Spring 2014

Hittingtime and PageRank

Shanthi Kannan
San Jose State University

HITTING TIME AND PAGERANK

A Thesis

Presented to

The Faculty of the Department of Mathematics and Statistics

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Shanthi Kannan

May 2014

The Designated Thesis Committee Approves the Thesis Titled

HITTING TIME AND PAGERANK

by

Shanthi Kannan

APPROVED FOR THE DEPARTMENT OF MATHEMATICS AND STATISTICS

SAN JOSÉ STATE UNIVERSITY

May 2014

Dr. Wasin So	Department of Mathematics and Statistics
Dr. Maurice Stanley	Department of Mathematics and Statistics
Dr. Slobodan Simic	Department of Mathematics and Statistics

ABSTRACT

HITTING TIME AND PAGERANK

by Shanthi Kannan

In this thesis, we study convergence of finite state, discrete, and time homogeneous Markov chains to a stationary distribution. Expressing the probability of transitioning between states as a matrix allows us to look at the conditions that make the matrix primitive. Using the Perron-Frobenius theorem we find the stationary distribution of a Markov chain to be the left Perron vector of the probability transition matrix.

We study a special type of Markov chain — random walks on connected graphs. Using the concept of fundamental matrix and the method of spectral decomposition, we derive a formula that calculates expected hitting times for random walks on finite, undirected, and connected graphs.

The mathematical theory behind Google's vaunted search engine is its PageRank algorithm. Google interprets the web as a strongly connected, directed graph and browsing the web as a random walk on this graph. PageRank is the stationary distribution of this random walk. We define a modified random walk called the lazy random walk and define personalized PageRank to be its stationary distribution. Finally, we derive a formula to relate hitting time and personalized PageRank by considering the connected graph as an electrical network, hitting time as voltage potential difference between nodes, and effective resistance as commute time.

ACKNOWLEDGEMENTS

To my thesis advisor Dr Wasin So, thank you for being a truly exceptional advisor and mentor. Your energetic support and sharp insights have contributed immeasurably both to this thesis and to my mathematical education over this past year.

I am also extremely grateful to Dr Slobodan Simic and Dr Maurice Stanley, my committee members. Your comments and suggestions are invaluable. I would also like to thank the many professors at San Jose State I had the pleasure of learning from over the past few years.

Many thanks to my friends and fellow students at SJSU for your support and camaraderie. The study groups and happy hours were both equally valuable. I miss you all!!

Finally, thanks to my family and friends at large whose encouragement and confidence helped me tremendously throughout this endeavor.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

CHAPTER 1

SUMMARY

A Markov chain on a finite or countable set of states is a stochastic process with the special property of "memorylessness." The sequence of states that the process transitions through is the Markov chain. Given any starting distribution and the probability transition matrix, we easily find the distribution after any number of transitions.

In Chapter 2, we study Markov chains with a single communication class and the additional property of recurrence. Such Markov chains eventually reach a stationary distribution. We relate irreducibility and recurrence of the Markov chain to irreducibility and primitivity of the probability transition matrix. The stationary distribution is then the left Perron vector of the probability transition matrix. Using the fundamental matrix of the Markov chain, we derive a formula for the hitting time between states.

In Chapter 3, we study a special kind of Markov chain, namely random walk on an undirected, unweighted, and connected graph. The main result in this chapter is a formula for the hitting time between vertices of the connected graph. We define a symmetric form of the probability transition matrix and find its spectral decomposition. Using this spectra, we compute the hitting time and commute time between vertices of the graph.

If we look at the World Wide Web as a large but finite connected graph and a user browsing the web as taking a random walk on this graph, then the concepts developed in chapters 2 and 3 are easily applied to this web graph. Google's

PageRank is the left Perron vector of a modified probability transition matrix, the Google matrix, designed to be primitive. In Chapter 4, we also mention a simple iterative algorithm for computing PageRank and how it translates to the large web graph.

The last two chapters draw an intriguing analogy between random walks on connected weighted graphs and electrical networks. In Chapter 5, drawing on the initial work of [AC96] and [PGD06], we relate the vertices of a graph to nodes in an electrical network and the edges of the graph to connectors between electrical nodes. In this context, the flow of electrons is similar to a random walk on the graph. Using harmonic functions, we establish that the voltage potential between nodes and hitting time between vertices are indeed the same function. Finally, in Chapter 6 we follow the work of [FC10] to modify the regular random walk and design a lazy random walk. The personalized PageRank is the stationary distribution of this lazy random walk. We study the normalized Laplacian of the lazy random walk and its inverse, the Green's function. By linking voltage potential in an electrical network to the normalized Laplacian, we derive a direct formula for the hitting time in terms of the personalized PageRank.

CHAPTER 2

MARKOV CHAIN

A Markov chain, named after Andrey Markov (1856-1922), is a random process that transitions from one state to another, among a finite or countable number of possible states. It is a mathematical model for a random process evolving with time, usually characterized as memoryless: the next state depends only on the current state and not on the sequence of states that precede it. We say that the past affects the future only through the present. This specific kind of "memorylessness" is called the Markov property. The time can be discrete (integers), continuous (the real numbers), or a totally ordered set like English words.

Markov chains model many interesting phenomena such as virus mutation, the spread of epidemics, and more. The lack of memory property makes it possible to build probabilistic models and predict how a Markov chain may behave. In our study, we shall focus our attention exclusively on Markov chains with discrete time and a finite set of states. We follow [Nor98] in this chapter.

Example 2.0.1. Consider a mouse in a cage with two cells: cell 1 with ripe cheese and cell 2 with fresh cheese as shown in Figure 2.1. A scientist observes the mouse and records its position every minute. If the mouse is in cell 1 at minute n, then at minute $n+1$ it has either moved to cell 2 or stays in cell 1. Statistical observations led the scientist to conclude that the mouse moved from cell 1 to cell 2 with probability $\alpha = 0.95$. Similarly, when in cell 2 it moved to cell 1 with probability $\beta = 0.01$. As we see, at any time, the mouse decides where to move only based on where it is now and not where it came from.

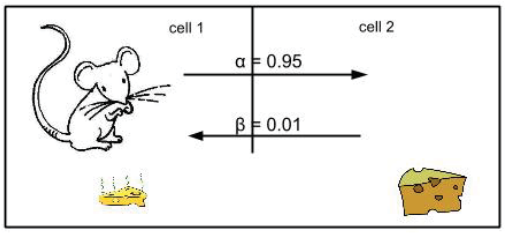

Figure 2.1: Cheesy dilemma

We represent the transition from cell 1 to cell 2 using a *probability transition* matrix P. In this scenario, P is a 2×2 matrix with the rows and columns indexed by 1 and 2 and each entry p_{ij} is the probability of the mouse moving from cell i to cell j. Since the mouse moves from cell 1 to cell 2 with probability $\alpha = 0.95$, it stays in cell 1 with probability $1 - \alpha = 0.05$. Similarly, it stays in cell 2 with probability $1 - \beta = 0.99$. The probability transition matrix is

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} = \begin{bmatrix} 0.05 & 0.95 \\ 0.01 & 0.99 \end{bmatrix}.$$

2.1 Basic definitions and theorems

Definition 2.1.1. Stochastic process.

A *stochastic process* is a sequence of random variables $(X_n)_{n\geq 0}$ having a common range in the finite state space I for the process.

Definition 2.1.2. Stochastic matrix.

A *stochastic matrix* is a nonnegative matrix $[x_{ij}]$ in which each row sum equals 1; $\sum_j x_{ij} = 1$ for every row i of the matrix.

Definition 2.1.3. Markov chain.

A*Markov chain* is a stochastic process $(X_n)_{n\geq 0}$ with an initial probability

distribution μ and probability transition matrix $P = [p_{ij}]$, if it satisfies the following properties:

i. the initial state X_0 has the initial distribution μ; that is

$$\mathbb{P}(X_0 = i) = \mu_i \text{ for all } i \in I.$$

ii. for $n \geq 0$, conditioning on $X_n = i_n$, X_{n+1} has distribution $(p_{i_n i_{n+1}} : i_{n+1} \in I)$ and is independent of $X_0, X_1, \cdots, X_{n-1}$. We write this as

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \cdots, X_n = i_n) = p_{i_n i_{n+1}}. \tag{2.1}$$

We are interested in the special case of *time-homogeneous* Markov chains, which means that the transition probabilities of $p_{ij}(n, n+1)$ do not depend on n. From here on we consider only time-homogeneous and finite state Markov chains. Notation: $Markov(\mu, P)$ represents a Markov chain with initial probability distribution μ and probability transition matrix P.

Definition 2.1.4. ***n*-step Transition probability.**
The probability of transitioning from state i to state j in n time steps is given by

$$\mathbb{P}(X_n = j | X_0 = i) = p_{ij}^{(n)}.$$

$p_{ij}^{(n)}$ is the *n-step transition probability* from i to j.

With these definitions, we are now ready to state our theorems on Markov chains.

Theorem 2.1.5. *(**Markov property**) A stochastic process $(X_n)_{n \geq 0}$ is $Markov(\mu, P)$ if for all states $i_k \in I$, $0 \leq k \leq N$*

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n). \tag{2.2}$$

Proof. Suppose $(X_n)_{n\geq 0}$ is Markov, applying Definition 2.1.3, the claim holds by the conditional independence of (X_{n+1}) and $(X_0, X_1, \cdots, X_{n-1})$ given X_n.

□

Theorem 2.1.6. *A stochastic process $(X_n)_{n\geq 0}$ is Markov(μ, P) if and only if for all states $i_0, i_1, \cdots, i_N \in I$*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}. \tag{2.3}$$

Proof. Suppose $(X_n)_{n\geq 0}$ is Markov(μ, P), then

$$\begin{aligned}&\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n)\\ &\quad = \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdots \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1} \cdots X_0 = i_0).\end{aligned} \tag{2.4}$$

By Markov property in Theorem 2.1.5, we have

$$\mathbb{P}(X_k = i_k | X_{k-1} = i_{k-1} \cdots X_0 = i_0) = \mathbb{P}(X_k = i_k | X_{k-1} = i_{k-1}).$$

Applying this in (2.4), we get

$$\mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdots \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1} \cdots X_0 = i_0)$$

$$= \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdots \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}).$$

The transition probability from state i to j is given by p_{ij}. So, we get

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}.$$

For the reverse, suppose (2.3) holds for all $i \in I$. By induction, we establish that

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}.$$

From probability theory, we know that given two events A and B with $P(A) > 0$, the conditional probability $\mathbb{P}(B|A)$ is given by

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Using this, we write

$$\begin{aligned}\mathbb{P}(X_{n+1} = i_{n+1}|X_0 = i_0, \cdots, X_n = i_n) &= \frac{\mathbb{P}(X_0 = i_0, \cdots, X_{n+1} = i_{n+1})}{\mathbb{P}(X_0 = i_0, \cdots, X_n = i_n)} \\ &= \frac{\mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} p_{i_n i_{n+1}}}{\mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}} \\ &= p_{i_n i_{n+1}}.\end{aligned}$$

So $(X_n)_{n \geq 0}$ is Markov.

□

Theorem 2.1.7. *Suppose $(X_n)_{n \geq 0}$ is Markov(μ, P). Then*

i. $\mathbb{P}(X_n = j) = (\mu^T P^n)_j$.

ii. $\mathbb{P}(X_{n+m} = j | X_m = i) = p_{ij}^{(n)}$ *where m, n are any two positive integers.*

Note: Here $p_{ij}^{(n)}$ refers to the (i, j)th entry of the matrix power P^n.

Proof. i. The probability that $X_n = j$ is the sum of the probability of all possible paths starting at any state, based on the initial probability distribution μ, and navigating to state j after $n - 1$ steps. We write this as

$$\begin{aligned}\mathbb{P}(X_n = j) = \mathbb{P}(X_0 = i_1)\mathbb{P}(X_1 = i_2) \cdots \mathbb{P}(X_n = j) + \\ \cdots + \mathbb{P}(X_0 = i_{n-1}) \cdots \mathbb{P}(X_n = j). \quad (2.5)\end{aligned}$$

Writing this using summations notation and applying Theorem 2.1.6, we get

$$\mathbb{P}(X_n = j) = \sum_{i_1 \in I} \cdots \sum_{i_{n-1} \in I} \mu_{i_1} p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} j} = [\mu^T P^n]_j.$$

ii. The Markov property in Theorem 2.1.5 proves that the future states depend only on the current state and not the states that precede it. Given that $X_m = i$, the probability distribution after step m is $\mu^m = [0, 0, \cdots, 1, 0, 0]$, where 1 is in the i^{th} position. Using (i) above, we get

$$\mathbb{P}(X_{n+m} = j | X_m = i) = [(0, 0, \cdots, 1, 0, 0) \cdot P^n]_j = p_{ij}^{(n)}.$$

We call $p_{ij}^{(n)}$ as the n-step transition probability from state i to j.

□

Lemma 2.1.8. *Chapman-Kolmogorov equation.*

$$p_{ij}^{(m+n)} = \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)}.$$

Proof. Method 1. By Theorem 2.1.7 (ii),

$$\begin{aligned}
p_{ij}^{(m+n)} &= \mathbb{P}(X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in I} \mathbb{P}(X_m = k, X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in I} \mathbb{P}(X_m = k | X_0 = i) \mathbb{P}(X_{m+n} = j | X_m = k, X_0 = i) \\
&= \sum_{k \in I} p_{ik}^{(m)} \mathbb{P}(X_{m+n} = j | X_m = k, X_0 = i) \\
&= \sum_{k \in I} p_{ik}^{(m)} \mathbb{P}(X_{m+n} = j | X_m = k), \text{ since } (X_n)_{n \geq 0} \text{ is Markov} \\
&= \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)}.
\end{aligned}$$

Method 2. By matrix multiplication, we have $P^{m+n} = P^m P^n$. Thus,

$$p_{ij}^{(m+n)} = [P^{(m+n)}]_{ij} = \sum_{k \in I} [P^m]_{ik} [P^n]_{kj} = \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)}.$$

□

Corollary 2.1.9. *Based on the above lemma we have these two results:*

i. $p_{ij}^{(m+n)} \geq p_{ik}^{(m)} p_{kj}^{(n)}$, *for any* $k \in I$.

ii. $p_{ij}^{(a+b+c)} \geq p_{ik}^{(a)} p_{kl}^{(b)} p_{lj}^{(c)}$, *for any* $k, l \in I$.

Example 2.1.10. Consider a three state, $I = \{1, 2, 3\}$, Markov chain (μ, P) as shown in Figure 2.2 with

$$\mu = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right] \quad \textit{and} \quad P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

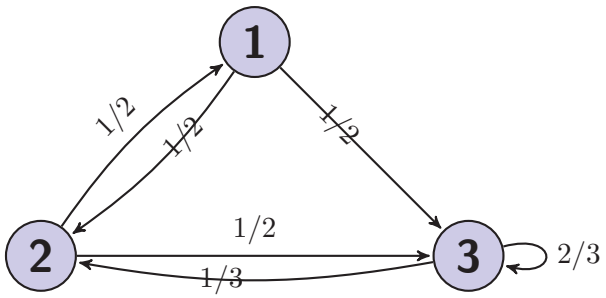

Figure 2.2: State transition diagram.

Using (2.5), we compute the probability that the chain is in state 1 at the second step as

$$\begin{aligned}
\mathbb{P}(X_2 = 1) &= \mathbb{P}(X_0 = 1, X_1 = 2, X_2 = 1) + \mathbb{P}(X_0 = 3, X_1 = 2, X_2 = 1) \\
&= \mathbb{P}(X_0 = 1)\mathbb{P}(X_1 = 2|X_0 = 1)\mathbb{P}(X_2 = 1|X_1 = 2) \\
&+ \mathbb{P}(X_0 = 3)\mathbb{P}(X_1 = 1|X_0 = 3)\mathbb{P}(X_2 = 1|X_1 = 2) \\
&= \mu(1)p_{12}p_{21} + \mu(3)p_{32}p_{21} \\
&= \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{36} = \left[\mu^T P^2\right]_1.
\end{aligned}$$

Using Theorem 2.1.7 (ii), we compute the conditional probability $\mathbb{P}(X_5 = 3|X_2 = 1)$ as

$$p_{13}^{(3)} = \left[\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} P^3\right]_3 = \frac{43}{72}.$$

□

2.2 Stationary distribution

Let $(X_n)_{n\geq 0}$ be Markov(μ^0, P) with state space I and let μ^k be the distribution of $(X_n)_{n\geq 0}$ at step k.

$$\mu_i^k = \mathbb{P}(X_k = i) \text{ for all } i \in I.$$

By conditioning on the possible predecessors of the $(k+1)$-th state, we see that

$$\mu_j^{k+1} = \mathbb{P}(X_{k+1} = j) = \sum_{i\in I} \mathbb{P}(X_k = i)p_{ij} = \sum_{i\in I} \mu_i^k p_{ij} \text{ for all } j \in I.$$

Rewriting this in vector form gives

$$[\mu^{k+1}]^T = [\mu^k]^T P \text{ for } k \geq 0.$$

Since P is stochastic and μ^0 is a distribution, μ^k is a distribution for all k. Hence, by Theorem 2.1.7

$$[\mu^k]^T = [\mu^0]^T P^k \text{ for } k \geq 0.$$

Does this sequence of distributions $\{\mu^0, \mu^1, \cdots\}$ have a limiting value? If such a limiting distribution π exists, then

$$\pi^T P^{n+1} = \pi^T P^n P.$$

Hence, by taking limit,

$$\pi^T = \pi^T P. \tag{2.6}$$

Eigenvector interpretation: The equation $\pi^T = \pi^T P$ signifies that π is a left eigenvector of the matrix P with eigenvalue 1. In addition, π must be a distribution:

$$\sum_{i \in I} \pi(i) = 1.$$

The matrix P always has the eigenvalue 1 because P is stochastic, i.e.

$$\sum_{j \in I} p_{ij} = 1.$$

In matrix notation we write this as $P\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is a column vector whose entries are all 1; hence, $\mathbf{1}$ is a (right) eigenvector of P corresponding to eigenvalue 1.

Definition 2.2.1. Stationary distribution.

A *steady-state vector* or *stationary distribution* for a finite state Markov chain with transition matrix P, is a vector π that satisfies

$$\pi^T = \pi^T P, \text{ where } \sum_{i \in I} \pi_i = 1 \text{ and } \pi_i \geq 0, \text{ for all } i \in I. \tag{2.7}$$

Markov chain theory ensures that this sequence of distributions has a limiting stationary distribution for certain types of random processes. From Perron-Frobenius Theorem B.0.22, we know that an irreducible and primitive matrix has such a limiting distribution. We now look at the conditions under which the probability transition matrix of a Markov chain is irreducible and primitive.

2.3 Irreducible Markov chain

Consider a Markov chain with state space I and probability transition matrix P. Suppose i, j are any two distinct states. We say that j is reachable from i if there exists an integer $n \geq 0$ such that $p_{ij}^{(n)} > 0$. Suppose i is also reachable from j, i.e., $p_{ji}^{(n')} > 0$, for some positive integer n', then states i and j are said to communicate with each other. We write this as $i \leftrightarrow j$. By convention, all states are defined to communicate with themselves: $i \leftrightarrow i$.

Theorem 2.3.1. *Communication is an equivalence relation.*

1. *Reflexive: $i \leftrightarrow i$ for all states i.*

2. *Symmetric: If $i \leftrightarrow j$, then $j \leftrightarrow i$.*

3. *Transitive: If $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$.*

Proof. Since P^0 is the identity matrix, $p_{ii}^{(0)} = 1$ for all states i. Hence, $i \leftrightarrow i$ for all states i.

If $i \leftrightarrow j$, then for some positive integers n, n', $p_{ij}^{(n)} > 0$ and $p_{ji}^{(n')} > 0$. And if $j \leftrightarrow k$, then for some positive integers m, m', $p_{jk}^{(m)} > 0$ and $p_{kj}^{(m')} > 0$. By Chapman-Kolmogorov equation, we have

$$p_{ik}^{(n+m)} = \sum_{l \in I} p_{il}^{(n)} p_{lk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0.$$

Similarly, it is easy to show that $p_{ki}^{(n'+m')} > 0$ and so $i \leftrightarrow k$.

□

All states that communicate with each other belong to the same communication class and communication classes do not overlap. Thus, the communication classes partition the state space I.

Example 2.3.2. Consider a Markov chain on $I = \{1, 2, 3\}$ with the following:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}. \text{ Then } P^2 = \begin{pmatrix} 1/2 & 3/8 & 1/8 \\ 3/8 & 19/48 & 11/48 \\ 1/6 & 19/36 & 11/36 \end{pmatrix}.$$

Since P^2 is a positive matrix, all states communicate and there is a single communication class.

