CS 860 Spectral graph theory. Spring 2019, Waterloo.

Lecture 6: Random walks

We study random walks on graphs and analyze the mixing time using eigenvalues.

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.

Some of the basic mathematical questions 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.

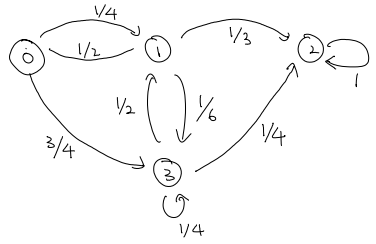
We will study the spectral approach in this course.

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

Today we will study questions (1) and (2) and discuss some applications.

Markov Chain

We analyze the general problem of random walk on a directed graph.



each vertex corresponds to a state

an arc corresponds to the transition prob. from state i to state j.

We can also formulate the problem as a matrix problem

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/3 & 1/4 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \end{pmatrix} \end{array},$$

i.e. P_{ij} = transition prob. from state i to state j.

Let X_t be the state at time t.

Let $p_t(i)$ be the probability of being at state i at time t.

For example, $\vec{p_0} = (1, 0, \ldots, 0)$ if the walk starts at state 0 at time 0,

or $\vec{p_0} = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ if the walk starts at a random state at time 0.

By the definition, it follows that $p_{t+1}(j) = \sum_{i=0}^{n-1} p_t(i) \cdot P_{i,j} \quad \forall j$, and we can write it compactly as $\vec{p_{t+1}} = \vec{p_t} P$ and more generally $\vec{p_{t+m}} = \vec{p_t} P^m$, where the vectors are row vectors.

This random process is called a Markov chain because it forgets about the past, i.e.

$$\Pr(X_t = a_t \mid X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \ldots, X_1 = a_1, X_0 = a_0) = \Pr(X_t = a_t \mid X_{t-1} = a_{t-1}) = P_{a_{t-1}, a_t}.$$

States

We will assume all Markov chains are finite in this course.

A Markov chain is <u>irreducible</u> if the underlying directed graph is strongly connected, i.e. there is a directed path from i to j and from j to i for every pair of vertices $i, j \in V$.

In other words, for every i, j, there exists ℓ such that $\Pr(X_{t+\ell} = s_j \mid X_t = s_i) > 0$.

The <u>period</u> of a state i is defined as $d(s_i) := \gcd\{ t \mid P^t_{i,i} > 0 \}$.

A state i is <u>aperiodic</u> if $d(s_i) = 1$.

A Markov chain is aperiodic if all states are aperiodic; otherwise it is periodic.

 is periodic.

For example, random walks on a bipartite graph is aperiodic as $d(s_i) = 2 \quad \forall i \in V$.

We use irreducibility and aperiodicity to get the following property.

<u>Theorem</u> For any finite, irreducible, and aperiodic Markov chain, there exists $T < \infty$ such that

$$(P^t)_{i,j} > 0 \quad \text{for any } i, j \quad \text{for any } t \geq T.$$

Stationary distribution

A stationary distribution of a Markov chain is a probability distribution $\vec{\pi}$ such that $\vec{\pi} = \vec{\pi} P$.

Informally, $\vec{\pi}$ is a "steady / equilibrium / fixed point" state, as $\vec{\pi} = \vec{\pi} P^t$ for any $t \geq 0$.

Intuitively, given any finite, irreducible, and aperiodic Markov chain, after we run it long enough, then we will completely forget about the history and "converge" to the same distribution.

Given two probability distributions $\vec{p} = (p_1, p_2, \ldots, p_n)$ and $\vec{q} = (q_1, q_2, \ldots, q_n)$, the <u>total variation distance</u>

of $\vec{p}$ and $\vec{q}$ is defined as $d_{TV}(p,q) = \frac{1}{2}\sum_{i=1}^{n} |p_i - q_i|$.

We say that $\vec{p}_t$ converges to $\vec{q}$ if $\lim_{t\to\infty} d_{TV}(\vec{p}_t, q) = 0$.

One more definition: the hitting time from i to j is defined as $H_{i,j} := \min\{ t \geq 1 \mid X_t = s_j, X_0 = s_i \}$.

