(u)}$.

In matrix form, we have $p_t = \left(\frac{1}{2} I + \frac{1}{2} A D^{-1}\right) p_{t-1}$ for lazy random walks.

We denote $W = \frac{1}{2} I + \frac{1}{2} A D^{-1}$ as the _walk matrix_, and so $p_{t+1} = W p_t$ and $p_t = W^t p_0$.

Analysis for regular graphs

We first assume the graph G is d-regular, and prove that there is a unique limiting distribution if
G is connected and non-bipartite, and then analyze the convergence rate.

Then, we will explain how to generalize to general undirected graphs, and finally mention the directed case.

For d-regular graphs, the random walk matrix is simply $\frac{A}{d} = \mathcal{A}$, the normalized adjacency matrix,
and the lazy walk matrix is $W = \frac{1}{2} I + \frac{1}{2} \frac{A}{d} = \frac{1}{2} I + \frac{1}{2} \mathcal{A}$.

For d-regular graphs, the uniform distribution $\frac{\vec{1}}{n}$ is a stationary distribution.

We are interested in proving that $\lim_{t \to \infty} \mathcal{A}^t p_0 = \frac{\vec{1}}{n}$ when G is connected and non-bipartite, and
$\lim_{t \to \infty} W^t p_0 = \frac{\vec{1}}{n}$ as long as G is connected.

Spectrum

To study the behavior of the repeated applications of a linear transformation like $\mathcal{A}^t p_0$, it is usually very
useful to know the spectrum of the matrix.

Let $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of $\mathcal{A}$ and $v_1, v_2, \ldots, v_n$ be the corresponding eigenvectors.

Since $\mathcal{A}$ is symmetric, by the spectral theorem in L01, we know that all eigenvalues are real, and we
can choose v_i so that they form an orthonormal basis.

Recall that $\alpha_1 = 1$ and $v_1 = \frac{\vec{1}}{\sqrt{n}}$.

Also, recall that G is connected if and only if $\alpha_2 < 1$.

Furthermore, recall that G is non-bipartite if and only if $\alpha_n > -1$.

It turns out that $\alpha_2 < 1$ and $\alpha_n > -1$ is a necessary and sufficient condition for unique limiting distribution,
and so we could prove that a connected non-bipartite graph has a unique limiting distribution.

Uniqueness

We would like to show that $\lim_{t \to \infty} A^t p_0 = \frac{\vec{1}}{n}$ when G is connected non-bipartite.

Since the eigenvectors of A form an orthonormal basis, we can write $p_0 = c_1 v_1 + \ldots + c_n v_n$ where $c_i = \langle p_0, v_i \rangle$.

Then, $A^t p_0 = A^t (\sum_{i=1}^n c_i v_i) = \sum_{i=1}^n A^t c_i v_i = \sum_{i=1}^n c_i A^t v_i = \sum_{i=1}^n c_i \alpha_i^t v_i$.

Now, as G is connected and non-bipartite, we have $\alpha_2 < 1$ and $\alpha_n > -1$,

and so $\lim_{t \to \infty} A^t p_0 = \lim_{t \to \infty} \sum_{i=1}^n c_i \alpha_i^t v_i = \sum_{i=1}^n \lim_{t \to \infty} c_i \alpha_i^t v_i = c_1 \alpha_1 v_1$ (since $\alpha_i^t \to 0$ for $2 \leq i \leq n$).

Recall that $\alpha_1 = 1$ and $v_1 = \frac{\vec{1}}{\sqrt{n}}$, so $c_1 = \langle p_0, v_1 \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n p_0(i) = \frac{1}{\sqrt{n}}$ as p_0 is a probability distribution,

and thus $\lim_{t \to \infty} A^t p_0 = c_1 \alpha_1 v_1 = \frac{1}{\sqrt{n}} \cdot 1 \cdot \frac{\vec{1}}{\sqrt{n}} = \frac{\vec{1}}{n}$, as desired.

<u>Lazy random walk</u> : The analysis is almost the same. When we consider $W = \frac{1}{2} I + \frac{1}{2} A$, we have $W \succeq 0$

and so we don't need to worry about the smallest eigenvalue (and so don't need to worry about

bipartiteness). The condition G is connected is used to guarantee that the second largest eigenvalue

is less than one, and so we get the same uniqueness conclusion.

Mixing time

We would like to quantify how fast is p_t converging to the uniform distribution.

