Chapter 6

Random Walks on Graphs

Given an undirected graph $G = (V, E)$, a random walk is a simple stochastic process where it starts from a vertex, and in each step the walk moves to a uniformly random neighbor of the current vertex. Some of the basic mathematical questions about random walks are:

1. (Stationary Distribution:) Is there a limiting distribution of the random walks?
2. (Mixing Time:) How long does it take for the current distribution to be close to the limiting distribution?
3. (Hitting Time:) Starting from a vertex s, how long does it take to first reach a vertex t?
4. (Cover Time:) How long does it take to reach every vertex in the graph at least once?

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 transition matrix. We study the spectral approach in this chapter and refer the reader to [Häg02, LPW06] for the coupling approach.

Questions (3) and (4) are best answered by viewing the graph as an electrical network. This is a topic that we used to study but probably not in this offering. We refer the reader to [DS84, AF02, LP16, Spi19] for the interesting connection between random walks and electrical networks.

6.1 Markov Chains

In this section, we consider the more general setting of a finite Markov chain and state the fundamental theorem.

A finite Markov chain is defined by a finite state space and a transition matrix.

Definition 6.1 (Transition Matrix)**.** *Let $[n]$ be the state space. A matrix $P \in \mathbb{R}^{n\times n}$ is a probability transition matrix if P is non-negative and $\sum_{j\in[n]} P_{i,j} = 1$ for each $i \in [n]$. For $1 \le i, j \le n$, the entry P_{ij} is the transition probability from state i to state j.*

Definition 6.2 (Markov Chain)**.** *A sequence of random variables $(X_0, X_1, \ldots)$ is a Markov chain with state space $[n]$ and transition matrix $P \in \mathbb{R}^{n\times n}$ if for all $i, j \in [n]$ and $t \ge 1$,*

$$\Pr\left[X_{t+1} = j \mid X_t = i \cap X_{t-1} = i_{t-1} \cap \ldots \cap X_0 = i_0\right] = \Pr\left[X_{t+1} = j \mid X_t = i\right] = P(i, j).$$

The first equality is called the Markov property, which states that the transition probability from i to j is the same regardless of the states $X_0, \ldots, X_{t-1}$ that precedes the current state X_t.

We can simply think of a Markov chain as a random walk on a weighted directed graph $G = ([n], w)$, where the transition probability from state i to state j is proportional to the edge weight $w(i, j)$ such that $P_{i,j} = w(i, j)/\sum_{j \in [n]} w(i, j)$.

Irreducibility and Aperiodicity

The following are two properties that will imply the existence of a unique limiting distribution.

Definition 6.3 (Irreducibility)**.** *A Markov chain defined by transition matrix $P \in \mathbb{R}^{n \times n}$ is called irreducible if for any two states i, j, there exists an integer t such that $\Pr[X_t = j \mid X_0 = i] > 0$.*

An equivalent definition is that the underlying directed graph $G = ([n], E)$ where $E(G) := \{ij \mid P_{i,j} > 0\}$ of P is strongly connected.

This property is called irreducibility, because if it is not satisfied then the Markov chain is reducible to a smaller one for the study of the limiting distribution, as the limiting distribution if exists will only have support on a strongly connected component.

Definition 6.4 (Aperiodicity)**.** *The period of a state i is defined as $\gcd\{t \mid \Pr[X_t = i \mid X_0 = i] > 0\}$, the greatest common divisor of the set of times when it is possible to return to the starting state i. A state i is called aperiodic if its period is equal to 1. A Markov chain is called aperiodic if all states are aperiodic; otherwise it is called periodic.*

For examples, random walks on an undirected bipartite graph is periodic as every state has period 2, and random walks on a directed cycle of length $k > 1$ is periodic as every state has period k. In general, there is no limiting distribution if the Markov chain is periodic.

Irreducibility and aperiodicity together imply the following property that after enough number of steps, the probability to transit from any state to any other state is positive.

Proposition 6.5 (Reachability)**.** *For any finite, irreducible, and aperiodic Markov chain, there exists an integer $\tau < \infty$ such that $\Pr[X_t = j \mid X_0 = i] > 0$ for all i, j and all $t \geq \tau$.*

The proof uses aperiodicity and a simple number-theoretic argument to prove the statement for all $i = j$, and then uses irreducibility to prove the statement for all $i \neq j$. See [Häg02, LPW06] for a proof.