Suppose we change the last row of the probability transition matrix as shown below:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then, P^n is not a positive matrix for any positive integer n. P has two communication classes containing the appropriate states, namely $C_1 = \{1, 2\}$ and $C_2 = \{3\}$.

Definition 2.3.3. Irreducible Markov chain.

A Markov chain for which there is only one communication class is called an *irreducible* Markov chain; all states communicate.

Theorem 2.3.4. *If $Markov(\mu, P)$ is irreducible, then its probability transition matrix P is also irreducible.*

Proof. Suppose $Markov(\mu, P)$ is irreducible. Then for every pair of states $(i, j), i \neq j$, there exists a positive integer k (depending on i, j) such that $p_{ij}^{(k)} > 0$. Suppose matrix P is reducible. By Definition B.0.14,

$$U^T P U = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

for some $n \times n$ permutation matrix U. Note that $UU^T = I$. By matrix block multiplication,

$$(U^T P^k U) = \underbrace{U^T P U \cdot U^T P U \cdots}_{k \text{ times}} = (U^T P U)^k = \begin{bmatrix} B^k & * \\ 0 & D^k \end{bmatrix}.$$

Clearly, $[U^T P^k U]_{n1} = 0$. Since the 1^{st} column and n^{th} row of U are the standard vectors e_j, e_i

$$0 = [U^T P^k U]_{n1} = [U^T_{row\ n}] P^k [U_{col\ 1}] = e_i^T P^k e_j = [P^k]_{ij}.$$

By irreducibility of the Markov chain, $[P^k]_{ij} = p_{ij}^{(k)} > 0$, a contradiction. Hence, P is irreducible.

□

2.4 Recurrent states

The probability that the chain reenters state i after n steps is given by

$$\mathbb{P}(X_n = i | X_0 = i) = p_{ii}^{(n)}.$$

Consider the random variable

$$L_n = \begin{cases} 1, & \text{if } X_n = i. \\ 0, & \text{if } X_n \neq i. \end{cases}$$

Then, the number of visits to state i is

$$\sum_{n=0}^{\infty} L_n.$$

The expected value of the number of visits to i is given by

$$\begin{aligned} & E(\text{number of visits to } i | X_0 = i) \\ &= E\left(\sum_{n=0}^{\infty} L_n\right) \\ &= \sum_{n=0}^{\infty} E(L_n | X_0 = i) \\ &= \sum_{n=0}^{\infty} \mathbb{P}(L_n = 1 | X_0 = i) \\ &= \sum_{n=0}^{\infty} \mathbb{P}(X_n = i | X_0 = i) \\ &= \sum_{n=0}^{\infty} p_{ii}^{(n)}. \end{aligned}$$

Definition 2.4.1. Recurrent state.

A state i is said to be *recurrent* if

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty;$$

transient if

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty.$$

Theorem 2.4.2. *For any communication class C, if a state $i \in C$ is recurrent, then all states in C are recurrent. If not, all states are transient.*

Proof. Suppose $i \in C$ is recurrent. Let $j \in C$. By definition of communicating class, $i \leftrightarrow j$. So, there exists positive integers a, b such that $p_{ij}^{(a)} > 0$ and $p_{ji}^{(b)} > 0$. Using Chapmann-Kolmogrov, we compute

$$p_{jj}^{(n+a+b)} \geq p_{ji}^{(b)} p_{ii}^{(n)} p_{ij}^{(a)}, \text{ for any } n,$$

and so

$$\sum_{k\geq 0} p_{jj}^{(k)} \geq \sum_{a,b,n\geq 0} p_{jj}^{(n+a+b)} \geq p_{ji}^{(b)} p_{ij}^{(a)} \sum_{n\geq 0} p_{ii}^{(n)} = \infty.$$

Hence, j is also recurrent. Since this is true for any $j \in C$, all states in C are recurrent.

□

Definition 2.4.3. Recurrent Markov chain.

If all states in a Markov chain are recurrent, then the Markov chain is termed recurrent; it is transient otherwise.

Theorem 2.4.4. *A finite state Markov chain cannot have all transient states.*

Proof. Let $I = \{i_k\}, 1 \leq k \leq m$, be the set of possible states of the Markov chain. We start by generating a sequence of non-communicating states. If state i_1 is transient, then there exists state i_2 such that $i_1 \to i_2$ but $i_2 \not\to i_1$. If i_2 is also transient, then for some state i_3, $i_2 \to i_3$ but $i_3 \not\to i_2$ and $i_3 \neq i_1$. Thus, successive states, $i_1, i_2, \cdots, i_k$ are distinct and transient. If for some state i_{k+1}, $i_k \to i_{k+1}$ but $i_{k+1} \not\to i_k$, then $i_{k+1} \neq i_j$ for $1 \leq j \leq k$.

Suppose the Markov chain is in state i_{k+1} having visited all other states in I but without revisiting any state. Then, in the next step, the chain must re-visit some state $i_j, 1 \leq j \leq k$. So, $p_{(k+1)j} > 0$. The chain then revisits the sequence of states $\{i_j, i_{j+1}, \cdots, i_k, i_{k+1}\}$. So, there is a path from $i_j \to i_{k+1}, k > j$ and $p^n_{j(k+1)} > 0$ for some positive integer n. So, $j \leftrightarrow (k+1)$ and states j and $k+1$ communicate. By transitivity of $\leftrightarrow$, this sequence of states $\{j, j+1, \cdots, k, k+1\}$ form a communication class. It is now easy to see that

$$\sum_{n=0}^{\infty} p^{(n)}_{(k+1)(k+1)} = \infty,$$

and state $k+1$ is recurrent. By Theorem 2.4.2, all states in this communication class are recurrent. Hence, a finite state Markov chain cannot have all transient states.

□

Corollary 2.4.5. *An irreducible and finite state Markov chain has all recurrent states.*

Proof. From Theorem 2.4.4, the Markov chain must have at least one communication class with recurrent states. But the chain is irreducible. By Definition 2.3.3, it has only one communication class. Since all states communicate, the chain revisits all states. So, all states are recurrent.

2.5 Periodic and aperiodic Markov chains

Definition 2.5.1. Period of a Markov chain.

The *period* $r(i)$ of a recurrent state $i \in I$ is defined to be the greatest common divisor of all steps n at which the Markov chain returns to i.

$$D(i) = \{n \in \mathbb{Z}^+ | p_{ii}^{(n)} > 0\}.$$

$$r(i) = gcd\Big(D(i)\Big).$$

Theorem 2.5.2. *If two states $i, j \in I$ communicate, then $r(i) = r(j)$.*

Proof. Suppose two states, $i, j \in I$, communicate. Then, for some positive integers x, y, $p_{ij}^{(x)} > 0$, $p_{ji}^{(y)} > 0$, and $p_{jj}^{(x+y)} \geq p_{ji}^{(y)} p_{ij}^{(x)} > 0$. Hence, $r(j) \mid (x + y)$. If $n \in D(i)$ is such that $p_{ii}^{(n)} > 0$, then $p_{jj}^{(x+y+n)} \geq p_{ji}^{(y)} p_{ii}^{(n)} p_{ij}^{(x)} > 0$. Hence, $r(j) \mid (x + y + n)$. A number that divides any two numbers must divide their difference as well, so $r(j) \mid n$ for all $n \in D(i)$. Since $r(i)$ is the gcd of $D(i)$, we must have $r(j) \leq r(i)$. Similarly, it is straightforward to show that $r(i) \leq r(j)$. Hence, $r(i) = r(j)$.

□

Corollary 2.5.3. *Period is a class property.*

Corollary 2.5.4. *An irreducible and recurrent Markov chain has the same period for all states $i \in I$.*

Proof. An irreducible Markov chain has a single communication class. By Theorem 2.5.2, all states have the same period.

□

Definition 2.5.5. Aperiodic Markov chain.

An irreducible, recurrent Markov chain with period one is *aperiodic.*

Theorem 2.5.6. *If an irreducible Markov chain is aperiodic, its probability transition matrix P is primitive.*

Proof. A matrix P is primitive, if for some positive integer L, P^L is a positive matrix as defined in B.0.13.

Suppose $(X_n)_{n\geq 0}$ Markov(μ, P) is irreducible and aperiodic. By Lemma D.0.40, for every state i, there exists $m_i \in M_i$ such that for any $m \geq m_i$, $p_{ii}^{(m)} > 0$. Set $M = max_{i\in I}(m_i)$.

By irreducibility of the Markov chain, for every pair $(i, j), i \neq j$, there exists $r_{ij} \in \mathbb{Z}^+,$ such that $p_{ij}^{(r_{ij})} > 0$. Set $R = max_{i,j\in I, i\neq j}(r_{ij})$.

Let $L = M + R$. Then, for every state i, $L \geq m_i$. So, $p_{ii}^{(L)} > 0$. For every pair (i, j), $L \geq r_{ij} + m_i$. Hence, $p_{ij}^{(L)} \geq p_{ii}^{(L-r_{ij})} p_{ij}^{(r_{ij})} \geq p_{ii}^{(m_i)} p_{ij}^{(r_{ij})} > 0$. Thus, P is primitive.

□

Theorem 2.5.7. *Suppose M is a stochastic matrix. Then the spectrum of M is contained in the unit disc.*

Proof. Let v be an eigenvector of M and λ the corresponding eigenvalue. Then for any matrix norm $||\cdot||$, by Theorem 5.6.8 in [RAH85], we have

$$|\lambda|||v|| = ||\lambda v|| = ||Mv|| \leq ||M||||v||.$$

Since v is a nonzero vector,

$$|\lambda| \leq ||M||_\infty = 1,$$

where $||\cdot||_\infty$ is the max row sum norm. Hence, the eigenvalues of the probability transition matrix P lie in $[-1, 1]$.

□

Theorem 2.5.8. ***The fundamental stability theorem for Markov chains.*** *Suppose* $Markov(\mu, P)$ *is irreducible and aperiodic. Then*

i. $Markov(\mu, P)$ *has an unique stationary distribution* π.

ii. The probability transition matrix P *converges to a matrix with rows all equal to* π^T.

$$\lim_{m\to\infty} P^m = \mathbf{1}\pi^T \text{ where } \lim_{m\to\infty} p_{ij}^{(m)} = \pi(j),\ i,j \in I.$$

iii. $\lim_{m\to\infty} \mathbb{P}(X_m = j) = \pi(j)$, *for any initial distribution* μ.

Proof. By Theorem 2.5.6, P is primitive. So, we apply the Perron-Frobenius Theorem B.0.22 to P.

i. P has only one eigenvalue $\lambda > 0$ on its spectral radius. By Theorem 2.5.7, $\lambda = 1$ is the spectral radius of P. Hence, $\lambda = 1$ is the largest, simple eigenvalue of P. By Perron-Frobenius Theorem, P has unique left and right Perron vectors corresponding to λ. Since P is stochastic, we see that $\mathbf{1}$ is the right eigenvector with eigenvalue 1. Suppose we denote π to be the left eigenvector. By Perron-Frobenius Theorem, we know that $\pi^T\mathbf{1} = \sum_{i=1}^{n} \pi(i) = 1$. Thus, π is a probability distribution. Since $\pi^T P = \pi^T$, by Definition 2.2.1, π is the unique stationary distribution vector of P.

ii. π^T and $\mathbf{1}$ are the left and right Perron vectors of P corresponding to eigenvalue

1 and are strictly positive. By Perron-Frobenius Theorem iv, P has a limit

$$\lim_{m\to\infty}[P]^m = \mathbf{1}\pi^T = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \cdots & \cdots & \pi_n \\ \vdots & & \vdots & \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{bmatrix}.$$

iii. By 2.1.7(ii), $\mathbb{P}(X_m = j) = p_{ij}^{(m)}$. By 2.1.8, $p_{ij}^{(m)} = [P^m]_{ij}$. Using (ii) above, we have the desired result.

□

For Markov chains, the past and the future are independent of the present. This property is symmetrical in time and suggests we look at the reverse of a Markov chain, running backwards. But we have looked at Markov chains that converge to a limiting invariant distribution. This suggests that if we start with the invariant distribution, the Markov chain will be in equilibrium, i.e., a Markov chain running forward and backward are symmetric in time. A Markov chain running backwards is also a Markov chain, but with a different probability transition matrix.

Theorem 2.5.9. *Let P be irreducible with π as its stationary distribution. Suppose (X_n) is Markov. Set $Y_n = X_{N-n}$, for some fixed N. The reverse chain (Y_n) is also Markov$(\pi, \hat{P})$, where $\hat{P} = [\hat{p}_{ij}]$ is given by*

$$\pi_j \hat{p}_{ji} = \pi_i p_{ij} \quad \textit{for all } i, j \in I. \tag{2.8}$$

Furthermore, $\hat{P}$ is also irreducible with stationary distribution π.

Proof. First, we show that $\hat{P}$ is stochastic. $\hat{p}_{ji} = \frac{\pi_i}{\pi_j} p_{ij}$. We write $\hat{P}$ as

$$\hat{P} = \mathbf{D}(\pi^{-1}) P^T \mathbf{D}(\pi),$$

where $\mathbf{D}(\pi)$ is the diagonal matrix with π_i on the diagonals and $\mathbf{D}(\pi^{-1})$ is the diagonal matrix with $1/\pi_i$ on the diagonals. Then

$$\hat{P}\mathbf{1} = [\mathbf{D}(\pi^{-1})P^T\mathbf{D}(\pi)]\mathbf{1} = \mathbf{D}(\pi^{-1})P^T\pi = \mathbf{D}(\pi^{-1})\pi = \mathbf{1}.$$

Next, we show that (Y_n) is Markov.

$$\begin{aligned}
&\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \cdots, Y_N = i_N) \\
&\quad = \mathbb{P}(X_N = i_N, X_{N-1} = i_{N-1}, \cdots, X_0 = i_0) \\
&\quad = \pi_{i_N} p_{i_N i_{N-1}} p_{i_{N-1} i_{N-2}} \cdots p_{i_1 i_0} \\
&\quad = \pi_{i_0} \hat{p}_{i_0 i_1} \hat{p}_{i_1 i_2} \cdots \hat{p}_{i_{N-1} i_N}.
\end{aligned}$$

By Theorem 2.1.6, (Y_n) is Markov$(\pi, \hat{P})$.

To show that $\hat{P}$ is also irreducible, consider any two distinct states i, j. There exists a chain of states $i_0 \to i_1 \to \cdots \to i_k$ and $p_{i_0 i_1} \cdots p_{i_{k-1} i_k} > 0$. Then

$$\hat{p}_{i_N i_{N-1}} \cdots \hat{p}_{i_{N-k-1} i_{N-k}} = \frac{1}{\pi_{i_N}} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{k-1} i_k} > 0.$$

So, there is only one communication class. Hence, $\hat{P}$ is also irreducible.

Finally, we show that π is indeed the stationary distribution of $\hat{P}$.

$$\pi^T \hat{P} = \pi^T [\mathbf{D}(\pi^{-1})P^T\mathbf{D}(\pi)] = \mathbf{1}^T P^T \mathbf{D}(\pi) = \mathbf{1}^T \mathbf{D}(\pi) = \pi^T. \tag{2.9}$$

□

Definition 2.5.10. Time-reversed Markov chains.

The chain $(Y_n)_{n \geq 0}$ is called the *time-reversal* of $(X_n)_{n \geq 0}$.

Definition 2.5.11. Detailed balance.

A stochastic matrix P and stationary distribution π are said to be in *detailed*

balance if $\hat{P} = P$, i.e.

$$\mathbf{D}(\pi)P = P^T\mathbf{D}(\pi)$$
$$P = \mathbf{D}(\pi^{-1})P^T\mathbf{D}(\pi) = \hat{P}.$$

We write the condition in (2.8) as $\pi_j p_{ji} = \pi_i p_{ij}$.

Example 2.5.12. Consider a Markov chain with state transition diagram as shown in Figure 2.3.

$$\pi^T = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 2/3 & 1/3 & 0 \end{bmatrix}.$$

$$\hat{P} = \mathbf{D}(\pi^{-1})P^T\mathbf{D}(\pi) = \begin{bmatrix} 0 & 1/3 & 2/3 \\ 2/3 & 0 & 1/3 \\ 1/3 & 2/3 & 0 \end{bmatrix}.$$

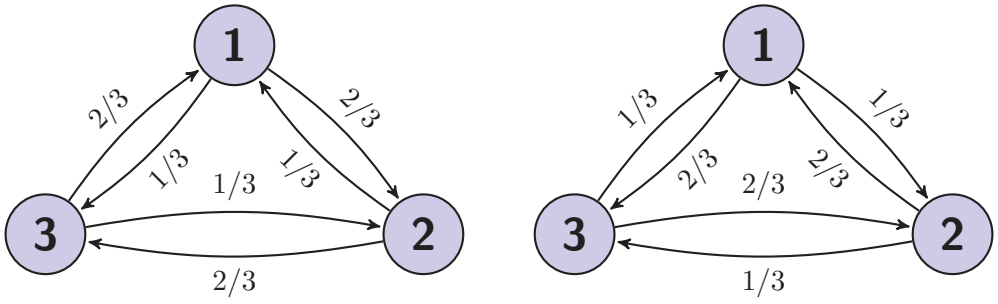

Figure 2.3: Markov chain not in detailed balance $P \neq \hat{P}$.

Example 2.5.13. Consider a Markov chain with state transition diagram as shown below in Figure 2.4.

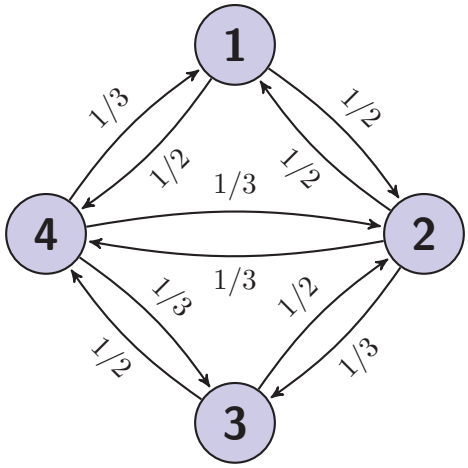

Figure 2.4: Markov chain in detailed balance $P = \hat{P}$.

$$\pi^T = \begin{bmatrix} 1/5 & 3/10 & 1/5 & 3/10 \end{bmatrix} \text{ and } P = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix} = \hat{P}.$$

2.6 Hitting time of a connected graph

Definition 2.6.1. Hitting time.

Given an irreducible and aperiodic Markov chain $(X_n)_{n\geq 0}$, and two distinct states $i, j \in I$, the *hitting time* or *mean first passage time* is the expected number of steps to reach state j starting from state i, for the first time.

$$H(i,j) = \sum_{t=1}^{\infty} t \cdot \mathbb{P}(X_t = j | X_0 = i, X_k \neq j, k < t).$$

Definition 2.6.2. Return time.

Given an irreducible and aperiodic Markov chain $(X_n)_{n\geq 0}$, and state $i \in I$, the *return time* or *mean recurrence time* is the expected number of steps to return to i

for the first time.

$$R(i) = \sum_{t=1}^{\infty} t \cdot \mathbb{P}(X_t = i | X_0 = i, X_k \neq i, k < t).$$

Since there are numerous paths from $i \to j$, the probabilistic computation of hitting time and return time is laborious. So, we look at matrix based techniques for simplifying such computations. Consider the first step to any state k from i with $p_{ik} > 0$. Then from k we navigate to j. We write this as

$$H(i,j) = 1 + \sum_{k \neq j} p_{ik} H(k,j). \tag{2.10}$$

Similarly, starting at i, the chain takes at least one step to some state $j \neq i$ and returns to i. Considering all possible first steps, we get

$$R(i) = \sum_{k} p_{ik}(H(k,i) + 1) = 1 + \sum_{k} p_{ik} H(k,i). \tag{2.11}$$

Define two matrices $\mathbf{H}$, where $\mathbf{H}_{ij} = H(i,j)$, $\mathbf{H}_{ii} = 0$ and $\mathbf{R}$, a diagonal matrix with $\mathbf{R}_{ii} = R(i)$. We combine the above two equations in to a single matrix form as

$$\mathbf{H} = P\mathbf{H} + J - \mathbf{R}, \text{where } J \text{ is the all one matrix.} \tag{2.12}$$

Equivalently,

$$(I - P)\mathbf{H} = J - \mathbf{R}. \tag{2.13}$$

Theorem 2.6.3. *If $Markov(\mu, P)$ is irreducible and aperiodic, then the return time for any state $i \in I$ is $R(i) = 1/\pi(i)$, where π is the stationary distribution.*

Proof. Multiplying both sides of (2.13) by π^T gives

$$\pi^T(I - P)\mathbf{H} = \pi^T J - \pi^T \mathbf{R}.$$

Since $\pi^T P = \pi^T$ and $\pi^T J = \mathbf{1}^T$, we get

$$\vec{0} = (1, 1, \cdots, 1) - (\pi_1 R(1), \pi_2 R(2), \cdots, \pi_n R(n)),$$

yielding $R(i) = 1/\pi(i)$ for all states i.