Given two probability distribution p and q, the <u>total variation distance</u> is defined as

$$d_{TV}(p, q) := \sum_{i=1}^n |p_i - q_i| = \|p - q\|_1.$$

The (total variation) <u>mixing time</u> is defined as the smallest t such that $d_{TV}(p^t, \pi) \leq \frac{1}{4}$,

where π is the limiting distribution (in the regular case, $\pi = \vec{1}/n$).

To bound the mixing time, we will use the same approach, and it should be clear that we need α_2 and

$|\alpha_n|$ to be bounded away from 1 so that α_i^t converges to zero quickly for $2 \leq i \leq n$.

Define $\alpha := \min \{1 - \alpha_2, 1 - |\alpha_n|\}$ be the spectral gap of A.

By the calculations above, we have $A^t p_0 = \sum_{i=1}^n c_i \alpha_i^t v_i = \frac{\vec{1}}{n} + \sum_{i=2}^n c_i \alpha_i^t v_i$.

So, $d(A^t p_0, \pi) = \|A^t p_0 - \frac{\vec{1}}{n}\|_1 = \|\sum_{i=2}^n c_i \alpha_i^t v_i\|_1 \leq \sqrt{n} \|\sum_{i=2}^n c_i \alpha_i^t v_i\|_2$ by Cauchy-Schwarz.

Since v_i are orthonormal, $\|\sum_{i=2}^n c_i \alpha_i^t v_i\|_2^2 = \sum_{i=2}^n \|c_i \alpha_i^t v_i\|_2^2 = \sum_{i=2}^n c_i^2 \alpha_i^{2t} \|v_i\|^2 \leq (1-\alpha)^{2t} \sum_{i=2}^n c_i^2$ by spectral gap.

Note that $\sum_{i=2}^n c_i^2 < \sum_{i=1}^n c_i^2 = \|p_0\|^2 \leq 1$, we have $d(A^t p_0, \pi) \leq \sqrt{n}(1-\alpha)^t$.

Therefore, if $t \geq \frac{\ln n}{\alpha}$, then we have $d(A^t p_0, \pi) \leq \sqrt{n}(1-\alpha)^t \leq \sqrt{n} e^{-\alpha t} \leq \sqrt{n} e^{-\ln n} \leq \frac{1}{\sqrt{n}} \ll \frac{1}{4}$.

We conclude that the mixing time is bounded by $\ln n / \alpha$.

<u>Lazy random walk</u>

The same analysis shows that the mixing time of lazy random walks is $\leq \dfrac{\ln n}{1 - \frac{\alpha_2}{2}} = \dfrac{2 \ln n}{\lambda_2}$.

By Cheeger's inequality $\lambda_2 \geq \dfrac{\phi(G)^2}{2}$, we conclude that the mixing time $\leq \dfrac{4 \ln n}{\phi^2(G)}$.

This implies that the mixing time is $O(\ln n)$ for an expander graph (with $\phi(G) = \Omega(1)$).

This result is very important in studying Markov chain mixing time with many applications in random sampling, and we will briefly discuss it later today.

<u>General undirected graphs</u>

The random walk matrix for general graphs is AD^{-1} and the lazy walk matrix is $W = \frac{1}{2} I + \frac{1}{2} AD^{-1}$.

The main difference is that these matrices are not necessarily symmetric, and so the spectral theorem does not apply directly.

Fortunately, W is similar to a symmetric matrix, where we say a matrix X is similar to Y if there exists a non-singular matrix B such that $X = BYB^{-1}$ (so the linear transformations of X and Y are the same up to change of basis B).

In our case, $W = \frac{1}{2} I + \frac{1}{2} AD^{-1} = D^{\frac{1}{2}} \left(\frac{1}{2} I + \frac{1}{2} D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) D^{-\frac{1}{2}} = D^{\frac{1}{2}} \left(\frac{1}{2} I + \frac{1}{2} \mathcal{A} \right) D^{-\frac{1}{2}}$, and so W is similar to $\frac{1}{2} I + \frac{1}{2} \mathcal{A}$, which is a symmetric matrix.

Note that two similar matrices have the same spectrum. One way to see it is to see that they have the same characteristic polynomial, as $\det(\lambda I - X) = \det(\lambda I - BYB^{-1}) = \det(B^{-1}(\lambda I - Y)B) = \det(\lambda I - Y)$.

In our case, we know that the spectrum of $\frac{1}{2} I + \frac{1}{2} \mathcal{A}$ is $1 = \frac{1}{2} + \alpha_1 \geq \frac{1}{2} + \alpha_2 \geq \dots \geq \frac{1}{2} + \alpha_n \geq 0$, where $1 = \alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n \geq -1$ is the spectrum of the normalized adjacency matrix $\mathcal{A}$, and so W also has the same spectrum.