The expected hitting time is defined as $h_{i,j} := E[H_{i,j}]$.

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

Theorem For any finite, irreducible, and aperiodic Markov chain, the followings hold.

1. There exists a stationary distribution $\vec{\pi}$.
2. The distribution $\vec{p}_t$ will converge to $\vec{\pi}$ as $t \to \infty$, no matter what is the initial distribution $\vec{p}_0$.
3. There is a unique stationary distribution.
4. $\pi_i = \lim_{t\to\infty} (P^t)_{i,i} = \frac{1}{h_{i,i}}$.

We will see a spectral proof of this theorem for undirected graphs.

Matrix Formulation

Let $G = (V, E)$ be an undirected graph.

Let $p_t : V \to \mathbb{R}$ be the probability distribution after t steps, i.e. $p_t(v) \geq 0$ and $\sum_{v\in V} p_t(v) = 1$.

The initial distribution is p_0, e.g. if we start at vertex i, then p_0 is the i-th standard unit vector.

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

Then, $p_t(u) = \sum_{v : v\sim u} p_{t-1}(v) \cdot \frac{1}{d(v)}$ 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 $p_{t+1} = AD^{-1}p_t$, and thus $p_t = (AD^{-1})^t p_0$.

Stationary Distribution

A probability distribution π is a stationary distribution if $\pi = (AD^{-1})\pi$. 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 $\pi(v) = \frac{d(v)}{\sum_{u \in V} d(u)} = \frac{d(v)}{2m}$. Let $\vec{d}$ be the vector with the i-th entry being $d(i)$. Then $\pi = \frac{\vec{d}}{2m}$.

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

Is it the unique stationary distribution? Will we always reach a stationary 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.

But it turns out that these are the only obstacles.

Theorem Every connected, non-bipartite undirected graph has a unique stationary distribution, and it will be reached regardless of the initial distribution.

The non-bipartiteness condition is to ensure that the Markov chain is aperiodic.

There is a simple modification of the random walks so that we can remove this assumption.

Consider the lazy random walk where in each step we stay at the same vertex with probability $1/2$,

i.e. $p_t(v) = \frac{1}{2} p_{t-1}(v) + \frac{1}{2} \sum_{u:uv} p_{t-1}(u) \cdot \frac{1}{d(u)}$.

In matrix form, we have $p_t = \left(\frac{1}{2}I + \frac{1}{2}AD^{-1}\right)^t p_0$

We write $W = \frac{1}{2}I + \frac{1}{2}AD^{-1}$ as the (lazy) random walk matrix.

By doing the lazy random walk, we make the graph non-bipartite, and the above theorem implies that the lazy random walk of any connected graph converges to the unique stationary distribution regardless of the initial probability distribution 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 non-regular undirected graphs, and finally mention the directed case.

For d-regular graphs, the random walk matrix W is simply $W = \frac{A}{d} = \mathcal{A}$, the normalized adjacency matrix,

and the lazy random 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.

Our goal is to prove that $\lim_{t\to\infty} W^t p_0 = \frac{\vec{1}}{n}$ regardless of the initial distribution p_0, as long as the graph is connected and non-bipartite. or we do the lazy random walk on a connected graph.

Spectrum

To compute $W^t p_0$, it is 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, we know from L01 that all eigenvalues are real and the eigenvectors form an orthonormal basis.

- We know that $\alpha_1 = 1$ and $v_1 = \frac{\vec{1}}{\sqrt{n}}$.
- Recall from L02 that G is connected if and only if $\alpha_2 < 1$.
- Furthermore, G is non-bipartite if and only if $\alpha_n > -1$ (homework).

It will turn out $\alpha_2 < 1$ and $\alpha_n > -1$ are the necessary and sufficient conditions for unique limiting distributions.

Following from the above combinatorial characterizations of α_2 and α_n, this will show that random walks on a connected non-bipartite graph have a unique limiting distribution.

In the regular case, the limiting distribution is the uniform distribution.

For lazy random walk, the spectrum of the walk matrix W is $\frac{1}{2}(1+\alpha_1) \geq \ldots \geq \frac{1}{2}(1+\alpha_n)$.

This implies that the smallest eigenvalue is at least $0 > -1$, and so the same proof works.