□

2.6.1 Fundamental matrix

The matrix $(I - P)$ is not invertible, since it has row sum zero. So, we consider a rank-one update $\Pi = \mathbf{1}\pi^T$ to $(I - P)$. The new matrix $(I - P + \Pi)$ is invertible. We define $\mathbf{Q} = (I - P + \Pi)^{-1}$ to be the *fundamental matrix* of the Markov chain. We now look at some properties of the fundamental matrix $\mathbf{Q}$.

Proposition 2.6.4. $\mathbf{Q}J = J$.

Proof. Since $PJ = J$ and $\Pi J = J$, $(I - P + \Pi)J = J$. Hence, $\mathbf{Q}J = \mathbf{Q}(I - P + \Pi)J$ gives us $\mathbf{Q}J = J$.

□

Proposition 2.6.5. $\mathbf{Q}\mathbf{1} = \mathbf{1}$.

Proof. Since $(I - P + \Pi)\mathbf{1} = \mathbf{1}$, $\mathbf{1} = \mathbf{Q}\mathbf{1}$.

□

Proposition 2.6.6. $\mathbf{Q}(I - P) = (I - P)\mathbf{Q} = (I - \Pi)$.

Proof. Note that $P\Pi = P\mathbf{1}\pi^T = \mathbf{1}\pi^T = \Pi$. Similarly, $\Pi P = \mathbf{1}\pi^T P = \mathbf{1}\pi^T = \Pi$. Also, $\Pi^2 = \mathbf{1}(\pi^T\mathbf{1})\pi^T = \mathbf{1}\pi^T = \Pi$. So, we have,

$$(I - P + \Pi)(I - \Pi) = I - P + \Pi - \Pi + P\Pi - \Pi^2 = (I - P).$$

And

$$(I - \Pi)(I - P + \Pi) = I - \Pi - P + \Pi + \Pi - \Pi = (I - P).$$

Hence, $\mathbf{Q}(I - P) = (I - P)\mathbf{Q} = (I - \Pi)$.

□

Theorem 2.6.7. *The hitting time in terms of the fundamental matrix* $\mathbf{Q}$ *is given by*

$$\mathbf{H}_{ij} = \frac{\mathbf{Q}_{jj} - \mathbf{Q}_{ij}}{\pi_j}.$$

Proof. From (2.13), we have

$$(I - P)\mathbf{H} = J - \mathbf{R}.$$

Multiplying on the left by $\mathbf{Q}$ yields

$$\mathbf{Q}(I - P)\mathbf{H} = \mathbf{Q}(J - \mathbf{R}).$$

From propositions 2.6.4 and 2.6.6, we get $(I - \Pi)\mathbf{H} = J - \mathbf{QR}$. Hence

$$\mathbf{H} = J - \mathbf{QR} + \Pi\mathbf{H}. \tag{2.14}$$

So,

$$0 = \mathbf{H}_{jj} = 1 - \mathbf{Q}_{jj}R(j) + [\pi^T\mathbf{H}]_j, \tag{2.15}$$

and

$$\mathbf{H}_{ij} = 1 - \mathbf{Q}_{ij}R(j) + [\pi^T\mathbf{H}]_j. \tag{2.16}$$

Subtracting (2.15) from (2.16) results in

$$\mathbf{H}_{ij} = (\mathbf{Q}_{jj} - \mathbf{Q}_{jj})R(j).$$

Substituting $R(j) = 1/\pi(j)$, allows us to express the hitting time between any two distinct states as

$$\mathbf{H}_{ij} = \frac{\mathbf{Q}_{jj} - \mathbf{Q}_{ij}}{\pi_j}. \tag{2.17}$$

□

CHAPTER 3

RANDOM WALK ON GRAPHS

A random walk is a mathematical formalization of a path that consists of a succession of random steps. For example, the path traced by a molecule as it travels in liquid or gas, the search path of a foraging animal, the price of a fluctuating stock, or the financial status of a gambler can all be modeled as random walks, although they may not be truly random in reality. The term random walk was first introduced by Karl Pearson in 1905. Random walks have been used in many varied fields: ecology, economics, psychology, computer science, physics, chemistry, biology, finance, and more. Random walks explain the observed behaviors of processes in these fields and thus serve as a fundamental model for the recorded stochastic activity. Though many types of random walks exist, we are interested in random walks that are time-homogeneous Markov chains. Random walks occur on graphs, integer lines, planes, or even on topological structures of higher dimensions. Our study focuses on time-homogenous random walks on finite, connected graphs. Appendix C contains basic definitions and theorems on graphs.
In this chapter we follow [Lov93].

Example 3.0.8. Let us consider a simple random walk on the integer line as shown in Figure 3.1.

Suppose our random walk starts on 0. The probability of getting to 1 and -1 are the same, equal to $\frac{1}{2}$. This is true for the transition from any integer n to $n \pm 1$. This is an example of a simple random walk and is sometimes referred to as the drunkard's walk.

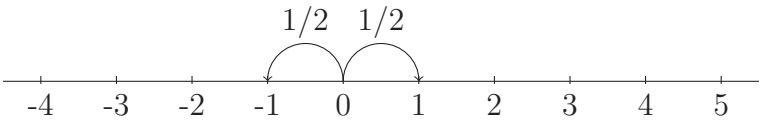

Figure 3.1: Drunkard's walk on an integer line.

3.1 Random walk on graphs

Given a finite, connected graph $\mathcal{G}(V, E)$, and a starting vertex v_0, choose any adjacent vertex v_1 at random and move to this neighbor. Then select v_2, a neighbor of v_1 at random and move to v_2, and so on. The sequence of vertices, so chosen, $\{v_0, v_1, \cdots, v_k\}$, constitute a random walk on $\mathcal{G}$.

At each step k, we assign to the random variable X_k, a value from V. Hence, the random sequence $X_0, X_1, X_2, \cdots X_k, \cdots$, is a discrete time stochastic process defined on the state space V.

The choice of vertex v_i, at any step k, depends only on reaching its neighbor v_j in step $(k-1)$ and not how v_j is reached. In an unweighted graph, the probability of taking an edge depends only on the degree of the current vertex and is the same for all edges from a vertex. Suppose $d(v_i)$ denotes the degree of vertex v_i and p_{ij} denotes the probability of moving from vertex v_i to vertex v_j. Then

$$p_{ij} = \mathbb{P}(X_{k+1} = v_j | X_k = v_i) = \begin{cases} \frac{1}{d(v_i)}, & \text{if } (ij) \in E. \\ 0, & \text{otherwise.} \end{cases}$$

The transition probabilities p_{ij} are independent of time k. If at time k we are at vertex v_i, we choose v_j uniformly from the neighbors of v_i and move to it. The process is thus "memoryless;" the future choice of vertex depends only on the current vertex. We denote $v_i \sim v_j$ if v_j is a neighbor of v_i. In this chapter, we focus on finite, connected, unweighted graphs.

Definition 3.1.1. Adjacency matrix.

For a graph $\mathcal{G}$ with $V = \{v_1, v_2, \cdots\}$, the *adjacency matrix* A is given by

$$[A]_{ij} = \begin{cases} 1, & \text{if } v_i \sim v_j. \\ 0, & \text{otherwise.} \end{cases}$$

Definition 3.1.2. Degree matrix.

For a graph $\mathcal{G}$, the *degree matrix* D is the diagonal $|V| \times |V|$ matrix given by

$$[D]_{ii} = d(v_i), \text{ where } d(v_i) = \sum_j [A]_{ij}.$$

Definition 3.1.3. Probability transition matrix.

For a graph $\mathcal{G}$, the *probability transition matrix* P is the $|V| \times |V|$ matrix given by

$$P = D^{-1}A.$$

Suppose μ^0 is the initial probability distribution, the random sequence of vertices visited by the walk $X_0, X_1, \cdots, X_k, \cdots$, is $Markov(\mu^0, P)$ with state space V. The probability distribution μ^t at any time t is given by

$$[\mu^t]^T = [\mu^0]^T P^t.$$

Theorem 3.1.4. *A random walk on a graph $\mathcal{G}$ with probability transition matrix P is Markov and Theorem 2.1.5 holds; i.e.,*

$$\mathbb{P}(X_{k+1} = v_j | X_k = v_i, X_{k-1} = v_{k-1}, \cdots, X_1 = v_1, X_0 = v_0) = \mathbb{P}(X_{k+1} = v_j | X_k = v_i) = p_{ij}. \quad (3.1)$$

Proof. First, we show that the probability transition matrix $P = [p_{ij}]$ of a random walk is stochastic. For any row i of the matrix P,

$$\sum_j p_{ij} = \sum_{v_j \sim v_i} \frac{1}{d(v_i)} = 1,$$

since there are $d(v_i)$ entries in each row i. To see that the Markov property holds, we use the definition of $\mathbb{P}(A|B)$ and compute

$$\begin{aligned}
&\mathbb{P}(X_{k+1} = v_j | X_k = v_i, X_{k-1} = v_{i_{k-1}}, \cdots, X_0 = v_0) \\
&= \frac{p_{ij} p_{(k-1)i} p_{(k-2)(k-1)} \cdots}{p_{(k-1)i} p_{(k-2)(k-1)} \cdots} \\
&= p_{ij}.
\end{aligned}$$

□

Example 3.1.5. Let us consider an undirected, connected graph with five vertices as shown in Figure 3.2. The probability of transition between any two vertices depends on the degree of the current vertex.

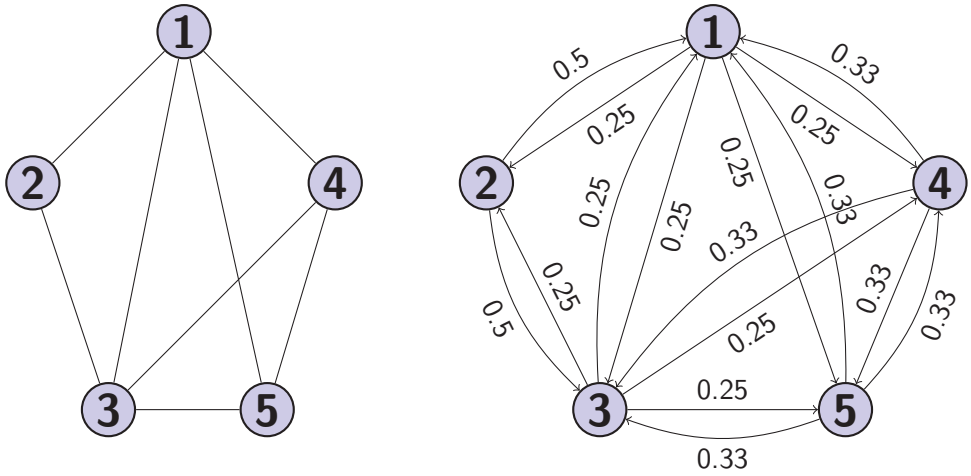

Figure 3.2: Undirected graph with state transition diagram.

$$P = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}.$$

□

We are interested in finding the stationary distribution of a random walk. From Markov chain theory, we know that if a Markov chain is irreducible and aperiodic, then the stationary distribution exists. The properties of the random walk reflect the properties of the underlying graph. Here, we show that a random walk on a connected, non-bipartite graph has a stationary distribution.

Theorem 3.1.6. *A random walk on a connected graph $\mathcal{G}$ is irreducible.*

Proof. Since the graph $\mathcal{G}$ is connected, for any two vertices $v_i, v_j \in V$, there exists a path from v_i to v_j: $v_i \to v_{i_1} \to \cdots \to v_j$ such that $p_{i_k i_{k+1}} > 0$ for every vertex v_{i_k} in the path. The transition probability from $v_i \to v_j$ is

$$\begin{aligned} \mathbb{P}(X_m = v_j | X_0 = v_i) &= \mathbb{P}(X_m = v_j | X_{m-1} = v_{i_{m-1}}, X_{m-2} = v_{i_{m-2}}, \cdots, X_0 = v_i) \\ &\geq p_{ii_1} p_{i_1 i_2} \cdots p_{i_{m-1} j} > 0. \end{aligned}$$

The final step above is using Theorem 2.1.7. Similarly, there is a path from v_j to v_i. So v_i and v_j communicate: $v_i \leftrightarrow v_j$. By Theorem 2.3.1, communication is an equivalence relation. Since graph $\mathcal{G}$ is connected, we conclude that all vertices in $\mathcal{G}$ communicate and hence belong to the same communication class. By Definition 2.3.3, the random walk is irreducible and by Theorem 2.3.4, the probability transition matrix P is also irreducible. By Corollary 2.4.5, all states are recurrent.

□

Since the graph is connected, by Theorem 2.5.2 and its Corollary 2.5.4, all vertices in $\mathcal{G}$ have the same period. Our next theorem shows that a random walk on a non-bipartite graph is aperiodic.

Theorem 3.1.7. *A random walk on a finite, connected graph is aperiodic if and only if the graph is non-bipartite.*

Proof. From Appendix C.0.38, a graph is bipartite if and only if it has no odd cycles. Suppose the walk is aperiodic, by definition of aperiodic, the graph has an odd cycle, or else two divides its period. Hence the graph is not bipartite.
On the other hand, suppose the graph is non-bipartite, it has at least one odd cycle. Any random walk on a connected, undirected graph has a walk with return time of two, i.e., you leave a vertex in any direction and return back in the next step. So, for each vertex, the walk also has a cycle of length two. Hence the gcd of the set of cycles of $\mathcal{G}$ is one. By Definition 2.5.1, the graph is aperiodic.

□

Theorem 3.1.8. *The stationary distribution vector π for a random walk on a finite connected graph $\mathcal{G}(V, E)$ exists and is given by $\pi_i = \dfrac{d(v_i)}{2m}$, where $m = |E|$.*

Proof. A random walk on a finite, connected graph is irreducible and aperiodic. By the fundamental stability theorem of Markov chains, Theorem 2.5.8, such a random walk has a stationary distribution and its probability transition matrix has a limiting value. Furthermore, by Perron-Frobenius theorem, P has right and left Perron vectors: $\mathbf{1}$ and π. Hence

- $\pi^T = \pi^T P$ and $\pi^T \mathbf{1} = 1$,
- $\lim_{m\to\infty} P^m = \Pi = \mathbf{1}\pi^T$.

Suppose $\pi_i = \frac{d(v_i)}{2m}$ for all $v_i \in V$, it is easily verified that $\pi^T P = \pi^T$.

$$\begin{aligned}\pi^T P &= \frac{1}{2m}[d(v_1)\cdots d(v_k)]D^{-1}A \\ &= \frac{1}{2m}[1\cdots 1]A \\ &= \frac{1}{2m}[d(v_1)\cdots d(v_k)] \\ &= \pi^T.\end{aligned}$$

□

Theorem 3.1.9. *A random walk on a finite, connected graph $\mathcal{G}$ is time reversible.*

Proof. From Definition 2.5.11, a Markov chain is time reversible if $\hat{P} = P$.

$$\mathbf{D}(\pi^{-1})P^T\mathbf{D}(\pi) = 2mD^{-1}(D^{-1}A)^T D\frac{1}{2m} = D^{-1}A^T D^{-1}D = D^{-1}A = P.$$

The random walk is in detailed balance. Hence is time reversible.

□

3.2 Access times on graphs

In a random walk, given any starting vertex, we choose any neighbor at random and proceed. This random choice is distributed evenly among the neighbors of the said vertex. For a finite, connected graph, there is a path between any two arbitrary vertices. This allows us to turn our focus to less qualitative questions; rather than asking whether or not a random walk will return to its starting vertex,

it is interesting to ask what is the expected number of steps the random walk would take to return to the starting vertex, reach a specific vertex, or to commute between any two vertices.

Definition 3.2.1. Hitting time.

Given graph $\mathcal{G}$, the *hitting time* $H(i,j), i \neq j$, from vertex v_i to v_j, is the expected number of steps it takes for a random walk that starts at vertex v_i to reach vertex v_j for the first time.

$$H(i,j) = \sum_{t=1}^{\infty} t \cdot \mathbb{P}(X_t = j | X_0 = i; X_k \neq j, k < t).$$

Definition 3.2.2. Commute time.

Given graph $\mathcal{G}$, the *commute time* $C(i,j), i \neq j$, between two vertices v_i and v_j is the expected number of steps that a random walk starting at v_i takes to reach v_j and return back to v_i.

$$C(i,j) = H(i,j) + H(j,i).$$

Usually, $H(i,j) \neq H(j,i)$. But $C(i,j) = C(j,i)$.

Definition 3.2.3. Return time.

Given graph $\mathcal{G}$, the return time to a vertex $R(i,i)$, is the number of steps that a random walk starting at v_i takes to return to v_i. Indeed, by Theorems 2.6.3 and 3.1.8, $R(i,i) = {\pi_i}^{-1} = \dfrac{2m}{d(v_i)}$.

Example 3.2.4. Let us look at a simple random walk on a path with $n+1$ nodes: $\{0, 1, 2, \cdots n\}$. We are interested in finding the hitting time $H(i,k)$, where i and k are any two nodes on the path. For $k \geq 1$, the hitting time $H(k-1,k)$, is equivalent to the expected return time of a random walk on a path with $k+1$ nodes, starting at an end node minus one. If we begin our random walk on node $k \geq 1$, then to

return to node k, it takes one fewer steps than had we started on node k. The return time for any node is given by $\frac{2m}{d(v_i)}$. The degree of the last node is *one*. Here we have k edges. Hence the return time is $2k$. Hence $H(k-1,k) = 2k-1$.
Now, let us look at hitting time $H(i,k), 0 \leq i \leq k \leq n$. To reach node k, we first have to first reach $k-1$. So we have the recurrence

$$\begin{aligned}
H(i,k) &= H(i,k-1) + 2k - 1 \\
&= H(i,k-2) + 2k - 3 + 2k - 1 \\
&= \cdots \\
&= H(i,i+1) + (2i+3) + \cdots + (2k-1) \\
&= (2i+1) + (2i+3) + \cdots + (2k-1) \\
&= (k-i)(2i) + (1 + 3 + \cdots + 2(k-i) - 1) \\
&= 2ki - 2i^2 + (k-i)^2 \\
&= k^2 - i^2
\end{aligned}$$

In particular $H(0,n) = n^2$.

□

Example 3.2.5. Let C be a cycle with n vertices. Then, the hitting time from any vertex v_i to a vertex that is l steps away is independent of v_i and is given by

$$H(i,i+l) = H_l = l(n-l).$$

Proof. From vertex v_i, the first step is either to vertex v_{i-1} or v_{i+1}, both with probability $\frac{1}{2}$.

$$\begin{aligned}
H(i,i+l) &= \frac{1}{2}(H(i-1,i+l) + H(i+1,i+l)) + 1 \\
&= \frac{1}{2}(H(i-1,i+l) + \frac{1}{2}(H(i+1,i+l)) + 1 \qquad (3.2)
\end{aligned}$$

Since $H(i, i+l)$ does not depend on i, but only on the distance l, we denote this by H_l and write

$$H_l = \frac{1}{2}H_{l-1} + \frac{1}{2}H_{l+1} + 1$$
$$-1 = \frac{1}{2}H_{l-1} + \frac{1}{2}H_{l+1} - H_l$$

We now setup a system of linear equations for $l = 1, l = 2, etcetra.$

$$\frac{1}{2}H_0 + \frac{1}{2}H_2 - H_1 = -1$$
$$\frac{1}{2}H_1 + \frac{1}{2}H_3 - H_2 = -1$$
$$\frac{1}{2}H_2 + \frac{1}{2}H_4 - H_3 = -1$$
$$\vdots$$
$$\frac{1}{2}H_{n-2} + \frac{1}{2}H_n - H_{n-1} = -1$$

The above set of equations are linearly independent. Suppose a linear combination of the above $n-2$ equation must result in $\mathbf{0}$. Since H_0 appears only in the first equation, that equation must have coefficient e_1 equal to 0 so that $e_1 H_0$ equals 0. Then, H_1 appears only in the second equation, hence this equation too must have coefficient e_2 equal 0 so that $e_2 H_1$ is 0. Proceeding in a similar manner, the coefficients of all the equation must be 0 to add up to $\mathbf{0}$. Hence this system must have an unique solution. We now verify $H_l = l(n-l)$ is indeed the right solution by checking (3.2). The length of the path $(i-1, l+i) = l+1$ and the length of $(i+1, l+i)$ is $l-1$.

$$\begin{aligned} H(i, i+l) &= \frac{1}{2}((l+1)(n-(l+1))+1) + \frac{1}{2}((l-1)(n-(l-1))+1 \\ &= \frac{1}{2}(nl + n - l^2 - 2l - 2 + 1 + nl - n - l^2 + 2l - 2 + 1) + 1 \\ &= nl - l^2 = l(n-l). \end{aligned}$$

□

Example 3.2.6. Consider a complete graph on vertices $(0, 1, \cdots, n-1)$. The hitting time is given by $H(i, j) = n - 1$.