Furthermore, let $v_1, \dots, v_n$ be an orthonormal basis of eigenvectors of $\mathcal{A}$ (and hence $\frac{1}{2} I + \frac{1}{2} \mathcal{A}$), then $D^{\frac{1}{2}} v_1, \dots, D^{\frac{1}{2}} v_n$ are linearly independent eigenvectors of W with eigenvalues $\frac{1}{2} + \alpha_1 \geq \dots \geq \frac{1}{2} + \alpha_n$.

To see it, note that $W(D^{\frac{1}{2}} v_i) = \left(D^{\frac{1}{2}} \left(\frac{1}{2} I + \frac{1}{2} \mathcal{A} \right) D^{-\frac{1}{2}} \right) \left(D^{\frac{1}{2}} v_i \right) = \left(\frac{1}{2} + \alpha_i \right) \left(D^{\frac{1}{2}} v_i \right)$ and so $D^{\frac{1}{2}} v_i$ is an eigenvector of W with eigenvalue $\frac{1}{2} + \alpha_i$. The eigenvectors $D^{\frac{1}{2}} v_i$ are linearly independent because v_i are and $D^{\frac{1}{2}}$ is a full rank matrix (simple exercise), and so they form a basis.

So, any initial probability distribution p_0 can be written as $p_0 = \sum_{i=1}^{n} c_i D^{\frac{1}{2}} v_i$.

With this setup, the remaining proof just follows the same lines as in the regular case, with some minor changes in the calculations.

It would be good to stop here and work out the details on your own. Below are the details for reference.

In the following, we assume $\alpha_2 \leq 1 - 2\varepsilon$, equivalently $\lambda_2 \geq 2\varepsilon$.

Uniqueness

$$W^t p_0 = \left(D^{\frac{1}{2}}\left(\frac{1}{2}I + \frac{1}{2}\mathcal{A}\right)D^{-\frac{1}{2}}\right)^t \left(\sum_{i=1}^n c_i D^{\frac{1}{2}} v_i\right) = \left(D^{\frac{1}{2}}\left(\frac{1}{2}I + \frac{1}{2}\mathcal{A}\right)^t D^{-\frac{1}{2}}\right)\left(\sum_{i=1}^n c_i D^{\frac{1}{2}} v_i\right) = \sum_{i=1}^n c_i \left(\frac{1}{2} + \frac{\alpha_i}{2}\right)^t D^{\frac{1}{2}} v_i.$$

All the terms would vanish when $t \to \infty$, except the first term $c_1 D^{\frac{1}{2}} v_1$.

For $\mathcal{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, note that $\mathcal{A}\left(D^{\frac{1}{2}}\vec{1}\right) = \left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\left(D^{\frac{1}{2}}\vec{1}\right) = D^{-\frac{1}{2}}A\vec{1} = D^{-\frac{1}{2}}\vec{d} = D^{\frac{1}{2}}\vec{1}$, so $D^{\frac{1}{2}}\vec{1}$ is an eigenvector of eigenvalue one, and thus $v_1 = D^{\frac{1}{2}}\vec{1}/\|D^{\frac{1}{2}}\vec{1}\| = D^{\frac{1}{2}}\vec{1}/\sqrt{2m}$ where $m = |E|$.

To compute c_1, note that $p_0 = \sum_{i=1}^n c_i D^{\frac{1}{2}} v_i$ and so $D^{-\frac{1}{2}} p_0 = \sum_{i=1}^n c_i v_i$. By orthonormality of v_i, we have $c_1 = \langle D^{-\frac{1}{2}} p_0, v_1 \rangle = \langle D^{-\frac{1}{2}} p_0, D^{\frac{1}{2}}\vec{1}/\sqrt{2m} \rangle = \langle p_0, \vec{1}/\sqrt{2m} \rangle = \frac{1}{\sqrt{2m}}$.

since p_0 is a probability distribution and so $\langle p_0, \vec{1} \rangle = 1$.

Therefore, $\lim_{t \to \infty} W^t p_0 = c_1 D^{\frac{1}{2}} v_1 = \frac{1}{\sqrt{2m}} D^{\frac{1}{2}}\left(D^{\frac{1}{2}}\vec{1}/\sqrt{2m}\right) = \vec{d}/2m$, the stationary distribution based on the degrees of the vertices.

Mixing time

Let $\pi = \vec{d}/2m$ be the unique limiting distribution.

We are interested in bounding $\|W^t p_0 - \pi\|_1$.