Stationary Distribution and Convergence

Definition 6.6 (Stationary Distribution)**.** *For a Markov chain defined by transition matrix $P \in \mathbb{R}^{n \times n}$, a probability distribution $\vec{\pi} \in \mathbb{R}^n$ is a stationary distribution if $\vec{\pi}P = \vec{\pi}$ when $\vec{\pi}$ is represented as a row vector.*

Informally, a stationary distribution $\vec{\pi}$ is a steady distribution as $\vec{\pi}P^t = \vec{\pi}$ for any $t \geq 1$, and a limiting distribution if exists is a stationary distribution.

We measure how close are two probability distributions by the total variation distance.

Definition 6.7 (Total Variation Distance)**.** *Given two probability distributions $\vec{p}, \vec{q} \in \mathbb{R}^n$, the total variation distance of $\vec{p}$ and $\vec{q}$ is defined as*

$$d_{\mathrm{TV}}(\vec{p}, \vec{q}) := \frac{1}{2} \sum_{i=1}^{n} |p(i) - q(i)| = \frac{1}{2} \|\vec{p} - \vec{q}\|_1.$$

Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n \times n}$. Let $\vec{p}_0 \in \mathbb{R}^n$ be an initial probability distribution on the states. Let $\vec{p}_t \in \mathbb{R}^n$ be the probability distribution $\vec{p}_t := \vec{p}_0 P^t$ after t steps of random walks. We say that $\vec{p}_t$ converges to a probability distribution $\vec{q}$ as $t \to \infty$ if $\lim_{t \to \infty} d_{\mathrm{TV}}(\vec{p}_t, \vec{q}) = 0$.

Fundamental Theorem of Markov Chains

One more definition before we state the fundamental theorem.

Definition 6.8 (Hitting Time)**.** *For a Markov chain defined by transition matrix $P \in \mathbb{R}^{n \times n}$, the hitting time from state i to state j is defined as*

$$H_{i,j} := \min\{t \geq 0 \mid X_t = j, X_0 = i\}.$$

The first return time to state i is defined as $H^+_{i,i} := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$.

Given any finite, irreducible, and aperiodic Markov chain defined by P, after we run it long enough, then it is possible to reach any state from any other state by Proposition 6.5. If two Markov chains $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$ of P meet at the same state at some time t such that $X_t = Y_t$, then we cannot distinguish the probability distributions of X_τ and Y_τ for $\tau > t$ anymore as the Markov chains forget about the history. By Proposition 6.5, any two Markov chains of P will eventually meet, and so they will converge to the same distribution as $t \to \infty$, and thus a unique limiting distribution exists. This is the intuition of the coupling proof of the following fundamental theorem.

Theorem 6.9 (Fundamental Theorem of Markov Chains)**.** *Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n \times n}$. Let $\vec{p}_0 \in \mathbb{R}^n$ be an initial probability distribution on the states. Let $\vec{p}_t \in \mathbb{R}^n$ be the probability distribution $\vec{p}_t := \vec{p}_0 P^t$ after t steps. If the Markov chain is finite, irreducible, and aperiodic, then the followings hold.*

1. *There exists a stationary distribution $\vec{\pi}$.*
2. *The distribution $\vec{p}_t$ converges 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. *$\pi(i) = \lim_{t \to \infty} (P^t)_{i,i} = (\mathbb{E}[H^+_{i,i}])^{-1}$, the inverse of the expected first return time to i.*

We will see a spectral proof of this theorem in the special case of random walks on undirected graphs. For the general result, see [Häg02] for a probabilistic proof using coupling, [LPW06] for a probabilistic and algebraic proof, and [HJ13] for a purely algebraic proof related to the Perron-Frobenius Theorem 2.16.

6.2 Random Walks on Undirected Graphs

We consider random walks on an unweighted undirected graph $G = (V, E)$ with $V = [n]$, where in each step the walk moves to a uniformly random neighbor of the current vertex. The fundamental theorem becomes simpler and easier in this special case, as there are simple characterizations of irreducibility and aperiodicity and also the limiting distribution.