Proof. Since all vertices are connected to each other, it is sufficient to find $H(0, 1)$. The probability that we choose vertex v_1 from any other vertex is $\frac{1}{n-1}$. Then for every step that we do not choose v_1, we choose any other vertex with probability $\frac{n-2}{n-1}$. Putting these together, the probability that we start at vertex v_0 and reach vertex v_1 in t steps is given by

$$\mathbb{P}(X_t = v_1 | X_0 = v_0, X_k \neq v_1 \text{ for } k < t) = \frac{1}{n-1}\left[\frac{n-2}{n-1}\right]^{t-1}.$$

The hitting time $H(0, 1)$ is

$$H(0,1) = \sum_{t=1}^{\infty} t \cdot \frac{1}{n-1}\left[\frac{n-2}{n-1}\right]^{t-1} = n - 1.$$

We take advantage of geometric series to prove this. Let $S = H(0, 1)$.

$$\begin{aligned}
S &= \sum_{t=1}^{\infty} t \cdot \frac{1}{n-1}\left[\frac{n-2}{n-1}\right]^{t-1} \\
&= \frac{1}{n-1} + \frac{2}{n-1}\frac{n-2}{n-1} + \frac{3}{n-1}\left[\frac{n-2}{n-1}\right]^2 + \cdots \\
\frac{n-2}{n-1}S &= \frac{1}{n-1}\frac{n-2}{n-1} + \frac{2}{n-1}\left[\frac{n-2}{n-1}\right]^2 + \frac{3}{n-1}\left[\frac{n-2}{n-1}\right]^3 + \cdots \\
S - \frac{n-2}{n-1}S &= \frac{1}{n-1} + \frac{n-2}{n-1}\frac{1}{n-1} + \left[\frac{n-2}{n-1}\right]^2\frac{1}{n-1} + \left[\frac{n-2}{n-1}\right]^3\frac{1}{n-1} + \cdots \\
\frac{1}{n-1}S &= \frac{1}{n-1} + \frac{n-2}{n-1}\frac{1}{n-1}\left(1 + \frac{n-2}{n-1} + \left[\frac{n-2}{n-1}\right]^2 + \cdots\right)
\end{aligned}$$

For $0 < r < 1$, the geometric sum $(1 + r + r^2 + \cdots)$ is given by $\frac{1}{1-r}$. If we set $r = \frac{n-2}{n-1}$, then the sum

$$\frac{1}{n-1}S = \frac{1}{n-1}\left(1 + \frac{n-2}{n-1} + \left[\frac{n-2}{n-1}\right]^2 + \cdots\right) = \frac{1}{n-1}\left[\frac{1}{1-\frac{n-2}{n-1}}\right] = 1.$$

And so,

$$S = n - 1.$$

□

3.2.1 Eigenvalue connection

In general, the probability transition matrix P for a random walk on graph $\mathcal{G}$ is not symmetric. Suppose we define

$$N = D^{1/2}PD^{-1/2} = D^{1/2}D^{-1}AD^{-1/2} = D^{-1/2}AD^{-1/2}$$

. Then N is symmetric and has a spectral decomposition of orthonormal eigenvectors.

$$N = D^{1/2}PD^{-1/2} = \sum_{k=1}^{n} \lambda_k \nu_k \nu_k^T, \tag{3.3}$$

where λ_k are the eigenvalues of N and ν_k are the corresponding eigenvectors. And

$$P = D^{-1/2}ND^{1/2} = D^{-1/2} \sum_{k=1}^{n} \lambda_k \nu_k \nu_k^T D^{1/2}. \tag{3.4}$$

Since P and N are similar, both P and N have the same eigenvalues, but different eigenvectors. Suppose v is an eigenvector of P with eigenvalue λ, we have $Pv = \lambda v$. For $\nu = D^{1/2}v$,

$$N\nu = D^{1/2}PD^{-1/2}\nu = D^{1/2}Pv = D^{1/2}\lambda v = \lambda\nu.$$

So, ν is an eigenvector of N with eigenvalue λ.

By Theorem 3.1.8, we know that P has right and left Perron vectors $\mathbf{1}$ and π respectively, where $\pi(i) = d(v_i)/2m$ is the the stationary distribution. Let

$\omega = [\omega_i] = [\sqrt{d(v_i)}]$, where $d(v_i)$ is the degree of vertex v_i. ω is an eigenvector of N corresponding to eigenvalue 1.

$$N\omega = D^{1/2}PD^{-1/2}\omega = D^{1/2}P\mathbf{1} = \omega.$$

From Theorem 2.5.7, the spectral radius of P is one and one is a simple eigenvalue of P and hence of N. Let $\lambda_1 = 1$. Then, we order the eigenvaules of N as $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n > -1$. The corresponding eigenvector, ν_1 is a unit vector:

$$\nu_1 = \frac{\omega}{\|\omega\|} = \left[\sqrt{\frac{d(v_i)}{2m}}\right] = \sqrt{\pi} = [\sqrt{\pi(i)}], \text{ since}$$

$$\|\omega\| = \sqrt{\sqrt{(d_1)}^2 + \sqrt{(d_2)}^2 + \cdots + \sqrt{(d_n)}^2} = \sqrt{2m}.$$

In Theorem 2.5.8, we used Perron-Frobenius theorem to show that P has a limiting value. We show the same using the symmetric matrix N.

$$N^t = \underbrace{D^{1/2}PD^{-1/2}\ D^{1/2}PD^{-1/2}\cdots}_{t\ times} = D^{1/2}P^tD^{-1/2}.$$

We rewrite this as

$$P^t = D^{-1/2}N^tD^{1/2} = \sum_{k=1}^{n} \lambda_k^t D^{-1/2}\nu_k\nu_k^T D^{1/2}.$$

For $k = 1$, we have

$$D^{-1/2}\lambda_1\nu_1\nu_1^T D^{1/2} = D^{-1/2}\begin{bmatrix} \vdots \\ \sqrt{\frac{d(v_i)}{2m}} \\ \vdots \end{bmatrix}\begin{bmatrix} \cdots & \sqrt{\frac{d(v_i)}{2m}} & \cdots \end{bmatrix} D^{1/2}.$$

So,

$$[P]_{ij}^t = \pi_j + \left[\sum_{k=2}^{n} \lambda_k^t D^{-1/2}\nu_k\nu_k^T D^{1/2}\right]_{ij}.$$

Since $|\lambda_k| < 1$, for $k > 1$, $\lim_{t\to\infty} \lambda_k^t = 0$. Hence

$$[P]_{ij}^t \to \pi_j,\ (t \to \infty).$$

3.2.2 Spectra and hitting time

The hitting time $H(i,j)$ between any two vertices $v_i, v_j \in V$ is the expected number of steps that a random walk starting at vertex v_i takes to reach vertex v_j. While this is a probability based definition, following Lovasz's survey [Lov93], we derive a spectral formula for $H(i,j)$, based on the properties of the probability transition matrix P.

Using (2.17), we express the hitting time in terms of the fundamental matrix $\mathbf{Q}$ as

$$\mathbf{H}_{ij} = \frac{\mathbf{Q}_{jj} - \mathbf{Q}_{ij}}{\pi_j}, \tag{3.5}$$

where $\mathbf{Q} = (I - P + \Pi)^{-1}$. Since $P = D^{-1/2} N D^{1/2}$ and $\Pi = D^{-1/2}\nu_1\nu_1^T D^{1/2}$,
$(I - P + \Pi) = D^{-1/2}\left(I - \sum_{k=1}^{n} \lambda_k \nu_k \nu_k^T + \nu_1\nu_1^T\right) D^{1/2}$,

$$\mathbf{Q} = (I - P + \Pi)^{-1} = D^{-1/2}\left(I - \sum_{k=1}^{n} \lambda_k \nu_k \nu_k^T + [\pi]\right)^{-1} D^{1/2}$$

, where ν_i are the orthonormal eigenvectors of N. Since $\lambda_1 = 1$ and $\nu_1\nu_1^T = [\pi]$, we simplify and write

$$\mathbf{Q} = D^{-1/2}\left(I - \sum_{k=2}^{n} \lambda_k \nu_k \nu_k^T\right)^{-1} D^{1/2}.$$

Since matrix inverses are unique, if we find the inverse of the middle term in the above equation, we have $\mathbf{Q}$. Suppose we set $X = \left(I - \sum_{k=2}^{n} \lambda_k \nu_k \nu_k^T\right)$, we are interested in finding $Y \in M_n$ such that $XY = YX = I$. Thus

$$\begin{aligned}
\left(I - \sum_{k=2}^{n} \lambda_k \nu_k \nu_k^T\right) \quad \cdot \quad & \left(I + \sum_{k=2}^{n} \frac{\lambda_k \nu_k \nu_k^T}{1-\lambda_k}\right) \\
&= I - \sum_{k=2}^{n} \lambda_k \nu_k \nu_k^T + \sum_{k=2}^{n} \frac{\lambda_k}{1-\lambda_k} \nu_k \nu_k^T - \sum_{k=2}^{n} \frac{\lambda_k^2}{1-\lambda_k} \nu_k \nu_k^T \\
&= I - \sum_{k=2}^{n} \lambda_k \nu_k \nu_k^T + \sum_{k=2}^{n} \frac{\lambda_k}{1-\lambda_k} \nu_k \nu_k^T (1-\lambda_k) \\
&= I,
\end{aligned}$$

allows us to write

$$\mathbf{Q} = D^{-1/2}\left(I + \sum_{k=2}^{n} \frac{\lambda_k \nu_k \nu_k^T}{1-\lambda_k}\right) D^{1/2} = I + D^{-1/2} \sum_{k=2}^{n} \frac{\lambda_k \nu_k \nu_k^T}{1-\lambda_k} D^{1/2}.$$

And

$$\mathbf{Q}_{jj} = 1 + \sum_{k=2}^{n} \frac{\lambda_k}{1-\lambda_k} \nu_{kj}^2. \tag{3.6}$$

$$\mathbf{Q}_{ij} = \sum_{k=2}^{n} \frac{\lambda_k}{1-\lambda_k} \nu_{ki}\nu_{kj} \sqrt{\frac{d(v_j)}{d(v_i)}}. \tag{3.7}$$

Since ν_i are orthonormal, $\langle \nu_i, \nu_i \rangle = 1$ for any i. So,

$$\sum_{k=2}^{n} \nu_{kj}^2 = 1 - \nu_{1j}^2 = 1 - [D^{1/2} v_1]_j^2 = 1 - \frac{d(v_j)}{2m}.$$

This allows us to write

$$\sum_{k=2}^{n} \frac{1-\lambda_k}{1-\lambda_k} \nu_{kj}^2 = 1 - \frac{d(v_j)}{2m},$$

yielding

$$1 + \sum_{k=2}^{n} \frac{\lambda_k}{1-\lambda_k} \nu_{kj}^2 = \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{kj}^2 + \frac{d(v_j)}{2m}. \tag{3.8}$$

Similarly, $\langle \nu_i, \nu_j \rangle = 0$ for any $i, j, i \neq j$. So,

$$\sum_{k=2}^{n} \nu_{ki}\nu_{kj} = -\nu_{1i}\nu_{1j} = -\sqrt{\frac{d(v_i)}{2m}} \sqrt{\frac{d(v_j)}{2m}}.$$

Hence

$$\sum_{k=2}^{n} \frac{1-\lambda_k}{1-\lambda_k} \nu_{ki}\nu_{kj} = -\sqrt{\frac{d(v_i)}{2m}} \sqrt{\frac{d(v_j)}{2m}}.$$

And

$$\sum_{k=2}^{n} \frac{\lambda_k}{1-\lambda_k} \nu_{ki}\nu_{kj} \sqrt{\frac{d(v_j)}{d(v_i)}} = \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{ki}\nu_{kj} \sqrt{\frac{d(v_j)}{d(v_i)}} + \frac{d(v_j)}{2m}. \tag{3.9}$$

Subtracting (3.8) from (3.9) gives

$$\begin{aligned} \mathbf{Q}_{jj} - \mathbf{Q}_{ij} &= \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{kj}^2 + \frac{d(v_j)}{2m} - \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{ki}\nu_{kj} \sqrt{\frac{d(v_j)}{d(v_i)}} - \frac{d(v_j)}{2m} \\ &= \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{kj}^2 - \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{ki}\nu_{kj} \sqrt{\frac{d(v_j)}{d(v_i)}}. \end{aligned} \tag{3.10}$$

We now derive formula for hitting time $H(i,j)$ by substituting (3.10) back into (3.5).

$$\begin{aligned} H(i,j) &= \frac{2m}{d(v_j)} \cdot (\mathbf{Q}_{jj} - \mathbf{Q}_{ij}) \\ &= \frac{2m}{d(v_j)} \left(\sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{kj}^2 - \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \nu_{ki}\nu_{kj} \sqrt{\frac{d(v_j)}{d(v_i)}} \right) \\ &= 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(\frac{\nu_{kj}^2}{d(v_j)} - \frac{\nu_{ki}\nu_{kj}}{\sqrt{d(v_i)d(v_j)}} \right). \end{aligned} \tag{3.11}$$

The spectral formula for commute time is computed to be

$$\begin{aligned} C(i,j) &= H(i,j) + H(j,i) \\ &= 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(\frac{\nu_{kj}^2}{d(v_j)} - \frac{\nu_{ki}\nu_{kj}}{\sqrt{d(v_i)d(v_j)}} \right) + 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(\frac{\nu_{ki}^2}{d(v_i)} - \frac{\nu_{ki}\nu_{kj}}{\sqrt{d(v_i)d(v_j)}} \right) \\ &= 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(\frac{\nu_{kj}^2}{d(v_j)} - \frac{\nu_{ki}\nu_{kj}}{\sqrt{d(v_i)d(v_j)}} \right) + \left(\frac{\nu_{ki}^2}{d(v_i)} - \frac{\nu_{ki}\nu_{kj}}{\sqrt{d(v_i)d(v_j)}} \right) \\ &= 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(\frac{\nu_{kj}}{\sqrt{d(v_j)}} - \frac{\nu_{ki}}{\sqrt{d(v_i)}} \right)^2. \end{aligned} \tag{3.12}$$

Example 3.2.7. Consider the graph in Example 3.1.5 with probability transition matrix:

$$P = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}$$

The symmetric matrix $N = D^{-1/2}PD^{1/2}$ is

$$N = \begin{bmatrix} 0 & 0.3536 & 0.2500 & 0.2887 & 0.2887 \\ 0.3536 & 0 & 0.3536 & 0 & 0 \\ 0.2500 & 0.3536 & 0 & 0.2887 & 0.2887 \\ 0.2887 & 0 & 0.2887 & 0 & 0.3333 \\ 0.2887 & 0 & 0.2887 & 0.3333 & 0 \end{bmatrix}.$$

Using the eigen decomposition of N we compute $\mathbf{Q}$ to be

$$\mathbf{Q} = \begin{bmatrix} 0.8219 & 0.0859 & 0.0219 & 0.0352 & 0.0352 \\ 0.1719 & 0.9609 & 0.1719 & -0.1523 & -0.1523 \\ 0.0219 & 0.0859 & 0.8219 & 0.0352 & 0.0352 \\ 0.0469 & -0.1016 & 0.0469 & 0.8789 & 0.1289 \\ 0.0469 & -0.1016 & 0.0469 & 0.1289 & 0.8789 \end{bmatrix}.$$

And the hitting time matrix $\mathbb{H}$ is

$$\mathbb{H} = \begin{bmatrix} 0 & 7.0000 & 3.2000 & 4.5000 & 4.5000 \\ 2.6000 & 0 & 2.6000 & 5.5000 & 5.5000 \\ 3.2000 & 7.0000 & 0 & 4.5000 & 4.5000 \\ 3.1000 & 8.5000 & 3.1000 & 0 & 4.0000 \\ 3.1000 & 8.5000 & 3.1000 & 4.0000 & 0 \end{bmatrix}.$$

The commute time $C(i,j) = H(i,j) + H(j,i)$ is

$$\mathbb{C} = \begin{bmatrix} 0 & 9.6000 & 6.4000 & 7.6000 & 7.6000 \\ 9.6000 & 0 & 9.6000 & 14.0000 & 14.0000 \\ 6.4000 & 9.6000 & 0 & 7.6000 & 7.6000 \\ 7.6000 & 14.0000 & 7.6000 & 0 & 8.0000 \\ 7.6000 & 14.0000 & 7.6000 & 8.0000 & 0 \end{bmatrix}.$$

Clearly, $H(i,j) \neq H(j,i)$ for every pair $(i,j) \in E$ but $C(i,j) = C(j,i)$ for every pair.

CHAPTER 4

GOOGLE PAGERANK

Sergey Brin and Larry Page introduced Google in 1998, a time when the pace at which the web was growing began to outstrip the ability of current search engines to yield useable results. One factor that set Google's search engine apart from others was that its search listings always listed the "good stuff" on the top. Within the first page of search results, most user's query was answered. Search engines like Inktomi, Alta Vista, etc. focused only on a page's content, meta tags, and density of keywords. But Google's additional focus on the hyperlink structure of the web allowed it to rank the popularity of every indexed page on the web, and thereby present the most popular pages at the top of the results. The number of active pages on the web as of October 2013 is 1.93 billion pages [cita] and it continues to grow. As new pages are added, an efficient search engine has the daunting task of indexing these pages so they are returned in a user's search query. As of October 2013, Google has the maximum number of indexed pages and 73% [citb] of the market share of all searches. The huge market share is due to the combination of Google's technology and the computational algorithms that support the search process. In particular, Google's success is due to its method for computing the popularity of a webpage, i.e., the PageRank of every page on the web. The importance of PageRank is emphasized in one of Google's web pages:

The heart of our software is PageRank$^{\text{TM}}$*, a system for ranking web pages developed by our founders Larry Page and Sergey Brin at Stanford*

University. And while we have dozens of engineers working to improve every aspect of Google on a daily basis, PageRank continues to provide the basis for all of our web search tools. [citc]

Google's search engine has three main components: web crawler, indexer, and query processor. Google's web crawlers are bots that navigate through domain servers everyday and fetch URLs of newly created and modified webpages. The indexer then parses each page for searchable words and stores the resulting index of words in its database. Anytime a user submits a query, the query processor uses this large database to compile a list of pages, in order of relevancy, to present to the user. The order of relevancy is decided by the PageRank of each of the webpage in the search result. In this chapter, we look at the core ideas behind how Google calculates the PageRank.
We follow [AL06] in this chapter.

4.1 PageRank

The roots of PageRank actually derive from bibliometrics research, the analysis of the citation structure among academic papers. Let $inN(i)$ be the set of pages linking into page P_i and $outN(i)$ be the set of pages that P_i links to. The PageRank of a page P_i, denoted $r(P_i)$, is the sum of the PageRank of all the pages linking into P_i.

$$r(P_i) = \sum_{P_j \in inN(i)} \frac{r(P_j)}{|outN(j)|}. \tag{4.1}$$

But $r(P_i)$ depends on the PageRank of other pages and is computed iteratively. The PageRank algorithm starts by assigning a rank of $\frac{1}{n}$ to every page, where n is the total number of pages on the web; $r_0(P_i) = 1/n$ for every page i. The PageRank at

the $k+1$ iteration is given by the equation

$$r_{k+1}(P_i) = \sum_{P_j \in inN(i)} \frac{r_k(P_j)}{|outN(j)|}. \tag{4.2}$$

Example 4.1.1. Let us consider a simple web as shown in Figure 4.1

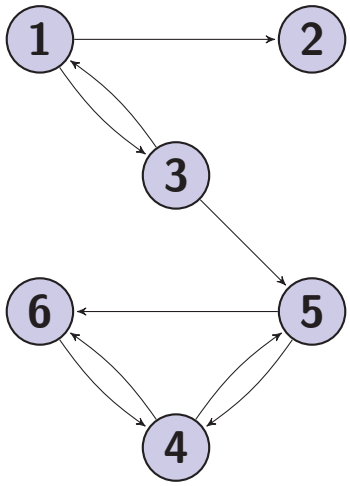

Figure 4.1: Web with six pages.