This is how the lazy walk allows us to remove the non-bipartiteness assumption.

Uniqueness

Let W be the walk matrix and $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be its eigenvalues with eigenvectors $v_1, \ldots, v_n$.

Claim If $\lambda_2 < 1$ and $\lambda_n > -1$, then $\lim_{t\to\infty} W^t p_0 = \frac{\vec{1}}{n}$ for any p_0.

proof Since the eigenvectors form an orthonormal basis, we can write $p_0 = c_1 v_1 + \ldots + c_n v_n$, where

$$c_i = \langle p_0, v_i \rangle \quad \text{for} \quad 1 \leq i \leq n.$$

Then, $W^t p_0 = W^t\left(\sum_{i=1}^n c_i v_i\right) = \sum_{i=1}^n c_i W^t v_i = \sum_{i=1}^n c_i \lambda_i^t v_i$

The assumption $\lambda_2 < 1$ and $\lambda_n > -1$ implies that $|\lambda_i| < 1$ for $2 \le i \le n$.

Hence, $\lim_{t\to\infty} W^t p_0 = \lim_{t\to\infty} \sum_{i=1}^{n} c_i \lambda_i^t v_i = c_1 v_1$.

In the regular case, we have $v_1 = \frac{\vec{1}}{\sqrt{n}}$, and thus $c_1 = \langle p_0, \frac{\vec{1}}{\sqrt{n}} \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} p_0(i) = \frac{1}{\sqrt{n}}$ as p_0 is a probability distribution.

It follows that $\lim_{t\to\infty} W^t p_0 = c_1 v_1 = \frac{1}{\sqrt{n}} \cdot \frac{\vec{1}}{\sqrt{n}} = \frac{\vec{1}}{n}$. □

Mixing time

First, we quantify how fast p_t converges to the limiting 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 $d_{TV}(p_t, \pi) \le \frac{1}{4}$, where π denotes the limiting distribution (in the regular case, $\pi = \frac{\vec{1}}{n}$).

To bound the mixing time, we use the same approach and it should be clear that λ_2 and $|\lambda_n|$ should be bounded away from one for λ_i^t converges to zero quickly for $2 \le i \le n$.

Define $\lambda := \min\{1 - \lambda_2, 1 - |\lambda_n|\}$ as the spectral gap of W.

<u>Claim</u> The mixing time is $O(\frac{\ln n}{\lambda})$.

<u>proof</u> Recall from the previous claim that $W^t p_0 = \frac{\vec{1}}{n} + \sum_{i=2}^{n} c_i \lambda_i^t v_i$.

So, $d_{TV}(W^t p_0, \pi) = \|W^t p_0 - \frac{\vec{1}}{n}\|_1 = \|\sum_{i=2}^{n} c_i \lambda_i^t v_i\|_1 \le \sqrt{n} \|\sum_{i=2}^{n} c_i \lambda_i^t v_i\|_2$ by Cauchy-Schwarz.

Since v_i are orthonormal, we have $\|\sum_{i=2}^{n} c_i \lambda_i^t v_i\|_2^2 = \sum_{i=2}^{n} c_i^2 \lambda_i^{2t} \le (1-\lambda)^{2t} \sum_{i=2}^{n} c_i^2$ by spectral gap.

Note that $\sum_{i=2}^{n} c_i^2 \le \sum_{i=1}^{n} c_i^2 = \|p_0\|_2^2 \le \|p_0\|_1^2 = 1$, and thus $d_{TV}(W^t p_0, \pi) \le \sqrt{n}(1-\lambda)^{2t} \le \sqrt{n} e^{-2\lambda t}$.

Therefore, by setting $t = O(\frac{\ln n}{\lambda})$, we have $d_{TV}(W^t p_0, \pi) \le \frac{1}{4}$ (in fact $\le \frac{1}{\text{poly}(n)}$). □

For lazy random walks, recall that $\lambda_n \ge 0$ and so λ is simply $1 - \lambda_2$.

We know from cheeger's inequality that $\lambda \ge \Omega(\phi(G)^2)$, and it follows that the mixing time of the lazy random walk is $O(\frac{\ln n}{\phi^2})$.