Matrix Formulation: The transition probability P_{ij} from a vertex i to a vertex j is simply $1/\deg(i)$, and so the transition matrix is $P = D^{-1}A$ where D is the diagonal degree matrix in Definition 3.13 and A is the adjacency matrix in Definition 3.1. Let $p_0 : V \to \mathbb{R}$ be an initial probability distribution, and p_t be the probability distribution after t steps of random walks. By definition, $p_{t+1}(i) := \sum_{j:ij\in E} p_t(j)/\deg(j)$ for all $1 \le i \le n$. Note that these equations can be written compactly as $p_{t+1} = P^T p_t = (AD^{-1})p_t$, and by induction $p_t = (AD^{-1})^t p_0$. We remark that it is common to write a probability distribution as a row vector, but in these notes we use the convention that p_t is a column vector.

Stationary Distribution: Recall that a probability distribution $\pi : V \to \mathbb{R}$ is a stationary distribution of P if $P^T\pi = \pi$. It is equivalent to saying that π is an eigenvector of P^T with eigenvalue 1. Given that $P = D^{-1}A$ for random walks on undirected graphs, it is not difficult to identify one such eigenvector.

Lemma 6.10 (Stationary Distribution of Undirected Graphs)**.** *Let $G = (V, E)$ be an undirected graph and $P = D^{-1}A$ be its transition matrix. The distribution $\pi : V \to \mathbb{R}$ with*

$$\pi(i) = \frac{\deg(i)}{\sum_{j\in V}\deg(j)} = \frac{\deg(i)}{2|E|}$$

for all $i \in V$ is a stationary distribution of P.

Irreducibility: Is π in Lemma 6.10 the unique 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). This corresponds to the irreducibility condition in the fundamental theorem. For undirected graphs, the irreducibility condition is simply equivalent to the condition that the graph is connected.

Aperiodicity: Even if the graph is connected, a limiting distribution may not exist. For example, in a connected bipartite graph, if the initial distribution p_0 is on a single vertex, then the distribution p_t depends on the parity of t, as the support of p_t oscillates between the two sides of the bipartite graph. This corresponds to the aperiodicity condition in the fundamental theorem. For connected undirected graphs, observe that the aperiodicity condition is equivalent to the condition that the graph is non-bipartite.

Fundamental Theorem: Given the simple characterizations of the conditions in the fundamental theorem, it reduces to the following statement for undirected graphs.

Theorem 6.11 (Fundamental Theorem for Undirected Graphs)**.** *Let G be a connected, non-bipartite undirected graph. Let $P = D^{-1}A$ be the transition matrix of random walks on G. The distribution π in Lemma 6.10 is the unique stationary distribution. Furthermore, $p_t := (P^T)^t p_0$ converges to π as $t \to \infty$ regardless of the initial distribution p_0.*

Lazy Random Walks: The non-bipartiteness condition is to ensure that the Markov chain is aperiodic. There is a simple modification of the random walks so that this assumption can be removed, by adding self-loops in the graph.

Definition 6.12 (Lazy Random Walks)**.** *Let G be an undirected graph. The transition matrix W of the lazy random walks is defined as $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$. In words, the lazy random walks stay at the current vertex with probability $1/2$ and moves to a uniform random neighbor of the current vertex with probability $1/2$.*

By doing the lazy random walks, we make the graph non-bipartite and obtain the following corollary of Theorem 6.11.

Corollary 6.13 (Fundamental Theorem for Lazy Undirected Graphs)**.** *Let G be a connected undirected graph. Let $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$ be the transition matrix of lazy random walks on G. The distribution π in Lemma 6.10 is the unique stationary distribution. Furthermore, $p_t := (W^T)^t p_0$ converges to π as $t \to \infty$ regardless of the initial distribution p_0.*

It will be clear from the spectral analysis why the constant $1/2$ is used.

6.3 Spectral Analysis of Mixing Time for Undirected Graphs

In this section, we will prove the fundamental theorem for undirected graphs in Theorem 6.11 using a spectral analysis. Besides that the spectral analysis is elegant, it has the very important feature that it can also be used to analyze the mixing time, which is the rate of convergence to the unique stationary distribution.