We compute the rank of each of these six pages using (4.2) as shown in Table 4.1.

Table 4.1: Ranking of pages after two iterations

Page	Iteration 0	Iteration 1	Iteration 2	Rank
1	1/6	1/12	1/24	4
2	1/6	1/12	1/24	4
3	1/6	1/12	1/24	4
4	1/6	1/4	1/4	1
5	1/6	1/6	1/6	3
6	1/6	1/6	1/5	2

The iterative method, while simple and straightforward for a small set of pages, is inefficient considering the size of the web. Suppose we visualize the web as a large graph, each page on the web as a vertex, the hyperlinks between pages as directed edges between two vertices, and a user surfing the web by making arbitrary clicks on the hyperlinks as taking a random walk on this web graph. In this context, we are able to apply the theory of random walk to model a user browsing the web. Let r be the PageRank vector. The iterative equation to compute the rank of page i at the kth iteration is

$$r_k(i) = \sum_{j \sim i} \frac{r_{k-1}(j)}{d(v_j)}.$$

We write this using the matrix notation as

$$r_{k+1}^T = r_k^T \cdot K, \text{ where } K \text{ is a probability transition matrix.} \tag{4.3}$$

If the sequence $r_1, r_2, \cdots r_k \cdots$ converges uniquely, then we have a stable PageRank vector. But there is no guarantee that this sequence of rankings converges uniquely. In Chapter 3, we saw that if the probability transition matrix of a Markov chain on a graph is stochastic, irreducible, and primitive, then it has a unique stationary distribution. Furthermore, the powers of the probability transition matrix also converges to a matrix with rows as the stationary distribution.

Since the importance of a page or its PageRank is measured by its popularity (how many incoming links it has), we view the importance of page i as the probability that a random surfer on the Internet opens a browser to any page and follows the hyperlinks, visits page i. We interpret the weights we assigned to the edges of the graph in a probabilistic way and model the process as a random walk on graphs. Each page has equal probability ($1/n$, where n is the number of indexed pages) to be chosen as a starting point. So, the initial probability distribution is given by the column vector: $r_0 = [1/n\ 1/n \cdots 1/n]^T$. The probability that page i is

visited after one step is equal to $r_0^T K$ and so on. The probability that page i is visited after k steps is equal to $r_0^T K^k$. The sequence $r_0^T K, r_0^T K^2, \cdots, r_0^T K^k, \cdots$ converges in this case to a unique probabilistic vector π. Moreover, the ith entry in the vector π is simply the probability that at each moment a random surfer visits page i. Hence, is the relative importance or rank of that page. In this context, π is the stationary distribution and is our PageRank vector.

4.2 Matrices of the webgraph

We define the following matrices for a webgraph.

Definition 4.2.1. Degree matrix.

Let D be the diagonal *out-degree matrix* with $D_{ii} = |outN(i)|$, the number of pages having an hyperlink on page i. We define the generalized inverse D^g as

$$D^g = \begin{cases} 1/D_{ii}, \text{ if } D_{ii} \neq 0. \\ 0, \text{ otherwise .} \end{cases}$$

Definition 4.2.2. Adjacency matrix.

A is the *adjacency matrix* of the web graph with

$$A_{ij} = \begin{cases} 1, & \text{if } j \in outN(i). \\ 0, & \text{otherwise.} \end{cases}$$

Definition 4.2.3. Hyperlink matrix.

The $n \times n$ *hyperlink matrix* $K = D^g A$ is the weighted probability transition matrix of the web graph with

$$K_{ij} = \begin{cases} 1/D_{ii}, & \text{if } j \in outN(i). \\ 0, & \text{otherwise.} \end{cases}$$

Most pages on the web are not linked to each other, leaving more *zeros* in every row. Furthermore, there are pages with no outlinks like image files, document files, etcetra. These pages called *dangling nodes*, create *zero* rows in the matrix leaving K sparsely populated. The random surfer model makes K the probability transition matrix of a random walk, but the *zero* rows make the hyperlink matrix K sub-stochastic.

4.3 Problems with the hyperlink matrix

The web graph is not connected and the hyperlink matrix K is sub stochastic and possibly reducible and periodic. Consider a small web graph with three nodes as shown in Figure 4.2. Suppose we start with the initial uniform distribution

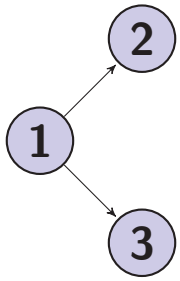

Figure 4.2: Dangling nodes.

$r_0(i) = 1/3$, for all nodes i. We see that in two iterations we arrive at the *zero* vector.

$$r_1^T = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/6 & 1/6 \end{bmatrix};$$

$$r_2^T = \begin{bmatrix} 0 & 1/6 & 1/6 \end{bmatrix} \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

The web graph may have parts that are not reachable after some time. In our previous example from Figure 4.1, we easily see that once we reach node 5, we are not able to get back to nodes $1, 2$ or 3. Consequently, by the twelfth iteration, we get $r_{12} = \begin{bmatrix} 0 & 0 & 0 & 0.2962 & 0.1481 & 0.2222 \end{bmatrix}$. Nodes 5 and 6 are rank sinks and start accumulating rank as the iterations continue.

It is also possible for the web graph to have cycles as shown in Figure 4.3.

Figure 4.3: Cycles.

In such a case, the initial distribution is $r_0 = [1\ 0]^T$, $r_1 = [0\ 1]^T$, $r_2 = [1\ 0]^T$. The iterates do not converge for any k, toggling between $[1\ 0]^T$ when k is even and $[0\ 1]^T$ when k is odd.

4.4 Adjustments to the model

The web graph is disconnected with more than one connected component and reducible. Parts of the graph are cycles, making the return times periodic. The dangling nodes have no outlinks and are isolated. The hyperlink matrix of the web graph is sub-stochastic, reducible, and periodic. So, the hyperlink matrix K does not have a unique stationary distribution, and the powers of the matrix may not converge. These problems caused Brin and Page to make adjustments to the basic model.

4.4.1 Stochastic adjustment

From a dangling node, a random surfer accesses any page with equal probability by typing in the page URL. In terms of the web graph, this implies that

there is a directed edge from every dangling node to every other page on the web. The first adjustment, the stochastic adjustment, replaces every row that has all *zero* entries, corresponding to a dangling node, with a row that has all $1/n$ entries.
Let $\mathbf{a}$ be a column vector defined as

$$\mathbf{a_i} = \begin{cases} 1, \text{ if } i \text{ is a dangling node.} \\ 0, \text{ otherwise.} \end{cases}$$

Define a new matrix S as

$$S = K + (1/n)\mathbf{a}\mathbf{1}^T.$$

This modified hyperlink matrix is stochastic. Hence, it is the probability transition matrix of a Markov chain. The stochastic adjustment is just a rank-one update to the original hyperlink matrix.

4.4.2 Primitivity adjustment

A random surfer on the web follows the hyperlink structure usually. But, at times, "the user teleports" to a randomly chosen page by typing in the URL of the new destination and follows the hyperlink structure until the next teleportation. In terms of the hyperlink matrix, this implies that there is a connectivity between any two pages on the web, however small the probability. Brin and Page modeled this mathematically using a teleportation factor α and designed a new matrix G, the Google matrix.

$$G = \alpha S + (1 - \alpha)1/n\mathbf{J}, \tag{4.4}$$

where α takes values between 0 and 1 and $\mathbf{J} = \mathbf{1} \cdot \mathbf{1}^T$ is the all one matrix. α indicates the time that a person spends following the hyperlink structure before teleporting. Suppose $\alpha = 0.6$, then the surfer follows the hyperlink structure 60% of

the time. Since the teleportation matrix $E = 1/n\mathbf{J}$ is positive and uniform, the surfer is equally likely to "teleport" to any page on the web.
The new Google matrix G has the following properties:

Proposition 4.4.1. G is stochastic.

Proof. G is the convex combination of two stochastic matrices S and $1/n\mathbf{J}$.

□

Proposition 4.4.2. G is irreducible.

Proof. Since $\mathbf{J}$ is positive and S is non-negative, G is positive. This implies that every page is directly connected to every other page. The entire web is one large connected component. So irreducibility is enforced by design.

□

Proposition 4.4.3. G is aperiodic.

Proof. Every page has a self-loop. So, for every page, there is a return time of one step; one leaves the page and returns in the next step. The set of return times for every page has 1 as the first entry. Hence the *gcd* of this set is 1 and consequently every page has period one.

□

Proposition 4.4.4. G is primitive.

Proof. $G^L > 0$ for some positive integer L. In fact, this is true for $L = 1$.

□

Proposition 4.4.5. *G has a stationary distribution vector.*

Proof. Since G is primitive, by Markov chain theory G has a stationary distribution vector π. So,

$$\pi^T G = \pi^T, \tag{4.5}$$

and

$$\lim_{m\to\infty} G^m \to \pi^T \mathbf{1}.$$

Using (4.4) the PageRank equation (4.5) be rewritten as

$$\begin{aligned} \pi^T &= \pi^T \alpha S + (1-\alpha)\pi^T \frac{1}{n}\mathbf{1}\mathbf{1}^T \\ &= \pi^T \alpha S + \frac{(1-\alpha)}{n}\mathbf{1}^T, \text{ since } \pi^T\mathbf{1} = 1. \end{aligned} \tag{4.6}$$

□

Proposition 4.4.6. *G is completely dense, but be written as a sum of the sparse hyperlink matrix K and a rank-one matrix.*

Proof.

$$\begin{aligned} G &= \alpha S + (1-\alpha)\frac{1}{n}\mathbf{J} \\ &= \alpha(K + \mathbf{a}\frac{1}{n}\mathbf{1}^T) + (1-\alpha)\frac{1}{n}\mathbf{J}, \ \mathbf{a} \text{ is the dangling node vector} \\ &= \alpha K + \left(\frac{\alpha}{n}\mathbf{a} + \frac{(1-\alpha)}{n}\mathbf{1}\right)\mathbf{1}^T. \end{aligned} \tag{4.7}$$

□

Proposition 4.4.7. If the spectrum of the stochastic matrix S is $\{1, \lambda_2, \lambda_3, \cdots, \lambda_n\}$, then the spectrum of the Google matrix G is $\{1, \alpha\lambda_1, \alpha\lambda_2, \cdots, \alpha\lambda_n\}$.

Proof. Since S is stochastic, $(1, \mathbf{1})$ is an eigenpair of S. Set $\mathbf{M} = \begin{bmatrix} \mathbf{1} & \mathbf{X} \end{bmatrix}$, where $\mathbf{X}$ is a non-singular matrix. Let $\mathbf{M}^{-1} = \begin{bmatrix} y^T \\ \mathbf{Y}^T \end{bmatrix}$.

$$\mathbf{M}^{-1}\mathbf{M} = \begin{bmatrix} y^T\mathbf{1} & y^T\mathbf{X} \\ \mathbf{Y}^T\mathbf{1} & \mathbf{Y}^T\mathbf{X} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{I} \end{bmatrix}$$

. This gives us two identities $y^T\mathbf{1} = 1$ and $\mathbf{Y}^T\mathbf{1} = 0$. We look at the similarity transformation

$$\mathbf{M}^{-1}\mathbf{S}\mathbf{M} = \begin{bmatrix} y^T\mathbf{1} & y^T\mathbf{S}\mathbf{X} \\ \mathbf{Y}^T\mathbf{1} & \mathbf{Y}^T\mathbf{S}\mathbf{X} \end{bmatrix} = \begin{bmatrix} 1 & y^T\mathbf{S}\mathbf{X} \\ 0 & \mathbf{Y}^T\mathbf{S}\mathbf{X} \end{bmatrix}.$$

This triangulation of $\mathbf{S}$ reveals that $\mathbf{Y}^T\mathbf{S}\mathbf{X}$ contains the remaining eigenvalues $\lambda_2, \cdots, \lambda_n$ of $\mathbf{S}$. Applying the same similarity transformation to $\mathbf{G} = \alpha S + \frac{(1-\alpha)}{n}\mathbf{1}\mathbf{1}^T$, we get

$$\begin{aligned}
\mathbf{M}^{-1}\left(\alpha S + \frac{(1-\alpha)}{n}\mathbf{J}\right)\mathbf{M} &= \mathbf{M}^{-1}\left(\alpha S\right)\mathbf{M} + (1-\alpha)\left(\mathbf{M}^{-1}(\mathbf{1})(\frac{1}{n}\mathbf{1})^T\mathbf{M}\right) \\
&= \alpha \begin{bmatrix} 1 & y^T\mathbf{S}\mathbf{X} \\ 0 & \mathbf{Y}^T\mathbf{S}\mathbf{X} \end{bmatrix} + (1-\alpha) \begin{bmatrix} y^T\mathbf{1} \\ \mathbf{Y}^T\mathbf{1} \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{1}^T\mathbf{1} & \frac{1}{n}\mathbf{1}^T\mathbf{X} \end{bmatrix} \\
&= \alpha \begin{bmatrix} 1 & y^T\mathbf{S}\mathbf{X} \\ 0 & \mathbf{Y}^T\mathbf{S}\mathbf{X} \end{bmatrix} + (1-\alpha) \begin{bmatrix} 1 & \frac{1}{n}\mathbf{1}^T\mathbf{X} \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & y^T(\alpha\mathbf{S} + \frac{(1-\alpha)}{n}\mathbf{1}^T)\mathbf{X} \\ 0 & \alpha\mathbf{Y}^T\mathbf{S}\mathbf{X} \end{bmatrix}
\end{aligned}$$

Therefore, the eigenvalues of $\mathbf{G}$ are $\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\}$.

□

4.5 Computation of PageRank

The Google matrix G is irreducible and primitive. By the Markov chain theory, a unique stationary distribution vector exists. From Chapter 3, we know that the stationary distribution vector is the left eigenvector of G corresponding to the largest eigenvalue 1. In theory, the PageRank vector be computed in two ways:

1. Solve the eigenvector problem for π^T.

$$\pi^T = \pi^T G,$$

$$\pi^T \mathbf{1} = 1.$$

2. Solve the linear homogeneous system for π^T.

$$\pi^T (I - G) = 0^T.$$

$$\pi^T \mathbf{1} = 1.$$

The first system requires us to find the dominant eigenpair of G, while the second method requires us to solve n homogeneous linear equations. The Google matrix is large and dense. Hence the eigen decomposition of G is computationally intensive. Other more advanced and computationally efficient numerical methods exist to solve the same equation [W.S94].

The power method is one of the oldest and simplest iterative methods for finding the dominant eigenpair of a matrix. But this is also the slowest as the matrix may not converge fast. Appendix E contains the computational mechanics of the power method. There are three main reasons why Brin and Page chose to implement the power method for computing the PageRank.

1. We saw in (4.7) that the Google matrix G be expressed in terms of the sparse K. Hence

$$\begin{aligned}\pi_{k+1}^T &= \pi_k^T G \\ &= \alpha\pi_k^T K + \frac{1}{n}(\alpha\pi_k^T \mathbf{a} + 1 - \alpha)\mathbf{1}^T.\end{aligned}$$

   To implement the power method, the matrix multiplication is performed on the sparse hyperlink matrix K, making it computationally cheaper.

2. The power method is also storage friendly. The sparse hyperlink matrix K, dangling node vector $\mathbf{a}$ and the current iterate π_k are the only elements that are stored. The PageRank vector is completely dense. Given the size of the web, storage requirements are a major factor.

3. The Google matrix G converges in 50 iterations. By Theorem 4.4.7, the eigenvalues of G are $\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\}$ where λ_i are the eigenvalues of S. From Appendix E, we see that the convergence rate of G^k depends on the ratio $\frac{\alpha|\lambda_2|}{|\lambda_1|} = \alpha\lambda_2$. Both α and λ_2 are less than one. If we consider $|\lambda_2| \approx 1$, then we are interested in finding out when $\alpha^k \to 0$. α the teleportation constant is an artificial manipulator that Google controls and has been set to 0.85. $0.85^{50} \approx 0.00029576$, which is as close to zero as Google wants. Since the operations are performed on the sparse hyperlink matrix, each iteration of matrix-vector multiplication is of order less than $O(n^2)$.

Example 4.5.1. We now compute the PageRank for the small web graph in Example 4.1.1.

The hyperlink matrix $K = D^g A$ is

$$K = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Only node 2 is dangling. So, the dangling node vector is $\mathbf{a^T} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$.

The stochastic matrix $S = K + \frac{1}{6}\mathbf{a}\mathbf{1}^T$ is

$$S = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

For $\alpha = 0.85$, the Google matrix $G = 0.85S + 0.15\frac{1}{6}J$ is

$$G = \begin{bmatrix} 0.0250 & 0.4500 & 0.4500 & 0.0250 & 0.0250 & 0.0250 \\ 0.1667 & 0.1667 & 0.1667 & 0.1667 & 0.1667 & 0.1667 \\ 0.4500 & 0.0250 & 0.0250 & 0.0250 & 0.4500 & 0.0250 \\ 0.0250 & 0.0250 & 0.0250 & 0.0250 & 0.4500 & 0.4500 \\ 0.0250 & 0.0250 & 0.0250 & 0.4500 & 0.0250 & 0.4500 \\ 0.0250 & 0.0250 & 0.0250 & 0.8750 & 0.0250 & 0.0250 \end{bmatrix}.$$

Clearly, G is positive. Hence primitive. We now compute the PageRank vector as the stationary distribution of G using the power method to be

$$\pi = \begin{bmatrix} 0.0577 & 0.0577 & 0.0577 & 0.3504 & 0.2066 & 0.2699 \end{bmatrix},$$

in 15 iterations. Pages 3, 2 and 1 all have the same rank. So, one possible ranking of the pages of this small web is $(4, 6, 5, 3, 1, 2)$.

CHAPTER 5

ELECTRICAL NETWORKS ON GRAPHS

An electrical network is an interconnection of electrical elements such as resistors, inductors, capacitors, voltage sources, current sources, and switches. An electrical circuit is a network consisting of a closed loop, giving a return path for the current. Electric current is the flow of electric charge or the flow of electrons in the direction opposite to the flow of current. Electrical networks and random walks on graphs are both governed by graphs which have values attached to vertices and edges. We view the flow of electrons through a circuit to approximate a random walk through the nodes of the network. This analogy allows us to express hitting time and commute time between vertices in terms of resistance, voltage and current. Doyle and Snell [PGD06] showed that the hitting times of a random walk on an undirected graph $\mathcal{G}(V, E)$ are related to voltages in an electrical network. Current in an electrical circuit has both strength and direction. So we now look at weighted graphs and choose an orientation of the graph in the direction of electron flow.

5.1 Matrices of a weighted graph

Let $\mathcal{G}(V, E, W)$ be a finite, connected, weighted graph, where $V = \{v_i\}$ is the set of vertices, $E = \{(v_i, v_j)\}$, the set of edges, and $W = \{w_{ij}\}$, the set of weights of edges.

Definition 5.1.1. Weighted adjacency matrix.

For graph $\mathcal{G}$, the *weighted adjacency matrix* A is given by

$$[A]_{ij} = \begin{cases} w_{ij}, & \text{if } v_i \sim v_j. \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

Since $\mathcal{G}$ is undirected, A is symmetric.

Definition 5.1.2. Degree matrix.

For graph $\mathcal{G}$, the *degree matrix* D is the diagonal matrix given by

$$[D]_{ii} = d(v_i) = \sum_{v_j \sim v_i} w_{ij} = \sum_j A_{ij}.$$

Definition 5.1.3. Probability transition matrix.

For graph $\mathcal{G}$, the *probability transition matrix* $P = D^{-1}A$ is the matrix given by

$$[P]_{ij} = \frac{w_{ij}}{d(v_i)}.$$

Definition 5.1.4. Edge-Vertex incidence matrix.

Consider an arbitrary but fixed orientation of $\mathcal{G}$. Then B is the signed *edge-vertex incidence matrix* given by

$$B(e, v) = \begin{cases} 1, & \text{if } v \text{ is the tail of } e. \\ -1, & \text{if } v \text{ is the head of } e. \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

Definition 5.1.5. Combinatorial Laplacian.

For graph $\mathcal{G}$, the *combinatorial Laplacian* L is the matrix given by

$$[L]_{ij} = \begin{cases} d(v_i), & \text{if } v_i = v_j. \\ -w_{ij}, & \text{if } v_j \sim v_i. \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

The Laplacian matrix is clearly symmetric. An equivalent definition for the Laplacian is $L = D - A$.