This gives us a combinatorial condition that guarantees fast mixing.

In particular, for expander graphs (i.e. $\phi(G) = \Omega(1)$), the mixing time is $O(\ln n)$, i.e. after only $O(\log n)$ steps, the random walk reaches an almost uniformly random vertex.

This is very useful in designing random sampling algorithms.

For the purpose of (uniform) random sampling, the analysis for regular graphs is usually enough, as we can design the Markov chain (possibly adding self-loops) to make the underlying graph regular.

Analysis for general undirected graphs

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 spectrum theorem in L01 does not apply directly.

Spectrum

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

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

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

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

Then $D^{\frac{1}{2}}v_i$ is an eigenvector of W with eigenvalue $\frac{1}{2}(1+\alpha_i)$ for lazy random walk and $\frac{1}{2}\alpha_i$ for non-lazy, as for instance $W(D^{\frac{1}{2}}v_i) = D^{\frac{1}{2}}\left(\frac{1}{2}I + \frac{1}{2}\mathcal{A}\right)D^{-\frac{1}{2}}(D^{\frac{1}{2}}v_i) = D^{\frac{1}{2}}\left(\frac{1}{2}I + \frac{1}{2}\mathcal{A}\right)v_i = \frac{1}{2}(1+\alpha_i)(D^{\frac{1}{2}}v_i)$.

Note that $D^{\frac{1}{2}}v_1, \ldots, D^{\frac{1}{2}}v_n$ are linearly independent as $v_1, \ldots, v_n$ are linearly independent and D is non-singular (as the graph is connected and so $D_{ii} > 0 \ \forall i$).

Therefore, if $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of W, then $\lambda_i = \frac{1}{2}(1+\alpha_i)$ for lazy random walks and $\lambda_i = \alpha_i$ for non-lazy random walks.

And the initial distribution p_0 can be written as $\sum_{i=1}^{n} c_i D^{\frac{1}{2}}v_i$.

With this setup, the remaining proof is similar to that in the regular case, with suitable modifications. It would be good to stop here and work out the details on your own.

Uniqueness

Let W be the walk matrix and $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be its eigenvalues with $\lambda_i = \frac{1}{2}(1+\alpha_i)$.

Uniqueness

Let W be the walk matrix and $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be its eigenvalues with $\lambda_i = \frac{1}{2}(1+\alpha_i)$.

Claim If $\lambda_2 < 1$ and $\lambda_n > -1$, then $\lim_{t\to\infty} W^t p_0 = \frac{\vec{d}}{2m}$ for any p_0, where $\vec{d}$ is the degree vector with $\vec{d}(i) = \deg(i)$ and m is the number of edges in the graph.

Proof We write $p_0 = \sum_{i=1}^n c_i D^{\frac{1}{2}} v_i$ where $v_1, \ldots, v_n$ are the orthonormal basis of eigenvectors of $\mathcal{A}$.

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

Since $\lambda_2 < 1$ and $\lambda_n > -1$, we have $\lim_{t\to\infty} W^t p_0 = c_1 D^{\frac{1}{2}} v_1$.

For the normalized adjacency matrix $\mathcal{A}$, note that $\mathcal{A}(D^{\frac{1}{2}}\vec{1}) = \left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right)\left(D^{\frac{1}{2}}\vec{1}\right) = D^{-\frac{1}{2}} A \vec{1} = D^{-\frac{1}{2}}\vec{d} = D^{\frac{1}{2}}\vec{1}$,

and so $D^{\frac{1}{2}}\vec{1}$ is an eigenvector with eigenvalue 1, and thus $v_1 = D^{\frac{1}{2}}\vec{1} / \|D^{\frac{1}{2}}\vec{1}\|_2 = D^{\frac{1}{2}}\vec{1}/\sqrt{2m}$.

To compute c_1, note that $p_0 = \sum_{i=1}^n c_i D^{\frac{1}{2}} v_i$ implies that $D^{-\frac{1}{2}} p_0 = \sum_{i=1}^n c_i v_i$.

Since $\{v_1, \ldots, v_n\}$ is an orthonormal basis, we have $c_1 = \langle D^{-\frac{1}{2}} p_0, v_1 \rangle = \frac{1}{\sqrt{2m}} \langle D^{-\frac{1}{2}} p_0, D^{\frac{1}{2}}\vec{1} \rangle = \frac{1}{\sqrt{2m}}$.