As in chapter 4 for Cheeger's inequality, we will first assume the given undirected graph is d-regular and prove Theorem 6.11 and then define and bound the mixing time. After that, we outline the modifications needed for general undirected graphs.

Spectrum for Regular Graphs

For d-regular graphs, the transition matrix P for random walks is simply $P = D^{-1}A = \frac{1}{d}A = \mathcal{A}$, the normalized adjacency matrix. And the transition matrix W for lazy random walks is $W = \frac{1}{2}I + \frac{1}{2}\mathcal{A}$. This is the main simplification from the d-regular assumption, as the matrices P and W are still real symmetric. Another simplification is that for d-regular graphs the stationary distribution π in Lemma 6.10 is simply the uniform distribution $\vec{1}/n$.

Our goal is to prove that $\lim_{t\to\infty} P^t p_0 = \vec{1}/n$ regardless of the initial distribution p_0, as long as the graph is connected and non-bipartite. And, similarly, $\lim_{t\to\infty} W^t p_0 = \vec{1}/n$, as long as the graph is connected.

To compute $P^t p_0$ and $W^t p_0$, a repeated application of the same operator, it is very helpful to know the spectrum of the matrices P and W as discussed in chapter 2. Let $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of $\mathcal{A}$ and $v_1, \ldots, v_n$ be the corresponding orthonormal eigenvectors. Recall that

- $\alpha_1 = 1$ and $v_1 = \vec{1}/\sqrt{n}$ from Lemma 3.21,
- $\alpha_2 < 1$ if and only if G is connected from Proposition 3.18,
- and when G is connected, $\alpha_n > -1$ if and only if G is non-bipartite from Problem 3.7.

For the lazy random walk matrix W, the spectrum is $\frac{1}{2}(1+\alpha_1) \geq \frac{1}{2}(1+\alpha_2) \geq \ldots \geq \frac{1}{2}(1+\alpha_n)$, which implies that the smallest eigenvalue is at least $0 > -1$. This is why the non-bipartitness assumption can be removed when we consider lazy random walks.

Limiting Distribution

We translated the combinatorial conditions in the fundamental theorem into spectral conditions, and the fundamental Theorem 6.11 for d-regular undirected graphs can be restated as follows and the proof is relatively straightforward.

Proposition 6.14 (Limiting Distribution for Regular Graphs). *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $P = \mathcal{A}$ be the transition matrix of random walks on G and $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be its eigenvalues. If $\alpha_2 < 1$ and $\alpha_n > -1$, then $\lim_{t\to\infty} P^t p_0 = \vec{1}/n$.*

Proof. Let $v_1, v_2, \ldots, v_n$ be the corresponding orthonormal eigenvectors. For any initial distribution p_0, as $v_1, \ldots, v_n$ forms an orthonormal basis, we can write $p_0 = c_1 v_1 + \ldots + c_n v_n$ as a linear combination of $v_1, \ldots, v_n$, where $c_i = \langle p_0, v_i \rangle$ for $1 \leq i \leq n$. Then,

$$P^t p_0 = \mathcal{A}^t\Big(\sum_{i=1}^n c_i v_i\Big) = \sum_{i=1}^n c_i \mathcal{A}^t v_i = \sum_{i=1}^n c_i \alpha_i^t v_i.$$

The assumptions $\alpha_2 < 1$ and $\alpha_n > -1$ imply that $|\alpha_i| < 1$ for $2 \leq i \leq n$. Hence,

$$\lim_{t\to\infty} P^t p_0 = \lim_{t\to\infty} \sum_{i=1}^n c_i \alpha_i^t v_i = c_1 v_1,$$

as all but the first term go to zero as $t \to \infty$. Recall that in the d-regular case, $v_1 = \vec{1}/\sqrt{n}$ and thus $c_1 = \langle p_0, \vec{1}/\sqrt{n} \rangle = 1/\sqrt{n}$ as p_0 is a probability distribution. Therefore, we conclude that

$$\lim_{t\to\infty} P^t p_0 = c_1 v_1 = \frac{1}{\sqrt{n}} \cdot \frac{\vec{1}}{\sqrt{n}} = \frac{\vec{1}}{n}.$$

□

Basically, the proof says that under the assumption $|\alpha_i| < 1$ for $2 \leq i \leq n$, $P^t p_0$ converges to the first eigenvector which is proportional to the all-one vector.