Let W be the diagonal edge matrix, where $W(e, e) = w_e$ is the weight of an edge $e \in E$. We see directly from Definition 5.1.5, that for an edge $e = (v_i, v_j) \in E$,

$$[L]_{ij} = \sum_{e \in E} B(e, v_i) W(e, e) B(e, v_j).$$

We express this in matrix form as

$$L = B^T W B. \tag{5.4}$$

Definition 5.1.6. Normalized Laplacian.

The *normalized Laplacian* $\mathcal{L}$ is defined to be

$$\mathcal{L} = D^{-1/2} L D^{-1/2}. \tag{5.5}$$

Using the definition $L = D - A$, we write the above as

$$\mathcal{L} = D^{-1/2}(D - A)D^{-1/2} = I - D^{-1/2} A D^{-1/2}.$$

And

$$[\mathcal{L}]_{ij} = \begin{cases} 1, & \text{if } v_i = v_j. \\ -\frac{w_{ij}}{\sqrt{d(v_i)d(v_j)}} & v_j \sim v_i. \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

The normalized Laplacian is closely related to the probability transition matrix $P = D^{-1}A$.

$$D^{-1/2} \mathcal{L} D^{1/2} = D^{-1/2}(I - D^{-1/2} A D^{-1/2}) D^{1/2} = I - D^{-1} A = I - P.$$

So, $\mathcal{L}$ and $I - P$ are similar matrices. Since $D^{-1/2} A D^{-1/2} = N$, the symmetric form of P, we express $\mathcal{L}$ using N as $\mathcal{L} = I - N$.

5.1.1 Properties of the Laplacian.

Lemma 5.1.7. *Combinatorial Laplacian L is real and symmetric.*

Proof. Since D and A are real and symmetric, $L = D - A$ is real. Then

$$L^T = (D - A)^T = D^T - A^T = D - A = L,$$

and L is also symmetric.

□

By the spectral theory of real, symmetric matrices, the Laplacian L has real eigenvalues and an orthonormal basis of eigenvectors.

Lemma 5.1.8. *Combinatorial Laplacian L is positive semi-definite.*

Proof. From matrix theory, a matrix $M \in \mathbb{R}^n$ is *positive semi-definite* if for any non-zero vector, $x \in \mathbb{R}^n$, $x^T M x \geq 0$. Let x be any real, non-zero vector. From (5.4)

$$x^T L x = x^T B^T W B x = (Bx)^T W (Bx) = \sum_{e=(v_i,v_j)} w_e (x_i - x_j)^2 \geq 0, \tag{5.7}$$

since $w_e > 0$ for every edge.

□

Lemma 5.1.9. *The normalized Laplacian $\mathcal{L}$ is also real, symmetric and positive semi-definite.*

Proof. The first part follows directly from the properties of L. For the second part, let x be any real, non-zero vector. Then

$$x^T \mathcal{L} x = x^T D^{-1/2} L D^{-1/2} x.$$

Since $y = D^{-1/2}x > 0$, we have

$$(D^{-1/2}x)^T L(D^{-1/2}x) = y^T Ly \geq 0.$$

□

5.1.2 Spectrum of the Laplacian

In general, the properties of the the spectra of the two Laplacians are different, though they share some similarities.

Lemma 5.1.10. *Every eigenvalue of the combinatorial and normalized Laplacian is non-negative.*

Proof. Let μ be an eigenvalue of L associated with eigenvector y. Since L is positive semi-definite,

$$0 \leq y^T Ly = y^T \mu y = \mu y^T y.$$

$\mu = y^T Ly / y^T y \geq 0$, since it is a ratio of non-negative real numbers. Similarly, the normalized Laplacian $\mathcal{L}$ also has non-negative eigenvalues.

□

Lemma 5.1.11. *If graph $\mathcal{G}$ is connected, the null-space of L has dimension one, and is spanned by the vector* $\mathbf{1}$.

Proof. If $x \in null(L)$, then $Lx = 0$. From (5.7) we have

$$x^T Lx = \sum_{v_i \sim v_j} w_{ij}(x_i - x_j)^2 = 0.$$

Since $w_{ij} > 0$, we must have $(x_i - x_j) = 0$ and $x_i = x_j$ for every $(v_i, v_j) \in E$. Since $\mathcal{G}$ is connected, we infer that all x_i's are equal and x is a scalar multiple of $\mathbf{1}$. So the dimension of $null(L) = 1$.

Similarly, the dimension of null-space of $\mathcal{L}$ is also 1 and is spanned by the vector $D^{1/2}\mathbf{1}$.

□

Lemma 5.1.12. *Eigenvalues of L are $0 = \mu_1 < \mu_2 \leq \cdots \leq \mu_n$.*

Proof. The Laplacian L has row sum zero for every row. So, 0 is an eigenvalue of L.

$$L\mathbf{1} = \vec{0}.$$

Furthermore, since graph $\mathcal{G}$ is connected, from Lemma 5.1.11, we conclude that the eigenvalue 0 is simple. Since all the eigenvalues of L are non-negative, the result follows.

□

Lemma 5.1.13. *Eigenvalues of the normalized Laplacian $\mathcal{L}$ are*
$0 = \eta_1 < \eta_2 \leq \cdots \leq \eta_n \leq 2$

Proof. $D^{1/2}\mathbf{1}$ is an eigenvector of normalized Laplacian $\mathcal{L}$ corresponding to eigenvalue 0.

$$\mathcal{L}(D^{1/2}\mathbf{1}) = D^{-1/2}LD^{-1/2}(D^{1/2}\mathbf{1}) = D^{-1/2}L\mathbf{1} = \vec{0}.$$

Furthermore, by Lemma 5.1.11, the dimension of null space of $\mathcal{L}$ is one. Suppose y is an eigenvector of $\mathcal{L}$ corresponding to eigenvalue 0, we have

$$0 = y^T\mathcal{L}y = y^T D^{-1/2}LD^{-1/2}y.$$

Setting $z = D^{-1/2}y$, we get

$$0 = z^T L z.$$

Then z is in the null space of L. Therefore, by Lemma 5.1.11, we have $0 = \eta_1$ is simple.

Since $\mathcal{L}$ is real and symmetric, we apply the Rayleigh-Ritz theorem in pages 176-180 of [RAH85] to find the value of η_n.

$$\eta_n = \sup_{x \neq 0} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} = \sup_{x \neq 0} \frac{\left\langle x, D^{-1/2} L D^{-1/2} x \right\rangle}{\langle x, x \rangle} = \sup_{x \neq 0} \frac{\left\langle D^{-1/2}x, L D^{-1/2} x \right\rangle}{\langle x, x \rangle}.$$

For $y = D^{-1/2}x$,

$$\frac{\langle y, Ly \rangle}{\langle D^{1/2}y, D^{1/2}y \rangle} = \frac{\sum_{i \sim j} w_{ij}(y_i - y_j)^2}{\sum_j y_j^2 d(v_j)}.$$

We know that for any $a, b \in \mathbb{R}$, $(a-b)^2 \leq 2(a^2 + b^2)$. So, $(y_i - y_j)^2 \leq 2(y_i^2 + y_j^2)$. For an undirected, connected graph

$$\sum_{i \sim j} w_{ij}(y_i^2 + y_j^2) = \sum_{i \sim j} w_{ij} y_i^2 + \sum_{j \sim i} w_{ji} y_j^2 = \sum_j y_j^2 d(v_j).$$

Hence

$$\eta_n = \sup_{y \neq 0} \frac{\sum_{i \sim j} w_{ij}(y_i - y_j)^2}{\sum_j y_j^2 d(v_j)} \leq 2.$$

□

5.1.3 Eigenvalues of a graph

Let $\mathcal{G}$ be a connected graph. Suppose α_i are the eigenvalues of the adjacency matrix A, μ_i the eigenvalues of combinatorial Laplacian L, and η_i, the eigenvalues of normalized Laplacian $\mathcal{L}$, then

$$\alpha_1 > \alpha_2 \geq \cdots \geq \alpha_n.$$

We order the eigenvalues of L and $\mathcal{L}$ the reverse:

$$0 = \mu_1 < \mu_2 \leq \cdots \leq \mu_n$$

and

$$0 = \eta_1 < \eta_2 \cdots \leq \eta_n \leq 2.$$

Suppose λ_i are the eigenvalues of N. Since $\mathcal{L} = I - N$, $\lambda_i = 1 - \eta_i$, and

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1.$$

Suppose u is an eigenvector of $\mathcal{L}$ with eigenvalue η, we write

$$[\mathcal{L}u]_j = \frac{1}{\sqrt{d(v_j)}}\left[d(v_j)u_j - \sum_{v_i \sim v_j} u_j w_{ij}\right]\frac{1}{\sqrt{d(v_j)}} = u_j - \sum_{v_i \sim v_j} \frac{u_j w_{ij}}{\sqrt{d(v_i)d(v_j)}}$$

The normalized Laplacian is real symmetric. Hence we write it in terms of its spectral decomposition. Since $\eta_1 = 0$, we have

$$\mathcal{L} = \sum_{i=1}^{n} \eta_i \phi_i \phi_i^T = \sum_{i=2}^{n} \eta_i \phi_i \phi_i^T,$$

where $\{\eta_i\}$ is the set of eigenvalues and $\{\phi_i\}$ is an orthonormal basis of eigenvectors of $\mathcal{L}$.

[Chu94] has more details on the spectral properties of the normalized Laplacian $\mathcal{L}$.

5.2 Inverse of the Laplacian

The normalized Laplacian is more useful in the study of electrical network and effective resistance. So, we focus our attention here on the normalized Laplacian. In particular, we are interested in finding the inverse of $\mathcal{L}$.

5.2.1 Green's function

Green's function was first introduced in a celebrated essay by George Green in March 1828 [Gre]. Since then the concept of Green's function has been used in a wide range of areas, especially in the study of partial differential equations and quantum field theory.

Definition 5.2.1. The Green's function $\mathcal{R}$ [Chu00] denotes the symmetric matrix satisfying

$$y^T\mathcal{L}\mathcal{R} = y^T\mathcal{R}\mathcal{L} = y^T,$$

for all vectors y which are orthogonal to the normalized eigenvector ϕ_1 of $\mathcal{L}$, where $\phi_1 = \frac{D^{1/2}\mathbf{1}}{\sqrt{vol(\mathcal{G})}}$. We have seen earlier in Chapter 2 that ϕ_1 is the stationary distribution of a random walk on a connected graph. Equivalently, the normalized Green's function satisfies

$$\mathcal{R}\mathcal{L} = \mathcal{L}\mathcal{R} = I - \phi_1\phi_1^T, \tag{5.8}$$

and has the following form:

$$\mathcal{R} = \sum_{i=2}^{n} \frac{1}{\lambda_i}\phi_i\phi_i^T.$$

Lemma 5.2.2. *The normalized matrix $\bar{\mathbf{Q}} = D^{1/2}\mathbf{Q}D^{-1/2}$ is the symmetric matrix satisfying the definition of Green's function, where $\mathbf{Q}$ is the fundamental matrix of the connected graph.*

Proof. From subsection 2.6.1 of Chapter 2, we know that $\mathbf{Q} = [I - P + \Pi]^{-1}$. Hence,

$$\begin{aligned}
\bar{\mathbf{Q}} &= D^{1/2}\mathbf{Q}D^{-1/2} \\
&= D^{1/2}[I - P + \Pi]^{-1}D^{-1/2} \\
&= [D^{1/2}(I - P + \Pi)D^{-1/2}]^{-1} \\
&= [D^{1/2}(I - D^{-1}A + \Pi)D^{-1/2}]^{-1} \\
&= [I - N + D^{1/2}\Pi D^{-1/2}]^{-1} \\
&= [\mathcal{L} + D^{1/2}\Pi D^{-1/2}]^{-1}.
\end{aligned}$$

Since I, N and $D^{1/2}\Pi D^{-1/2}$ are symmetric, the matrix $[I - N + D^{1/2}\Pi D^{-1/2}]$ is symmetric. By Theorem B.0.17, $\bar{\mathbf{Q}}$ is also symmetric.

To show that $\bar{\mathbf{Q}}$ satisfies the definition of Green's function, we refer to the properties of the fundamental matrix in Subsection 2.6.1 of Chapter 1. Since $\mathbf{Q}(I - P) = (I - P)\mathbf{Q} = I - \Pi$,

$$\begin{aligned}
\bar{\mathbf{Q}}\mathcal{L} &= D^{1/2}\mathbf{Q}D^{-1/2}D^{1/2}(I - P)D^{-1/2} \\
&= I - D^{1/2}\Pi D^{-1/2} \\
&= I - \phi_1\phi_1^T.
\end{aligned}$$

We see that

$$D^{1/2}\Pi D^{-1/2} = D^{1/2}\mathbf{1}\pi^T D^{-1/2} = \frac{D^{1/2}\mathbf{1}}{\sqrt{vol(\mathcal{G})}}\left[\cdots d(v_i)\cdots\right]\frac{D^{-1/2}}{\sqrt{vol(\mathcal{G})}} = \phi_1\phi_1^T.$$

Similarly,

$$\begin{aligned}
\mathcal{L}\bar{\mathbf{Q}} &= D^{1/2}(I - P)D^{-1/2}D^{1/2}\mathbf{Q}D^{-1/2} \\
&= I - D^{1/2}\Pi D^{-1/2} \\
&= I - \phi_1\phi_1^T.
\end{aligned} \tag{5.9}$$

Hence $\bar{\mathbf{Q}}$ is the Green's function $\mathcal{R}$.

□

5.2.2 Inverse of normalized Laplacian

The normalized Laplacian $\mathcal{L}$ is singular and is not invertible. So, we consider the β-normalized Laplacian, denoted as $\mathcal{L}_\beta = \beta I + k\mathcal{L}$, for some scalars $\beta > 0, k > 0$. The spectrum of the β-normalized Laplacian is computed to be

$$\beta < \beta + k\eta_2 \leq \beta + k\eta_3 \leq \cdots \leq \beta + k\eta_n \leq \beta + 2k.$$

Since ϕ_i are orthonormal eigenvectors of $\mathcal{L}$,

$$\mathcal{L}_\beta = \sum_{i=1}^{n} \beta\phi_i\phi_i^T + k\sum_{i=1}^{n} \eta_i\phi_i\phi_i^T = \sum_{i=1}^{n}(\beta + k\eta_i)\phi_i\phi_i^T. \tag{5.10}$$

$\mathcal{L}_\beta$ has the same eigenvectors as $\mathcal{L}$ but with eigenvalues $\beta + k\eta_i$. Since the eigenvalues of $\mathcal{L}_\beta$ are non-zero, it is invertible. We define $\mathcal{R}_\beta$ to be the symmetric matrix satisfying

$$\mathcal{L}_\beta\mathcal{R}_\beta = I.$$

Since

$$\left(\sum_{i=1}^{n}(\beta + k\eta_i)\phi_i\phi_i^T\right)\left(\sum_{i=1}^{n}\frac{1}{(\beta + k\eta_i)}\phi_i\phi_i^T\right) = \sum_{i=1}^{n}(\beta + k\eta_i)\frac{1}{(\beta + k\eta_i)}\phi_i\phi_i^T\phi_i\phi_i^T = I,$$

$$\mathcal{R}_\beta = \sum_{i=1}^{n}\frac{1}{(\beta + k\eta_i)}\phi_i\phi_i^T. \tag{5.11}$$

$\mathcal{R}_\beta$ is referred to as the β normalized Green's function.

Example 5.2.3. Consider a weighted, undirected, connected graph as shown in Figure 5.1.

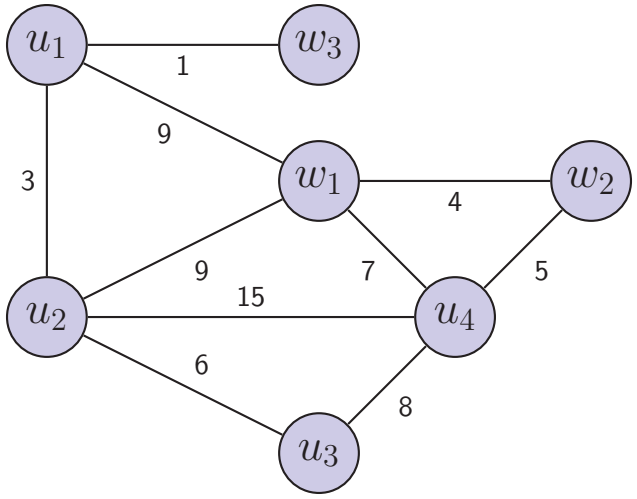

Figure 5.1: Weighted graph.

The matrices for this graph are

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 & 9 & 0 & 1 \\ 3 & 0 & 6 & 15 & 9 & 0 & 0 \\ 0 & 6 & 0 & 8 & 0 & 0 & 0 \\ 0 & 15 & 8 & 0 & 7 & 5 & 0 \\ 9 & 9 & 0 & 7 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 & 4 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, D = \begin{bmatrix} 13 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 33 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 35 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 29 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$P = D^{-1}A = \begin{bmatrix} 0 & 0.2308 & 0 & 0 & 0.6923 & 0 & 0.0769 \\ 0.0909 & 0 & 0.1818 & 0.4545 & 0.2727 & 0 & 0 \\ 0 & 0.4286 & 0 & 0.5714 & 0 & 0 & 0 \\ 0 & 0.4286 & 0.2286 & 0 & 0.2000 & 0.1429 & 0 \\ 0.3103 & 0.3103 & 0 & 0.2414 & 0 & 0.1379 & 0 \\ 0 & 0 & 0 & 0.5556 & 0.4444 & 0 & 0 \\ 1.0000 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

$$L = D - A = \begin{bmatrix} 13 & -3 & 0 & 0 & -9 & 0 & -1 \\ -3 & 33 & -6 & -15 & -9 & 0 & 0 \\ 0 & -6 & 14 & -8 & 0 & 0 & 0 \\ 0 & -15 & -8 & 35 & -7 & -5 & 0 \\ -9 & -9 & 0 & -7 & 29 & -4 & 0 \\ 0 & 0 & 0 & -5 & -4 & 9 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = \begin{bmatrix} 1 & -0.14 & 0 & 0 & -0.46 & 0 & -0.27 \\ -0.14 & 1 & -0.27 & -0.44 & -0.29 & 0 & 0 \\ 0 & -0.27 & 1 & -0.36 & 0 & 0 & 0 \\ 0 & -0.44 & -0.36 & 1 & -0.21 & -0.28 & 0 \\ -0.46 & -0.29 & 0 & -0.2 & 1 & -0.2476 & 0 \\ 0 & 0 & 0 & -0.28 & -0.24 & 1 & 0 \\ -0.27 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The spectra of the matrices are

$$spectra(A) = \begin{bmatrix} 26.34 & 6.34 & .91 & -0.1 & -5.04 & -12.16 & -16.29 \end{bmatrix},$$

$$spectra(L) = \begin{bmatrix} 0 & 1.08 & 9.55 & 10.56 & 24.04 & 39.15 & 49.61 \end{bmatrix},$$

$$spectra(\mathcal{L}) = \begin{bmatrix} 0 & 0.58 & 0.88 & 1.05 & 1.33 & 1.56 & 1.6 \end{bmatrix},$$

$$spectra(P) = \begin{bmatrix} 1 & 0.42 & 0.12 & -0.05 & -0.33 & -0.56 & -0.60 \end{bmatrix}.$$

□

5.3 Laws of electricity

Electrical circuits are governed by the physical laws of electricity. Here we look at two basic laws.

5.3.1 Ohm's law

The electric current I through a conductor is directly proportional to the voltage V applied to it. The ratio of voltage to current is the resistance R. We express this relationship as

$$V = IR. \tag{5.12}$$

5.3.2 Kirchoff's current law

The physical laws of conservation of charge tells us that electrical charge is neither created nor destroyed. By this principle of conservation, at any node in an electric circuit, the amount of charge entering the node equals the amount of charge leaving the node. Recalling that current is a signed quantity (positive or negative), as the current flows into or out of a node, we state this principle as

$$\sum_{i=1}^{k} I_i = 0, \tag{5.13}$$

where k is the total number of branches connecting into or out of a node.

5.4 Voltage potential

Consider a random walk along the line of positive integers, $N = 1, 2, \cdots, n$, starting at any vertex x. Let us look at the hitting probability function $k(x)$ of a

random walk starting at integer x and reaching n before reaching 1.

$$k(x) = \mathbb{P}(X_i = n | X_1 = x, X_l \neq 1 \text{ for } l < i).$$

The integer line ends at 1 and n. We have the following

(a) $k(1) = 0$,

(b) $k(n) = 1$,

(c) $k(x) = \frac{1}{2}k(x-1) + \frac{1}{2}k(x+1)$ for $x = 2, 3, \cdots n-1$.