Therefore, $\lim_{t\to\infty} W^t p_0 = c_1 D^{\frac{1}{2}} v_1 = \frac{1}{\sqrt{2m}} D^{\frac{1}{2}}\left(D^{\frac{1}{2}}\vec{1}/\sqrt{2m}\right) = \frac{1}{2m} D\vec{1} = \frac{\vec{d}}{2m}$. □

Mixing time

Define $\lambda := \min\{1-\lambda_2, 1-|\lambda_n|\}$ as the spectral gap of W.

Claim The mixing time is $O\left(\frac{\ln n}{\lambda}\right)$.

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

From the proof above, $W^t p_0 = \pi + \sum_{i=2}^n c_i \lambda_i^t D^{\frac{1}{2}} v_i$, and we would like to bound $\|W^t p_0 - \pi\|_1$.

As in the regular case, we bound $\|W^t p_0 - \pi\|_1 \leq \sqrt{n}\,\|W^t p_0 - \pi\|_2$ by Cauchy-Schwarz.

To bound $\|W^t p_0 - \pi\|_2$, we take advantage of the orthonormality of $v_1, \ldots, v_2$, and bound

$$\left\|D^{-\frac{1}{2}}(W^t p_0 - \pi)\right\|_2^2 = \left\|\sum_{i=2}^n c_i \lambda_i^t v_i\right\|_2^2 = \sum_{i=2}^n c_i^2 \lambda_i^{2t} \leq (1-\lambda)^{2t} \sum_{i=2}^n c_i^2.$$

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

Note also that $\|D^{-\frac{1}{2}}(W^t p_0 - \pi)\|_2 \geq \|D^{-\frac{1}{2}}\|\,\|W^t p_0 - \pi\|_2 \geq \frac{1}{\sqrt{d_{\max}}}\|W^t p_0 - \pi\|_2$, where $d_{\max}$ is max degree.

Combining, we have $\|W^t p_0 - \pi\|_1 \leq \sqrt{n}\,\|W^t p_0 - \pi\|_2 \leq \sqrt{n}\sqrt{d_{\max}}\,\|D^{-\frac{1}{2}}(W^t p_0 - \pi)\|_2 \leq \sqrt{n}\sqrt{\frac{d_{\max}}{d_{\min}}}(1-\lambda)^t$.

By setting $t = O\left(\frac{\log n}{\lambda}\right)$, we have $d_{TV}(W^t p_0 - \pi) \ll \frac{1}{4}$. □

The spectral approach can be extended to prove the fundamental theorem for directed graphs as well, but it is more involved and requires the Perror-Frobenius theorem and Jordan normal form.

Random sampling

An important application of random walks is in random sampling.

Consider the problem of generating a random spanning tree of an undirected graph.

The following is a simple algorithm:

- We start from an arbitrary spanning tree T_0.
- In each step $i \geq 0$, we randomly delete an edge e from T_i and randomly add an edge f that connects the two components of $T_i - e$, and set $T_{i+1} = T_i - e + f$.
- Return T_t for a large enough t.

To analyze this algorithm, it is equivalent to analyzing the random walk on a huge "spanning tree exchange graph", in which there is a vertex for each spanning tree, and two vertices have an edge if the corresponding trees T_1 and T_2 can be obtained from a local operation (i.e. $T_1 = T_2 - e + f$ for some edges e, f in the input graph).

If the original graph has n vertices, this exchange graph could have $\Omega(n^{\frac{n}{2}})$ vertices.

To prove that $t = O(\text{poly}(n))$ would work to return an almost uniform random spanning tree, we must prove that the random walk in the huge graph mixes in time logarithmic in the huge graph size.

That is, we need to prove that the spanning tree exchange graph is an expander graph.

It is usually a difficult task.

There are different approaches to prove fast mixing, one is called the coupling method, and another is called the canonical path (or multicommodity flow) method.

We won't discuss these in this course, but we will soon see a method based on high dimensional expansion, which will give a proof that the spanning tree exchange graph is an expander.

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

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