Check that the fundamental theorem for lazy random walks in Corollary 6.13 hold for d-regular graphs as well.

Mixing Time

The mixing time is to quantify how fast p_t converges to the limiting distribution. The following definition is for general Markov chains.

Definition 6.15 (Mixing Time)**.** *Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n\times n}$. Let $p_0 \in \mathbb{R}^n$ be an initial probability distribution on the states. Let $p_t \in \mathbb{R}^n$ be the probability distribution $p_t := (P^T)^t p_0$ after t steps. Suppose that the limiting distribution $\pi = \lim_{t\to\infty} p_t$ exists. For any $0 < \epsilon \leq 1$, the ϵ-mixing time $\tau_\epsilon(P)$ of P is defined as the smallest t such that $d_{\mathrm{TV}}(p_t, \pi) \leq \epsilon$ for any initial distribution p_0, where d_{TV} is the total variation distance in Definition 6.7.*

To bound the mixing time, we use the same approach and assume that α_2 and $|\alpha_n|$ are bounded away from one for α_i^t to converge to zero quickly for $2 \leq i \leq n$.

Theorem 6.16 (Bounding Mixing Time by Spectral Gap)**.** *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $P = \mathcal{A}$ be the transition matrix of random walks on G and $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be its eigenvalues. Let $g := \min\{1 - \alpha_2, 1 - |\alpha_n|\}$ be the spectral gap. Then the ϵ-mixing time of P is*

$$\tau_\epsilon(P) \lesssim \frac{1}{g} \ln\left(\frac{n}{\epsilon}\right).$$

Proof. Recall from Proposition 6.14 that $P^t p_0 = \frac{\vec{1}}{n} + \sum_{i=2}^n c_i \alpha_i^t v_i$ where $v_1, \ldots, v_n$ are the orthonormal eigenvectors, and the limiting distribution is $\pi = \vec{1}/n$. So,

$$d_{\mathrm{TV}}(p_t, \pi) = d_{\mathrm{TV}}(P^t p_0, \pi) = \frac{1}{2}\left\| P^t p_0 - \frac{\vec{1}}{n} \right\|_1 = \frac{1}{2}\left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_1 \lesssim \sqrt{n} \left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_2.$$

Since $v_1, \ldots, v_n$ are orthonormal, it follows that from the spectral gap assumption that

$$\left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_2^2 = \sum_{i=2}^n c_i^2 \alpha_i^{2t} \leq (1-g)^{2t} \sum_{i=2}^n c_i^2.$$

Note that $\sum_{i=2}^n c_i^2 \leq \sum_{i=1}^n c_i^2 = \|p_0\|_2^2 \leq \|p_0\|_1^2 = 1$, and thus

$$d_{\mathrm{TV}}(p_t, \pi) \lesssim \sqrt{n} \left\| \sum_{i=2}^n c_i \alpha_i^t v_i \right\|_2 \leq \sqrt{n(1-g)^{2t} \sum_{i=2}^n c_i^2} \leq \sqrt{n}(1-g)^t \leq \sqrt{n} e^{-gt}.$$

Therefore, by setting $t \gtrsim \frac{1}{g} \ln(\frac{n}{\epsilon})$, we have $d_{\mathrm{TV}}(p_t, \pi) \leq \epsilon$ for any initial distribution p_0. □

For the lazy random walk matrix W, recall that the smallest eigenvalue is at least 0, and so the spectral gap for W is simply $g = \frac{1}{2}(1 - \alpha_2)$. The following is an important corollary from Cheeger's inequality in Theorem 4.3 that $1 - \alpha_2 \gtrsim \phi(G)^2$

Corollary 6.17 (Bounding Mixing Time by Conductance)**.** *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $W = \frac{1}{2}I + \frac{1}{2}\mathcal{A}$ be the transition matrix of lazy random walks on G. Then the ϵ-mixing time of W is*

$$\tau_\epsilon(W) \lesssim \frac{1}{\phi(G)^2} \ln\left(\frac{n}{\epsilon}\right).$$

This provides a combinatorial condition that guarantees fast mixing. In particular, for expander graph where $\phi(G) = \Omega(1)$, the mixing time of lazy random walks is $O(\ln n)$, such that the random walk reaches an almost uniformly random vertex after $O(\ln n)$ steps. For many applications, it is important that the mixing time is polylogarithmic in the graph size, as we will discuss some examples soon.