Results (a) and (b) are based on our definition of $k(x)$. Result (c) is a direct application of probability theory.

Suppose we connect a series of resistors at each integer node x and suppose we maintain a unit potential at the ends by connecting a battery of potential 1 volt between the end nodes. Now, we view the integer line as an electrical circuit. By the laws of electricity, voltage $v(x)$ is established across each node x with $v(1) = 0$ and $v(n) = 1$. This satisfies conditions (a) and (b). We now show that $v(x)$ satisfies condition (c) as well.

By Kirchoff's law, the net current through the circuit is 0 and the current flowing into each node equals the current flowing out. By Ohm's law, if nodes x and y are connected by a resistor of r_{xy} ohms, then the voltage difference across the nodes is given by

$$v_x - v_y = i_{xy} \cdot r_{xy}$$

Thus, if there are n resistors, all of equal magnitude r ohms, the voltage potential across any two nodes $x, x+1$ is

$$v(x-1) - v(x+1) = [v(x-1) - v(x)] - [v(x) - v(x+1)].$$

The net current at each node is 0. Using Ohm's law we write the net current at node x as

$$0 = \frac{v(x-1) - v(x)}{r} + \frac{v(x+1) - v(x)}{r}$$

Multiplying through by r and collecting $v(x)$, we get,

$$v(x) = \frac{1}{2}v(x-1) + \frac{1}{2}v(x+1)$$

Thus $v(x)$ also satisfies property (c).

This simple example helps us see that the probability function $k(x)$ in a random walk behaves similar to the voltage potential $v(x)$ in an electrical network. To show that they are indeed the same, we refer to harmonic functions.

5.4.1 Harmonic functions on a graph

Let $\mathcal{G} = (V, E, W)$ be a connected, weighted graph. We split the vertices of the graph into two sets, U the set of internal vertices and W the set of external vertices, such that the two sets partition V i.e., $U \cup W = V$ and $U \cap W = \emptyset$. The set of external vertices W are considered the boundary points of the set V. Two vertices u and v that share an edge are neighbors, denoted by $u \sim v$.

Definition 5.4.1. Harmonic function.

A function $h : V \to \mathbb{R}$ is *harmonic* on U if

$$h(u) = \frac{1}{d(u)} \sum_{v \sim u} w_{uv} h(v), \tag{5.14}$$

for any $u \in U$, where $d(u) = \sum_{v \sim u} w_{uv}$.

Lemma 5.4.2. *Suppose $\mathcal{G}$ is a connected, weighted graph, $\{U, W\}$ is a partition of vertices of $\mathcal{G}$ as described above, and $h(w)$ is given for all $w \in W$. Then, there exists a unique harmonic function on U.*

Proof. Our proof takes advantage of basic graph theory and matrix theory theorems. As before, D is the degree matrix, A is the weighted adjacency matrix, and w_{uv} is the weight of the edge $(u, v) \in E$. From (5.14), the value of the harmonic function h at each vertex $u \in U$ is

$$h(u) = \frac{1}{d(u)} \sum_{v \sim u} w_{uv} h(v). \tag{5.15}$$

Equivalently,

$$d(u)h(u) - \sum_{v \sim u} w_{uv} h(v) = 0. \tag{5.16}$$

We partition the sum for the external vertices and express this equation as

$$\sum_{(v \sim u) \cap W} w_{uv} h(v) = d(u)h(u) - \sum_{(v \sim u) \cap U} w_{uv} h(v). \tag{5.17}$$

The sum on the left hand side of the above equation (5.17) is based on the given fixed values for the external vertices $w \in W$. Since there are $k = |U|$ linear equations for k unknowns, we represent this system using matrix notation as follows:

$$[D - A]_{k \times k} \mathcal{H} = \bar{W}, \tag{5.18}$$

where $[D - A]_{k \times k}$ represents the $k \times k$ submatrix of the Laplacian of $\mathcal{G}$ indexed by vertices $u \in U$, $\mathcal{H} = [h(u)], u \in U$ and $\bar{W} = \left[\sum_{(v \sim u) \cap W} w_{uv} h(v) \right]$. The left hand side of (5.18) is derived from the Laplacian L of $\mathcal{G}$. If we consider the subgraph $\mathcal{G}_U$ of $\mathcal{G}$ by restricting $\mathcal{G}$ to the set of vertices in U, we define the Laplacian of $\mathcal{G}_U$ as

$$L_U = D_{G_U} - A_U.$$

Let $\bar{D} = [D]_U - D_{G_U}$. We rewrite the matrix notation in (5.18) using the restricted Laplacian as

$$[L_U + \bar{D}]\mathcal{H} = \bar{W}. \tag{5.19}$$

We have seen before that the Laplacian of a graph is positive semi-definite. So L_U is positive semi-definite. Since $\bar{D}$ is diagonal with non-negative entries, it is also positive semi-definite. Thus $[L_U + \bar{D}]$ is positive semi-definite.

Claim 5.4.3. $[L_U + \bar{D}]$ is invertible.

Proof. Suppose $[L_U + \bar{D}]$ is not invertible. Then there exists a non-zero vector $x \in \mathbb{R}^k$ such that $[L_U + \bar{D}]x = 0$. And $x^T[L_U + \bar{D}]x = 0$. Since L_U and $\bar{D}$ are *positive semi-definite*, we must have $0 = x^T L_U x = x^T \bar{D} x$. So $x^T \bar{D} x = \sum_i \bar{d}_i x_i^2 = 0$, and $\bar{d}_i x_i^2 = 0$ for every i. Suppose $x_i \neq 0$, then $\bar{d}_i = 0$. This means that vertex v_i does not have a neighbor in W. But $\mathcal{G}$ is connected. So some vertex v_j has a neighbor in W and has $d_j \neq 0$. Hence $x_j = 0$. By Lemma 5.1.11, $x_i = x_j$. But $x_i \neq 0$ for every i and x is the zero vector, a contradiction. Hence $[L_U + \bar{D}]$ is invertible.

□

Matrices have unique inverses. Hence we conclude that there exists a solution to our matrix equation (5.19).

Corollary 5.4.4. *Harmonic functions with same boundary values are unique.*

Example 5.4.5. Consider the weighted graph as shown in Figure 5.1. The vertices of $\mathcal{G}$ are partitioned as $W = \{w_1, w_2, w_3\}$, the set of external vertices and $U = \{u_1, u_2, u_3, u_4\}$, the set of internal vertices. Suppose $h(W) = \begin{bmatrix} 4 & 7 & 5 \end{bmatrix}^T$. We compute the value of h at each internal vertex u_i using (5.15).

$h(u_1) = f(h(w_1), h(w_3), h(u_2)) = \frac{1}{13}(36 + 5 + 3h(u_2))$. We set up similar

equations for each vertex u_i and express this in matrix form as

$$\begin{bmatrix} 13 & -3 & 0 & 0 \\ -3 & 11 & -2 & -5 \\ 0 & -3 & 7 & -4 \\ 0 & -15 & -8 & 35 \end{bmatrix} \begin{bmatrix} h(u_1) \\ h(u_2) \\ h(u_3) \\ h(u_4) \end{bmatrix} = \begin{bmatrix} 41 \\ 12 \\ 0 \\ 63 \end{bmatrix}$$

yields

$$\mathcal{H} = \begin{bmatrix} 4.1927 & 4.5015 & 4.6702 & 4.7967 \end{bmatrix}^T.$$

We compute the same using the Laplacian in (5.19).

$$[L_U + \bar{D}] = \begin{bmatrix} 3 & -3 & 0 & 0 \\ -3 & 24 & -6 & -15 \\ 0 & -6 & 14 & -8 \\ 0 & -15 & -8 & 23 \end{bmatrix} + \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 \end{bmatrix} = \begin{bmatrix} 13 & -3 & 0 & 0 \\ -3 & 33 & -6 & -15 \\ 0 & -3 & 7 & -4 \\ 0 & -15 & -8 & 35 \end{bmatrix}$$

And

$$\bar{W} = \left[\sum_{u \sim v \cap W} w_{uv} h(v) \right] = \begin{bmatrix} 41 \\ 36 \\ 0 \\ 63 \end{bmatrix}$$

Hence, we get the same result for $\mathcal{H} = [h(u_i)] = [L_U + \bar{D}]^{-1} \bar{W}$.

□

5.5 Random walks and electrical networks

For any graph $\mathcal{G}(V, E, W)$, we view $\mathcal{G}$ as an electrical network by considering each edge as a resistor. For any edge (x, y), let r_{xy} be the resistance of that edge, i_{xy} the current, v_{xy} the voltage difference across the vertices, and $c_{xy} = \frac{1}{r_{xy}}$ the

conductance of that edge. For any vertex x, the conductance at that vertex c_x, is the sum of conductance of all the edges at x.

$$c_x = \sum_{y \sim x} c_{xy}.$$

First, we establish the voltage potential as a hitting probability by showing that both are harmonic functions on the same set with the same boundary conditions. Fix any two specific nodes a and b, connect a battery between them to establish a potential of 1 volt. So $v_a = 1$ and $v_b = 0$. The probability p_{xy} of taking the edge (x, y) from vertex x is $p_{xy} = \dfrac{c_{xy}}{c_x}$. By Ohm's law the current through any edge (x, y) is

$$i_{xy} = (v_x - v_y)c_{xy}.$$

By Kirchoff's current law, the total current through any vertex x other than a or b is 0. So, for any edge x with nodes not in $\{a, b\}$

$$0 = \sum_{x \sim y} i_{xy} = \sum_{y \sim x} (v_x - v_y)c_{xy}.$$

Alternatively,

$$\sum_{y \sim x} v_x c_{xy} = \sum_{y \sim x} v_y c_{xy}$$

or

$$v_x c_x = \sum_{y \sim x} v_y c_{xy},$$

yielding

$$v_x = \sum_{y \sim x} v_y \frac{c_{xy}}{c_x} = \sum_{y \sim x} v_y p_{xy}.$$

Since $\sum_y p_{xy} = 1$, we have stated v_x as a weighted sum using the probability of traversing each edge. From our setup, $v_a = 1$ and $v_b = 0$. These three essential factors allow us to conclude that $v(x)$ is a harmonic function at all points x other than a and b.

Now, let us take a random walk on this same weighted graph. We stipulate that the weight of each edge is the same as the conductance of the edge, i.e., $w_{xy} = c_{xy}$. Consider the hitting probability $k(x)$ that a random walk starting at x reaches a before b. Obviously $k(a) = 1$ and $k(b) = 0$. So, these hitting probabilities satisfy the same boundary conditions as the harmonic voltage function $v(x)$. Also, $k(x) = \sum_{y \sim x} p_{xy} k(y)$. As we have seen in Chapter 2, if we start a random walk at node x, our next step is towards any of its neighbors y, with the probability p_{xy}. Then $k(x)$ is also a weighted sum of the probability of taking an edge. So $k(x)$ is also a harmonic function defined on all vertices except a and b.

By Corollary 5.4.4, $v(x)$ and $k(x)$ are the same function. To sum up, we have the following interpretation of voltage. When a unit voltage is applied between nodes a and b, making $v_a = 1$ and $v_b = 0$, the voltage v_x at any point x represents the hitting probability that a walker starting from x returns to a before reaching b. Although we have chosen a unit voltage as the potential across a and b, it can be any arbitrary voltage v that the electrical circuit can support.

Theorem 5.5.1. *Hitting time between two vertices is the same as voltage potential across the vertices.*

Proof. Suppose $\mathcal{G}$ is an electrical network with $|V|$ nodes and $|E|$ edges with conductance w_{uv} on edge $(u, v) \in E$. Let v_u be the voltage at node u. Then $v_{uv} = v_u - v_v$ denotes the voltage potential across two vertices u and v. Suppose $d(u)$ units current is introduced at each vertex $u \in V - \{v\}$ and all the $vol(\mathcal{G})$ units of current are extracted at vertex v as shown in Figure 5.5.1. By Kirchoff's current law, the net current at node u is 0.

$$0 = i_1 - i_2 + d(u) + i_3 - i_4,$$

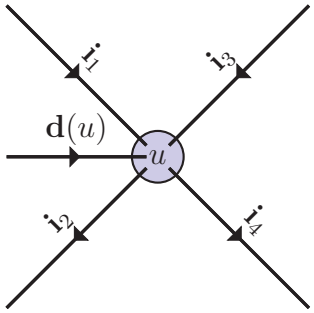

Figure 5.2: Current at node u

Rewriting for $d(u)$ yields

$$d(u) = i_2 + i_4 - i_1 - i_3.$$

For every node $u \in V - \{v\}$,

$$d(u) = \sum_{w \sim u} i_{uw}.$$

Using Ohm's law and $\sum_{w \sim u} w_{wu} = d(u)$, we have

$$\sum_{w \sim u} i_{uw} = \sum_{w \sim u} (w_{uw} v_{uv} - w_{wv} v_{wv}) = d(u) v_{uv} - \sum_{w \sim u} w_{uw} v_{wv}.$$

We rewrite the above equation as

$$v_{uv} = 1 + \frac{1}{d(u)} \sum_{w \sim u} w_{uw} v_{wv}. \tag{5.20}$$

Suppose we consider the electrical network as a graph $\mathcal{G}(V, E, W)$, with each node of the circuit as a vertex on the graph, the connections between each node as an edge and the conductance of the edge as its weight. We now apply (2.10) to express hitting time between any two vertices $u, v \in V$ as

$$\begin{aligned} H_{uv} = \sum_{w \sim u} (1 + H(w, v)) p_{uw} &= \sum_{w \sim u} (1 + H(w, v)) \frac{w_{uw}}{d(u)} \\ &= 1 + \frac{1}{d(u)} \sum_{w \sim u} w_{uw} H_{wv}. \end{aligned} \tag{5.21}$$

The above two equations are identical. Hence both hitting time and voltage potential are solutions to both the equations. From (5.15), we see that the summand on both the equations represent an harmonic function on the same graph. By uniqueness of harmonic functions in Corollary 5.4.4, these equations have the same solutions and we conclude $H_{uv} = v_{uv}$.

□

Example 5.5.2. The hitting time for the graph shown in Figure 5.1 is

$$H = \begin{bmatrix} 0 & 0.5571 & 1.7046 & 0.6913 & 0.3339 & 2.3007 & 19.8507 \\ 1.7563 & 0 & 1.3119 & 0.4218 & 0.6391 & 2.2867 & 21.6070 \\ 1.9866 & 0.3946 & 0 & 0.3300 & 0.8146 & 2.3137 & 21.8373 \\ 1.8981 & 0.4294 & 1.2549 & 0 & 0.6850 & 2.0727 & 21.7488 \\ 1.4046 & 0.5106 & 1.6034 & 0.5489 & 0 & 2.0732 & 21.2553 \\ 1.8280 & 0.6147 & 1.5590 & 0.3932 & 0.5298 & 0 & 21.6787 \\ 0.1492 & 0.7063 & 1.8539 & 0.8405 & 0.4832 & 2.4500 & 0 \end{bmatrix}$$

□

CHAPTER 6

PERSONALIZED PAGERANK AND HITTING TIME

In Chapter 3, we saw the exposition of PageRank and its use by Google to rank the importance of billions of webpages. Although the concept of PageRank was originally derived for webgraph, it can be applied to any connected graph. In this chapter, following Chung's paper [FC10], we modify the original PageRank equation to define a personalized PageRank. This allows us to define a generalized version of hitting time and commute time and express hitting time in terms of personalized PageRank.

We consider a connected, weighted graph $\mathcal{G}(V, E, W)$, where V is the set of vertices, E the set of edges, and W the weights of the edges of $\mathcal{G}$. We refer to the weight of an edge $(v_i, v_j) \in E$ by w_{ij}. The weights are positive for every edge. We choose an arbitrary, but fixed orientation of $\mathcal{G}$. As before, we use the notation $\sim$ to denote neighboring vertices; that is $v_i \sim v_j$ means v_j is a neighbor of v_i.

6.1 Personalized PageRank

In Proposition 4.4.5, we saw the definition of Google's PageRank π as the left eigenvector of the Google matrix G with eigenvalue 1.

$$\pi^T = \pi^T G,$$

which we defined in (4.6) to be

$$\pi^T = \pi^T \alpha S + \frac{(1-\alpha)}{n}\mathbf{1}^T, \tag{6.1}$$

where S is the stochastic probability transition matrix, α is the jumping constant and n is the number of vertices in the graph.

6.1.1 Lazy random walk

For a random walk on a connected graph, we saw that the sequence of distributions eventually converges to the stationary distribution π, where $\pi_j = \frac{d(v_j)}{2m}$. For a weighted graph $\mathcal{G}(V, E, W)$, we replace $2m$ by $vol(\mathcal{G})$ where $vol(\mathcal{G}) = \sum_{v_i \in V} d(v_i)$ and $d(v_i) = \sum_{v_j \sim v_i} w_{ij}$.

However, it is not necessary that if we start with any arbitrary distribution, the distribution converges to π. Indeed, by taking the graph to be a single edge, and starting at one of the vertices, the distribution is [1 0] and [0 1] alternately. The unique stationary distribution $[\frac{1}{2}\ \frac{1}{2}]$ is never reached.

The key problem here is that the walk is periodic; at even steps it is at one vertex and at odd steps at the other. The primitivity adjustment we saw in Chapter 3 was Google's fix to this periodicity issue. Here we look at an alternate method, a modified version of the original walk: the lazy random walk.

In a lazy random walk, we make a move only every other turn. We toss a fair coin. If the coin lands on head, we stay at the same vertex(hence lazy). If the coin lands on tail, we move at random to any neighbor. At any time

- we take a step of the original random walk with probability $\frac{1}{2}$, or
- we stay at the current vertex with probability $\frac{1}{2}$.

Definition 6.1.1. Probability transition matrix of a lazy random walk. The probability of transition at every step of a lazy random walk is $\frac{1}{2} + \frac{1}{2}p$, where p is the transition probability of the same step in the standard random walk. Hence the *probability transition matrix Z of a lazy random walk* equals $\frac{I+P}{2}$.

Theorem 6.1.2. *Let Z be the probability transition matrix of a lazy random walk. Then Z is stochastic and its eigenvalues lie between 0 and 1.*

Proof. Let P be the probability transition matrix of a standard random walk on $\mathcal{G}(V, E, W)$. Then

$$Z = \frac{I + P}{2}.$$

Since I and P are both stochastic, Z is also stochastic.

From Theorem 2.5.7, we know that the eigenvalues λ_i of P lie between $[-1, 1]$. Furthermore, the eigenvalue 1 is simple. So, $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1$. Using this, we compute the eigenvalues of Z. Let ω_i be the eigenvector of P with eigenvalue λ_i. Then

$$Z\omega_i = \frac{(I + P)}{2}\omega_i = \frac{1}{2}(\omega_i + \lambda_i\omega_i) = \frac{1}{2}(1 + \lambda_i)\omega_i.$$

So $\frac{1}{2}(1 + \lambda_i)$ are the eigenvalues of Z. Since $-1 \leq \lambda_i \leq 1$, we have

$$0 \leq \frac{1 + \lambda_i}{2} \leq 1.$$

The set, $\{(\frac{1}{2}(1 + \lambda_i), \omega_i)\}$, is the eigenpairs of Z. Since the eigenvalue 1 of P is simple, 1 is also a simple eigenvalue of Z.

□

6.1.2 Personalized PageRank

Definition 6.1.3. Personalized PageRank.

The *personalized PageRank vector* $pr_\alpha(s)$ is defined to be the unique solution to the equation

$$pr_\alpha(s)^T = \alpha pr_\alpha(s)^T Z + (1 - \alpha)s^T, \tag{6.2}$$

where Z is the probability transition matrix of a lazy random walk, $0 < \alpha < 1$ is the teleportation constant and s is the seed vector with $\sum_i s_i = 1$.

Proposition 6.1.4. Equation (6.2) has a unique solution.

Proof. We write (6.2) as

$$(1-\alpha)s^T = pr_\alpha(s)^T[I - \alpha Z]. \tag{6.3}$$

Since the eigenvalues of Z lie between 0 and 1 and $0 < \alpha < 1$, the matrix $[I - \alpha Z]$ does not have a zero eigenvalue. Hence $[I - \alpha Z]$ has a unique inverse.

□

Theorem 6.1.5. *$pr_\alpha(s)$ is the stationary distribution of the stochastic matrix $M = \alpha Z + (1-\alpha)\mathbf{1}s^T$.*

Proof. First note that M is the convex combination of two stochastic matrices: Z and $\mathbf{1}s^T$; hence M is stochastic. Since P is irreducible, $Z = (I+P)/2$ is primitive [RAH85] pages 100-120. Consequently, by Theorem B.0.21, M is primitive. Hence by the fundamental stability theorem of Markov chains, Theorem 2.5.8, M has a stationary distribution, and the result follows.