Note that $W^t p_0 - \pi = \sum_{i=2}^n c_i \left(\frac{1+\alpha_i}{2}\right)^t D^{\frac{1}{2}} v_i$

Taking advantage of the orthonormality, we first bound

$$\left\|D^{-\frac{1}{2}}W^t p_0 - D^{-\frac{1}{2}}\pi\right\|_2^2 = \left\|\sum_{i=2}^n c_i \left(\frac{1+\alpha_i}{2}\right)^t v_i\right\|_2^2 = \sum_{i=2}^n c_i^2 \left(\frac{1+\alpha_i}{2}\right)^{2t}\|v_i\|_2^2 \leq \sum_{i=2}^n c_i^2 (1-\varepsilon)^{2t} \leq (1-\varepsilon)^{2t}/d_{min}.$$

as $\sum_{i=1}^n c_i^2 = \|D^{-\frac{1}{2}} p_0\|_2^2 = p_0^T D^{-1} p_0 \leq \frac{1}{d_{min}}$ where d_{min} denotes the minimum degree

Note that $\left\|D^{-\frac{1}{2}}W^t p_0 - D^{-\frac{1}{2}}\pi\right\|_2^2 = \left\|D^{-\frac{1}{2}}(W^t p_0 - \pi)\right\|_2^2 \geq \frac{1}{d_{max}}\|W^t p_0 - \pi\|_2^2$.

Therefore, $\|W^t p_0 - \pi\|_1 \leq \sqrt{n}\|W^t p_0 - \pi\|_2 \leq \sqrt{n}\sqrt{\frac{d_{max}}{d_{min}}}(1-\varepsilon)^t$.

Again, if $t \geq \frac{2\ln n}{\varepsilon}$, then $d_{TV}(W^t p_0, \pi) \ll \frac{1}{4}$.

This implies that the mixing time is at most $O\left(\frac{\ln n}{\lambda_2}\right)$, and by Cheeger's inequality it is at most $O\left(\frac{\ln n}{\phi(G)^2}\right)$.

Directed graphs (optional)

Let $G = (V, E)$ be a directed graph. Let A be the adjacency matrix of G, where its ij-th entry is one if vertex j has an edge to vertex i, otherwise zero. Let D be the out-degree matrix, where $D_{ii} = d^{out}(i)$ and $D_{ij} = 0$ for $i \neq j$. Let $p_t : V \to \mathbb{R}$ be the probability distribution at time t. Then $p_t = AD^{-1} p_{t-1}$, and thus $p_t = (AD^{-1})^t p_0$.

First, note that D^{-1} is not defined if some vertex has zero out-degree. In fact, if there is such a vertex, the random walk "dies" if it reaches such a vertex. So, let's assume that there is no such vertex, possibly by adding a self-loop on each vertex, in which case the the random walk will get stuck at such a vertex. So, if a graph has two such vertices, then there is no unique stationary distribution.

More generally, we say a subset of vertices $S \subseteq V$ is <u>strongly connected</u> if there is a directed path for each ordered pair of vertices $u, v \in S$. We say S is a <u>strong component</u> if it is maximally strongly connected. Let $S_1, \ldots, S_k$ be the set of strong components of G. Let H be the directed graph on vertices $\{1, 2, \ldots, k\}$ that contains a directed edge from vertex i to vertex j if there is a directed edge in G from a vertex in S_i to a vertex in S_j. Observe that H has no directed cycles; otherwise we would have a bigger strong component, contradicting the maximality of the strong components. If H has two vertices with no outgoing edges, then G will not have a unique limiting distribution, because if we start a random walk in these strong components, the walk could not leave. Also, if some vertex in H has an outgoing edge, then eventually there will be no probabilities left in this component, because the probability will slowly leak out and could not come back.

So, to have a unique limiting distribution where each vertex has a positive probability, we must have that the graph G itself is one big strong component, in which case we say G is strongly connected. You can think of it as the corresponding statement for an undirected graph being connected.

Even if the directed graph is strongly connected, there may not exist unique stationary distribution. For example, a directed bipartite graph does not have unique stationary distribution, by the same

reason as in the directed case.

There are other examples in directed graphs. Consider a directed cycle of k vertices. Then it has a "periodicity" of k. You can also generalize this example to "directed k-partite" graphs. To avoid "periodicity", we assume that each vertex has some self-loop, similar to the lazy random walk in the undirected case.

The following theorem is known as the fundamental theorem of Markov chains.

Theorem If G is strongly connected and aperiodic, then there is a unique limiting distribution Π and $\Pi_i = \frac{1}{h_{i,i}}$ where $h_{i,i}$ is the expected hitting time from vertex i to vertex i.