Corollary 6.17 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 by adding self-loops) so that the underlying graph is regular.

Spectrum for General Graphs

The random walk matrix for general graphs is $P = D^{-1}A$ and the lazy random walk matrix is $W = \frac{1}{2}I + \frac{1}{2}P$. The main difference is that these matrices are not necessarily symmetric, and so we cannot directy apply the spectral theorem in Theorem 2.5 to reason about their eigenvalues and eigenvectors. A simple but important observation is that P and W are similar to a real symmetric matrix (see Definition 2.3 for the definition of similar matrices), and so the eignevalues of P and W are still all real numbers.

Lemma 6.18 (Spectrum of Random Walk Matrices)**.** *Let $G = (V, E)$ be a connected undirected graph with $V = [n]$ and $\mathcal{A}$ be its normalized adjacency matrix. Let the eigenvalues of $\mathcal{A}$ be $\alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n$ and $v_1, v_2, \ldots, v_n$ be the corresponding orthonormal basis of eigenvectors. Then the eigenvalues of the random walk matrix $P = D^{-1}A$ are $\alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n$, and the corresponding eigenvectors of $P^T = AD^{-1}$ are $D^{\frac{1}{2}}v_1, D^{\frac{1}{2}}v_2, \ldots, D^{\frac{1}{2}}v_n$. Also, the eigenvalues of the lazy random walk matrix $W = \frac{1}{2}I + \frac{1}{2}P$ are $\frac{1}{2}(1+\alpha_1) > \frac{1}{2}(1+\alpha_2) \geq \ldots \geq \frac{1}{2}(1+\alpha_n)$, and the corresponding eigenvectors of W^T are $D^{\frac{1}{2}}v_1, D^{\frac{1}{2}}v_2, \ldots, D^{\frac{1}{2}}v_n$.*

Proof. Note that $P = D^{-1}A = D^{-\frac{1}{2}}(D^{-\frac{1}{2}}AD^{-\frac{1}{2}})D^{\frac{1}{2}} = D^{-\frac{1}{2}}\mathcal{A}D^{\frac{1}{2}}$, and so P is similar to $\mathcal{A}$, as D is non-singular when the graph is connected. By the same argument, W is similar to $\frac{1}{2}I + \frac{1}{2}\mathcal{A}$. So, by Fact 2.4, P and $\mathcal{A}$ have the same spectrum, and W and $\frac{1}{2}I + \frac{1}{2}\mathcal{A}$ have the same spectrum. Note that $D^{\frac{1}{2}}v_i$ is an eigenvector of P^T with eigenvalue α_i, as $P^T\big(D^{\frac{1}{2}}v_i\big) = \big(D^{\frac{1}{2}}\mathcal{A}D^{-\frac{1}{2}}\big)\big(D^{\frac{1}{2}}v_i\big) = D^{\frac{1}{2}}\mathcal{A}v_i = \alpha_i\big(D^{\frac{1}{2}}v_i\big)$, and similarly $D^{\frac{1}{2}}v_i$ is an eigenvector of W_T with eigenvalue $\frac{1}{2}(1+\alpha_i)$. □

Note that $D^{\frac{1}{2}}v_1, \ldots, D^{\frac{1}{2}}v_n$ are linearly independent as D is non-singular for a connected graph. Therefore, any initial distribution p_0 can be written as $\sum_{i=1}^{n} c_i D^{\frac{1}{2}}v_i$, a linear combination of the eigenvectors of P^T and W^T. With this setup, we can adapt the proof in Proposition 6.14 to prove the following equivalent form of the fundamental theorem for undirected graphs in Theorem 6.11.