There are two different approaches to prove the theorem.

One is probabilistic and uses the idea of coupling; see e.g. [Hag02] for a proof.

Another is linear algebraic and uses the Perron-Frobenius theorem and the Jordan canonical form; see [07] of 2012 for more discussions and further references.

Pagerank (optional)

The brilliant idea of Google is to do page ranking based on the graph structure.

Suppose we have a directed graph representing the relationships between a set of webpages, i.e. there is an arc from page i to page j if page i has a link to page j.

The idea of the pagerank algorithm is that the webpages linked by many webpages are more important, and that the webpages linked by many important webpages are even more important.

Pagerank algorithm:

- Initially every page i has its pagerank value, $pr(i)$, equal to $1/n$.
- In each step, each page divide its current pagerank value equally to its outgoing links and send these equal shares to the pages that it links to. Each page updates its new pagerank value to be the sum of the shares it receives. Use these new values in the next step.

4/13
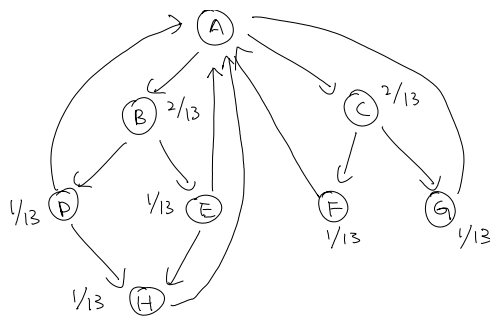
the "equilibrium" PageRank values.

This is just the same as random walk in a directed graph. To see this, observe that

$$pr_t(i) = \sum_{j: ji \in E} pr_{t-1}(j) / d^{out}(j).$$ Therefore we have $pr_t = (AD^{-1}) pr_{t-1}.$

So, to ask whether the pagerank algorithm converges to a set of unique values, it is the same as asking whether a random walk in the directed graph has a unique limiting distribution.

The directed graph of the webpages may not be strongly connected and aperiodic. To address this issue, the pagerank algorithm will jump to a uniformly random vertex with probability α, and with probability $1-\alpha$ go to a uniformly random out-neighbor. This guarantees that the resulting graph is strongly connected and aperiodic. Then, the algorithm will converge to the same values regardless of the initial distribution, showing that these values only depend on the graph structure.

Markov chain Monte Carlo   (optional)

One of the most important applications of random walks is to do random sampling.

Consider the problem of generating a random spanning tree in a graph.

The Markov chain Monte Carlo method would work as follows:

- Start from an arbitrary spanning tree.

- Define some simple local operations to move from one tree to another tree, e.g. add a random edge not in the tree and remove a random edge in the unique cycle created.

- Repeat for many steps and output the resulting tree.

Notice that this is just doing a random walk in a huge graph, in which there is a vertex for each spanning tree, and two spanning trees have an edge iff one can move from another by the simple local operations that we defined.

If we could show that this huge graph is an expander graph, then after $O(\log N) = O(poly\ n)$

steps where N and n are the number of vertices in the huge graph and the original graph respectively, the random walk mixes.

The huge graph can be usually defined in such a way that the graph is undirected and regular, so that the uniform distribution is the unique limiting distribution, and thus when the random walk mixes, we get a close to uniform random spanning tree.

The difficult part is to bound the mixing time.

One common approach is the coupling method that we didn't talk about.

Another is to argue that the graph is an expander graph and uses cheeger's inequality to bound the mixing time. To bound the expansion, one technique is to establish a multicommodity flow with low congestion, and this has been used successfully in sampling a random perfect matching in a bipartite graph, which is used to approximate the permanent of a matrix.

We won't discuss further about these topics. These are good project topics if you are interested.

For sampling random spanning trees, there are other algorithms to get an exact uniform sample.

One algorithm is to use the algorithm to count the number of spanning trees (see HW1), and thus we can compute the probabilities (and conditional probabilities) of having a specific edge.

There is a cool algorithm using random walks (but not by sampling).

Start a random walk from an arbitrary vertex v, until the random walk visits every vertex at least once (the cover time).

Now, for each vertex $u \neq v$, let wu be the first edge that visits u, then add wu to the tree. It can be proved that this is a uniform random spanning tree!

See [LPW08] for the proof and many more results like shuffling cards.

References   [Hag02]   Finite Markov chains and algorithmic applications, by Häggström, 2012.

[LPW08]   Markov chains and mixing times, by Levin, Peres, and Wilmer, 2008.