Problem 6.19 (Limiting Distribution for Undirected Graphs)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Let $P = D^{-1}A$ be the transition matrix of random walks on G and $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be its eigenvalues. If $\alpha_2 < 1$ and $\alpha_n > -1$, then $\lim_{t\to\infty}(P^T)^t p_0 = \frac{\vec{d}}{2|E|}$ where $\vec{d}$ is the degree vector with $\vec{d}(i) = \deg(i)$ for $1 \leq i \leq n$.*

Then the same approach as in Theorem 6.16 works to bound the mixing time for general undirected graphs. To bound $\|(P^T)^t p_0 - \pi\|_2$, it will be more convenient to bound $\|D^{-\frac{1}{2}}\big((P^T)^t p_0 - \pi\big)\|_2^2$, so as to take advantage of the orthonormality of $v_1, \ldots, v_n$. Then, one can bound $\|(P^T)^t p_0 - \pi\|_2 \leq \sqrt{n}\sqrt{\frac{d_{\max}}{d_{\min}}}(1-g)^t$ where g is the spectral gap, and extend the results for d-regular graphs to general undirected graphs.

Problem 6.20 (Bounding Mixing Time)**.** *Prove Theorem 6.16 and Corollary 6.17 for general undirected graphs.*

The same arguments work for weighted undirected graphs basically without modifications, by generalizing the definitions as in section 4.4.

We remark that this spectral approach can be extended to prove the fundamental theorem for directed graphs as well, but it is more involved and requires the Perron-Frobenius theorem and the Jordan normal form (see [HJ13] for proofs).

6.4 Random Sampling

An important application of random walks is in random sampling. Consider the following algorithm for generating a random spanning tree of an undirected graph.

Algorithm 3 Random Exchange Algorithm for Sampling Random Spanning Trees

Require: An undirected graph $G = (V, E)$.
1: Compute an arbitrary spanning tree T_0 of the graph.
2: **for** $1 \leq t \leq \tau$ **do**
3: Choose a uniform random edge $e \in E - T_{t-1}$ to T_{t-1}.
4: Choose a uniform random edge f in the unique cycle in $T_{t-1} + e$.
5: Set $T_t := T_{t-1} + e - f$.
6: **end for**
7: **return** T_τ.

To analyze this algorithm, it is equivalent to analyzing the random walks on a huge "spanning tree exchange graph" H, in which there is a vertex in H for each spanning tree of the graph, and two vertices in H have an edge if the corresponding spanning trees T and T' can be obtained from one step of the algorithm (i.e. $T' = T + e - f$ for some edges e, f in the input graph).

Note that if the original graph G has n vertices, then this exchange graph H could have as many as $\Omega(n^{n-2})$ vertices. So, to prove that $\tau \lesssim \text{poly}(n)$ would work to return an almost uniform random spanning tree, we must prove that the random walks in the exchange graph mix in polylogarithmic time in the size of H. In other words, we need to prove that the spanning tree exchange graph H is an expander graph.

This is usually a difficult task. There are different approaches to prove fast mixing of Markov chains. One is called the coupling method, which is the most common and versatile probabilistic technique in bounding mixing time (see [LPW06]). Another is called the canonical path method, which is based on using multicommodity flow to lower bound the graph conductance so as to use Corollary 6.17 to upper bound the mixing time. A very important application of the canonical path method is to approximate the permanent of a non-negative matrix [JSV04], which is equivalent to counting the number of perfect matchings in a bipartite graph.

We won't discuss these methods in this course, but we will see how to analyze the random exchange algorithm for sampling random spanning trees using the new techniques from high dimensional expanders in the third part of the course.

6.5 Local Graph Partitioning

Another useful application of random walks is in graph partitioning. As we have seen in the small-set expansion problem in section 5.2, the random walk distribution $W^t\chi_i$ for some starting vertex i can be used to find a small sparse cut of the graph. This idea was originally proposed by Spielman and Teng [ST13]. They proved that the performance of the random walk algorithm for graph partitioning is similar to that of the spectral partitioning algorithm in chapter 4. Furthermore, the random walk algorithm has the important advantage that it can be implemented *locally*, such that the running time depends only on output size but not on the original graph size, and this provides a sublinear time algorithm for graph partitioning for some instances. This idea can also be used to design approximation algorithms for the small-set expansion problem [KL12]. Local graph partitioning is an active research topic on its own, and there are several other algorithms such as using PageRank vector [ACL06] and evolving sets [AOPT16]. We won't discuss these results further in this course. This is a good project topic for those who are interested in these algorithms.

6.6 Problems

Problem 6.21 (Lower Bounding Mixing Time)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Let $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$ be the transition matrix of lazy random walks on G. Prove that the ϵ-mixing time of W is*

$$\tau_\epsilon(W) \gtrsim \frac{1}{1-\alpha_2} \ln\left(\frac{1}{\epsilon}\right),$$

where α_2 is the second largest eigenvalue of the normalized adjacency matrix $\mathcal{A}(G)$. A simpler problem is to prove that

$$\tau_\epsilon(W) \gtrsim \frac{1}{\phi(G)} \ln\left(\frac{1}{\epsilon}\right),$$

where $\phi(G)$ is the edge conductance of G. You may also consider the special case when G is d-regular.

Problem 6.22 (Page Ranking)**.** *Suppose someone searches a keyword (e.g. "car") and we would like to identify the webpages that are the most relevant for this keyword and the webpages that are the most reliable sources for this keyword (a page is a reliable source if it points to many most relevant pages). First we identify the pages with this keyword and ignore all other pages. Then we run the following ranking algorithm on the remaining pages. Each vertex corresponds to a remaining page, and there is a directed edge from page i to page j if there is a link from page i to page j. Call this directed graph $G = (V, E)$. For each vertex i, we have two values $s(i)$ and $r(i)$, where intentionally $r(i)$ represents how relevant is this page and $s(i)$ represents how reliable it is as a source (the larger the values the better). We start from some arbitrary initial values, say $s(i) = 1/|V|$ for all i, as we have no ideas at the beginning. At each step, we update s and r (where s and r are vectors of $s(i)$ and $r(i)$ values) as follows: First we update $r(i) = \sum_{j:ji\in E} s(j)$ for all i, as a page is more relevant if it is linked by many reliable sources. Then we update $s(i) = \sum_{j:ij\in E} r(j)$ for all i (using the just updated values $r(j)$), as a page is a more reliable source if it points to many relevant pages. To keep the values small, we let $R = \sum_{i=1}^{|V|} r(i)$ and $S = \sum_{i=1}^{|V|} s(i)$, and divide each $s(i)$ by S and divide each $r(i)$ by R. We repeat this step for many times to refine the values.*

Let $s, r \in \mathbb{R}^{|V|}$ be the vectors of the s and r values. Give a matrix formulation for computing s and r. Suppose G is weakly connected (when we ignore the direction of the edges the underlying undirected graph is connected) and there is a self-loop at each vertex. Prove that there is a unique

limiting s and a unique limiting r for any initial s as long as $s \geq 0$ and $s \neq 0$. You may use the Perror-Frobenius Theorem 2.16 to solve this problem.

6.7 References

[ACL06] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. Local graph partitioning using pagerank vectors. In *FOCS*, pages 475–486. IEEE Computer Society, 2006. 54

[AF02] David Aldous and James Allen Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2002. http://www.stat.berkeley.edu/~aldous/RWG/book.html. 45

[AOPT16] Reid Andersen, Shayan Oveis Gharan, Yuval Peres, and Luca Trevisan. Almost optimal local graph clustering using evolving sets. *J. ACM*, 63(2):15:1–15:31, 2016. 54

[DS84] Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks*. Number Book 22 in Carus Mathematical Monographs. Mathematical Association of America, Washington, DC, 1984. 45

[Häg02] Olle Häggström. *Finite Markov Chains and Algorithmic Applications*. London Mathematical Society Student Texts. Cambridge University Press, 2002. 45, 46, 47

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge; New York, 2nd edition, 2013. 12, 47, 53

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004. 53

[KL12] Tsz Chiu Kwok and Lap Chi Lau. Finding small sparse cuts by random walk. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX 2012*, volume 7408 of *Lecture Notes in Computer Science*, pages 615–626. Springer, 2012. 54

[LP16] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016. Available at https://rdlyons.pages.iu.edu/. 45

[LPW06] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006. 45, 46, 47, 53

[Spi19] Daniel A. Spielman. *Spectral and Algebraic Graph Theory*. 2019. 17, 45

[ST13] Daniel A. Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM J. Comput.*, 42(1):1–26, 2013. 54