□

Proposition 6.1.6. Suppose π is the PageRank of the standard random walk and pr is the PageRank of the lazy random walk, then

$$\pi\left(\frac{\alpha}{2-\alpha}, s\right) = pr_\alpha(s),$$

where $0 < \alpha < 1$ and $\sum_i s_i = 1$.

Proof. For a connected graph $\mathcal{G}$, the probability transition matrix P is already stochastic. Using the definition of Google's PageRank in (6.1), we write $\pi\left(\frac{\alpha}{2-\alpha}, s\right)$ as

$$\begin{aligned}\pi^T &= \frac{\alpha}{2-\alpha}\pi^T P + \left(1 - \frac{\alpha}{2-\alpha}\right)s^T \\ &= \frac{\alpha}{2-\alpha}\pi^T P + \frac{2(1-\alpha)}{2-\alpha}s^T.\end{aligned}$$

Multiplying by $(2-\alpha)$ and collecting terms yields

$$\begin{aligned}\pi^T(2-\alpha) &= \alpha\pi^T P + 2(1-\alpha)s^T \\ \pi^T &= \alpha\pi^T\frac{I+P}{2} + (1-\alpha)s^T \\ &= \alpha\pi^T Z + (1-\alpha)s^T.\end{aligned}$$

By Proposition 6.1.4, we know that the personalized PageRank equation has a unique solution. And we get the desired result.

□

Proposition 6.1.7. The Green's function $\mathcal{R}_\beta$ is a symmetric form of the personalized PageRank.

$$\frac{pr_\alpha(s)^T}{\beta} = s^T D^{-1/2}\mathcal{R}_\beta D^{1/2},$$

where $\beta = \dfrac{2(1-\alpha)}{2-\alpha}$.

Proof. From (6.2) we have

$$pr_\alpha(s)^T = \alpha pr_\alpha(s)^T Z + (1-\alpha)s^T,$$

where $0 < \alpha < 1$, s is the seed vector and $Z = (I+P)/2$.

Let $\beta = \dfrac{2-2\alpha}{2-\alpha}$. Since $0 < \alpha < 1$, $0 < \beta < 1$. Expressing α in terms of β, we get $\alpha = \dfrac{2(1-\beta)}{2-\beta}$. We now rewrite $pr_\alpha(s)$ as

$$pr_\alpha(s)^T = \frac{2(1-\beta)}{2-\beta} pr_\alpha(s)^T Z + \frac{\beta}{2-\beta} s^T.$$

Applying the definition of Z, we get

$$pr_\alpha(s)^T = \frac{1-\beta}{2-\beta} pr_\alpha(s)^T (I+P) + \frac{\beta}{2-\beta} s^T.$$

Collecting $pr_\alpha(s)^T$ and using $D^{-1/2}\mathcal{L}D^{1/2} = I - P$, gives

$$\frac{pr_\alpha(s)^T}{\beta} D^{-1/2}[(1-\beta)\mathcal{L} + \beta I]D^{1/2} = s^T.$$

Using the inverse function of $\mathcal{L}_\beta$, the beta normalized Green's function from (5.11) yields

$$\frac{pr_\alpha(s)^T}{\beta} = s^T D^{-1/2} \mathcal{R}_\beta D^{1/2}. \tag{6.4}$$

□

Example 6.1.8. For a connected graph as shown in Figure 6.1 below, we compute Google's PageRank and personalized PageRank.

Set $\alpha = 0.85$ and the seed vector

$$s = \begin{bmatrix} \frac{1}{12} & \frac{1}{12} & \frac{1}{6} & \frac{1}{3} & \frac{1}{4} & \frac{1}{12} \end{bmatrix}.$$

Method 1. Google's power method: Since the graph is connected, the probability

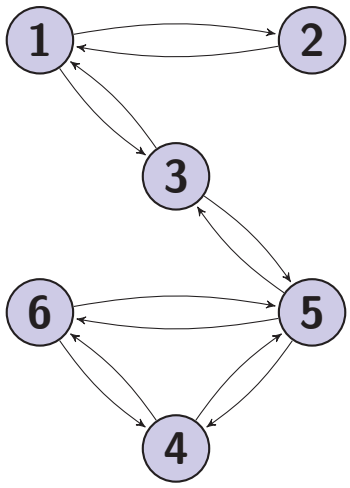

Figure 6.1: Connected graph with six pages.

transition matrix P is stochastic and the Google matrix G is primitive.

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix},$$

$$G = \begin{bmatrix} 0.0217 & 0.3913 & 0.4565 & 0.0435 & 0.0652 & 0.0217 \\ 0.7609 & 0.0217 & 0.0870 & 0.0435 & 0.0652 & 0.0217 \\ 0.3913 & 0.0217 & 0.0870 & 0.0435 & 0.4348 & 0.0217 \\ 0.0217 & 0.0217 & 0.0870 & 0.0435 & 0.4348 & 0.3913 \\ 0.0217 & 0.0217 & 0.3333 & 0.2899 & 0.0652 & 0.2681 \\ 0.0217 & 0.0217 & 0.0870 & 0.4130 & 0.4348 & 0.0217 \end{bmatrix}.$$

Google's PageRank, the stationary distribution vector computed using the power

method is

$$\pi^T = \begin{bmatrix} 0.1576 & 0.0800 & 0.2076 & 0.1587 & 0.2534 & 0.1428 \end{bmatrix}.$$

Method 2. Symmetric Green's function method: The adjacency matrix is symmetric. Hence the Laplacian $L = D - A$ is also symmetric.

$$L = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}.$$

The $\beta-$Laplacian $\mathcal{L}_\beta$ and its inverse the Green's function $\mathcal{R}_\beta$ are

$$\mathcal{L}_\beta = \begin{bmatrix} 1 & -0.5226 & -0.3696 & 0 & 0 & 0 \\ -0.5226 & 1 & 0 & 0 & 0 & 0 \\ -0.3696 & 0 & 1 & 0 & -0.3017 & 0 \\ 0 & 0 & 0 & 1-0.3017 & -0.3696 & \\ 0 & 0 & -0.3017 & -0.3017 & 1-0.3017 & \\ 0 & 0 & 0 & -0.3696 & -0.3017 & 1 \end{bmatrix},$$

$$\mathcal{R}_\beta = \begin{bmatrix} 1.7537 & 0.9166 & 0.7433 & 0.1510 & 0.3154 & 0.1510 \\ 0.9166 & 1.4790 & 0.3885 & 0.0789 & 0.1648 & 0.0789 \\ 0.7433 & 0.3885 & 1.4619 & 0.2969 & 0.6203 & 0.2969 \\ 0.1510 & 0.0789 & 0.2969 & 1.5406 & 0.7990 & 0.8105 \\ 0.3154 & 0.1648 & 0.6203 & 0.7990 & 1.6694 & 0.7990 \\ 0.1510 & 0.0789 & 0.2969 & 0.8105 & 0.7990 & 1.5406 \end{bmatrix}.$$

The personalized PageRank using the symmetric form of the $\beta-$Green's function in (6.4) is

$$pr_\alpha(s)^T = \begin{bmatrix} 0.1576 & 0.0800 & 0.2076 & 0.1587 & 0.2534 & 0.1428 \end{bmatrix}.$$

Since the graph is connected, $\pi^T = pr_\alpha(s)^T$.

6.2 Personalized PageRank and hitting time

Let $\mathcal{N}(\mathcal{G})$ be the electrical network of a graph $\mathcal{G}(V, E, W)$ having a node for every vertex $v \in V$ and conductance of w_{uv} on every edge $(u, v) \in E$. Here, we use the theory of electrical networks on graphs from Chapter 4.

Definition 6.2.1. Effective resistance.
In any electrical circuit, the *effective resistance* between any two nodes u and v is defined as the voltage that develops between them when a unit current is maintained through them (i.e., enters one and leaves the other). In the quantitative sense, the resistance between two points is defined to be the voltage difference that is required to take a unit current across the defined two nodes.
Formally, effective resistance R_{uv} between nodes u and v is the voltage potential difference induced between them when a current of one ampere is introduced at u and extracted at v. Suppose $f : V \to \mathbb{R}$ is a voltage potential function. Then $R_{uv} = f_u - f_v$.

Theorem 6.2.2. *For graph $\mathcal{G}(V, E, W)$ with vertices $u, v \in V$*

$$C_{uv} = vol(\mathcal{G}) R_{uv},$$

where C_{uv} is the commute time, R_{uv} is the effective resistance between the vertices u, v, and $vol(\mathcal{G}) = \sum_{v \in V} d(v)$.

Proof. We follow the proof in Chandra et al in [AC96]. Suppose $\mathcal{N}(\mathcal{G})$ has w_e ohm resistance at each edge $e \in E$. Let $d(x)$ units current be introduced at each vertex $x \in V$ and all the $vol(\mathcal{G})$ units of current be extracted at vertex v.

In Theorem 5.5.1, we established that the voltage potential f between two vertices is the same as the hitting time between the same two vertices: $f_u - f_v = H_{uv}$.

In an electrical network, current flows in the direction of voltage gradient. So, if we reverse the flow of current and induce $vol(\mathcal{G})$ unit of current at u and remove $d(x)$ at every vertex $x \in V - \{u\}$, we get the voltage at u with respect to v to be H_{vu}.

If we superimpose both these circuits, the current at each vertex $V - \{u, v\}$ cancels and we are left with the voltage difference between u and v when $vol(\mathcal{G})$ units of current are introduced at u and removed at v. By Ohm's law, $f_u - f_v = vol(\mathcal{G})R_{uv}$. By Theorem 5.5.1,

$$H_{uv} + H_{vu} = vol(\mathcal{G})R_{uv}. \tag{6.5}$$

Since $C_{uv} = H_{uv} + H_{vu}$, from (6.5), we have

$$R_{uv} = \frac{H_{uv} + H_{vu}}{vol(\mathcal{G})} = \frac{C_{uv}}{vol(\mathcal{G})}. \tag{6.6}$$

□

For graph $\mathcal{G}$, $B^T W$ gives the conductance between any two nodes $u, v \in V$, where B is the signed edge-vertex incidence matrix and W is the diagonal edge weight matrix. Using Kirchoff's law, we write the injected current function $i_V : V \to \mathbb{R}$ as

$$i_V^T = i_E^T B, \tag{6.7}$$

where $i_V(v)$ is the sum of all induced current on edges entering at vertex v. By Ohm's law, the induced current flow through any edge $(u, v) \in E$ with conductance w_{uv} is given by

$$i_{uv} = w_{uv}(f_u - f_v).$$

We write this in matrix form as

$$i_E^T = f^T B^T W. \tag{6.8}$$

Using the above two equations, we get

$$i_V^T = (f^T B^T W)B = f^T (B^T W B).$$

Applying the definition of the Laplacian from (5.4) gives us

$$i_V^T = f^T L = f^T D^{1/2} \mathcal{L} D^{1/2}.$$

Suppose we only consider potential functions such that $\sum_{v \in V} f = 0$; then $\mathbf{1}^T f = 0$. Notice that for ϕ_1, the eigenvector corresponding to eigenvalue 0 of the normalized Laplacian $\mathcal{L}$, $\phi_1 f = D^{1/2}\mathbf{1}^T f = 0$. Hence by definition of Green's function

$$f^T = i_V^T D^{-1/2} \mathcal{R} D^{-1/2}. \tag{6.9}$$

Consider a unit current injected at vertex u and extracted at vertex v. Suppose we represent the current vector by χ, where χ_u has 1 in the u^{th} entry and 0 elsewhere. Then $i_V = \chi_u - \chi_v$. By Ohm's law, the effective resistance between vertices u and v is

$$\begin{aligned} R(u, v) &= f^T(\chi_v - \chi_u) \\ &= (\chi_v - \chi_u)^T D^{-1/2} \mathcal{R} D^{-1/2} (\chi_v - \chi_u). \end{aligned} \tag{6.10}$$

Lemma 6.2.3. *For all vertices $u, v \in V$, the hitting time $H(u,v)$ is expressed in terms of the Greens's function $\mathcal{R}$ as follows:*

$$\frac{H(u,v)}{vol(\mathcal{G})} = (\chi_v - \chi_u)^T D^{-1/2} \mathcal{R} D^{-1/2} \chi_v. \tag{6.11}$$

Proof. Using the fundamental matrix $\mathbf{Q}$ in (2.17), we expressed hitting time as

$$H(u,v) = \frac{\mathbf{Q}_{vv} - \mathbf{Q}_{uv}}{\pi_v},$$

where $\mathbf{Q} = [I - P + \Pi]^{-1}$ and $\pi(v) = \dfrac{d(v)}{vol(\mathcal{G})}$. Hence

$$\begin{aligned} \frac{H(u,v)}{vol(\mathcal{G})} &= \frac{\mathbf{Q}_{vv}}{d(v)} - \frac{\mathbf{Q}_{uv}}{d(v)} \\ &= \left[\mathbf{Q}D^{-1}\right]_{vv} - \left[\mathbf{Q}D^{-1}\right]_{uv}. \end{aligned}$$

Using the unit function χ we rewrite the above equation as

$$\begin{aligned} \left[\mathbf{Q}D^{-1}\right]_{vv} - \left[\mathbf{Q}D^{-1}\right]_{uv} &= \chi_v^T \mathbf{Q} D^{-1} \chi_v - \chi_u^T \mathbf{Q} D^{-1} \chi_v \\ &= (\chi_v - \chi_u)^T \mathbf{Q} D^{-1} \chi_v. \end{aligned} \tag{6.12}$$

Using Lemma 5.2.2, we find

$$\mathbf{Q} = D^{-1/2}\left[D^{1/2}\mathbf{Q}D^{-1/2}\right]D^{1/2} = D^{-1/2}\mathcal{R}D^{1/2}.$$

Applying the above result back into (6.12), we get the desired result.

□

Definition 6.2.4. Generalized hitting time.

The *generalized hitting time* $h_\alpha(u,v)$ between any two vertices $u, v \in V$ with an additional parameter $0 < \alpha < 1$ is given by

$$h_\alpha(u,v) \stackrel{def}{=} \beta \frac{H(u,v)}{vol(\mathcal{G})} = \beta(\chi_v - \chi_u)^T D^{-1/2} \mathcal{R}_\beta D^{-1/2} \chi_v, \tag{6.13}$$

where $\beta = \dfrac{2(1-\alpha)}{2-\alpha}$.

Lemma 6.2.5. *The generalized hitting time in terms of the personalized PageRank* $pr_\alpha(\chi_v)$*, where* $0 < \alpha < 1$ *and* $\sum_i [\chi_v]_i = 1$ *is given by*

$$h_\alpha(u,v) = \frac{[pr_\alpha(\chi_v)^T](v)}{d(v)} - \frac{[pr_\alpha(\chi_v)^T](u)}{d(u)}. \tag{6.14}$$

Proof. Using (6.4), we have

$$pr_\alpha(\chi_v)^T = \beta \chi_v^T D^{-1/2} \mathcal{R}_\beta D^{1/2}.$$

So, the right hand side of the above equation becomes

$$\beta \frac{\chi_v^T D^{-1/2} \mathcal{R}_\beta D^{1/2}}{d(v)}[v] - \beta \frac{\chi_v^T D^{-1/2} \mathcal{R}_\beta D^{1/2}}{d(u)}[u] \tag{6.15}$$

Note that $\frac{1}{d(u)} = (D^{-1}\chi_u)[u]$ gives the u^{th} coordinate. Hence, we rewrite the above equation as

$$\beta \chi_v^T D^{-1/2} \mathcal{R}_\beta D^{1/2} (D^{-1}\chi_v) - \beta \chi_v^T D^{-1/2} \mathcal{R}_\beta D^{1/2} (D^{-1}\chi_u).$$

$D^{-1/2}\mathcal{R}_\beta D^{-1/2}$ is symmetric. And χ_x is a vector with 1 in the x^{th} coordinate. So $\chi_v^T [D^{-1/2}\mathcal{R}_\beta D^{-1/2}]\chi_u = [D^{-1/2}\mathcal{R}_\beta D^{-1/2}]_{uv} = \chi_u^T [D^{-1/2}\mathcal{R}_\beta D^{-1/2}]\chi_v$. So the right hand side of (6.14) becomes

$$\begin{aligned}
&\beta \chi_v^T D^{-1/2} \mathcal{R}_\beta D^{-1/2} \chi_v - \beta \chi_u^T D^{-1/2} \mathcal{R}_\beta D^{-1/2} \chi_v \\
&= \beta (\chi_v - \chi_u)^T D^{-1/2} \mathcal{R}_\beta D^{-1/2} \chi_v \\
&= h_\alpha(u,v).
\end{aligned}$$

□

Corollary 6.2.6. *The generalized effective resistance* $R_\alpha(u,v)$ *in terms of the Green's function* $\mathcal{R}_\beta$ *is given by*

$$R_\alpha(u,v) = h_\alpha(u,v) + h_\alpha(v,u).$$

Proof. Using (6.10), we define the generalized effective resistance $R_\alpha(u,v)$ to be

$$\begin{aligned} R_\alpha(u,v) &= \beta(\chi_v - \chi_u)^T D^{-1/2}\mathcal{R}_\beta D^{-1/2}(\chi_v - \chi_u) \\ &= \beta(\chi_v - \chi_u)^T D^{-1/2}\mathcal{R}_\beta D^{-1/2}\chi_v - \beta(\chi_v - \chi_u)^T D^{-1/2}\mathcal{R}_\beta D^{-1/2}\chi_u \\ &= \beta(\chi_v - \chi_u)^T D^{-1/2}\mathcal{R}_\beta D^{-1/2}\chi_v + \beta(\chi_u - \chi_v)^T D^{-1/2}\mathcal{R}_\beta D^{-1/2}\chi_u \\ &= h_\alpha(u,v) + h_\alpha(v,u). \end{aligned}$$

□

BIBLIOGRAPHY

[AC96] W. L. Ruzzo R. Smolensky P. Tiwari A. Chandra, P. Raghavan, *The electrical resistance of a graph captures its commute and cover times*, Computational Complexity **6** (1996), no. 4, 312–340.

[AL06] C. Meyer A. Langville, *Google's pagerank and beyond: The science of search engine rankings*, Princeton University Press, 2006.

[Chu94] F. Chung, *Spectral graph theory*, American Mathematical Society Publication, 1994.

[Chu00] ———, *Discrete green's functions*, Journal of Combinatorial Theory, Series A (2000), 191–214.

[cita] http://www.worldwidewebsize.com, October 14, 2013.

[citb] http://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4, October 14, 2013.

[citc] http://mygreatname.com/google-pagerank/google-pagerank-technology.html, October 14, 2013.

[FC10] W. Zhao F. Chung, *Pagerank and random walk on graphs*, Fete of Combinatorics and Computer Science **20** (2010), 43–62.

[Gre] http://www-history.mcs.st-and.ac.uk/Biographies/Green.html, October 14, 2013.

[Lov93] L. Lovăsz, *Random walks on graphs: A survey*, Combinatorics, Paul Erdos is Eighty **2** (1993), 1–46.

[Nor98] J. R. Norris, *Markov chains*, Cambridge University Press, 1998.

[PGD06] J. Laurie Snell Peter G. Doyle, *Random walks and electric networks*, The Mathematical Association of America, 2006.

[RAH85] C. R. Johnson R. A. Horn, *Matrix analysis*, Cambridge University Press, 1985.

[Ros12] S M. Ross, *Introduction to probability models, 9th edition*, Pearson, 2012.

[W.S94] W.Stewart, *Introduction to the numerical solution of markov chains*, Princeton University Press, 1994.

APPENDIX A

PROBABILITY THEORY

In probability theory, a probability space or a probability triple is a mathematical construct that models a real-world process (or "experiment") consisting of outcomes that occur randomly. A probability space is constructed with a specific kind of situation or experiment in mind. Each time a situation of the specific kind arises, the set of possible outcomes is the same and the probabilities are also the same.

A probability space consists of three parts:

- A sample space Ω, which is the set of all possible outcomes.
- A set of events $\mathcal{F}$, where each event is a set containing zero or more outcomes.
- A probability function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$, which assigns probabilities to the events.

Definition A.0.7. State-space.

Let I be a countable set. Each $i \in I$ is called a *state* and I the *state-space.*

Definition A.0.8. Random variable.

A *random variable* is a real-valued function defined on a set of possible outcomes: the sample space Ω. A random variable X with values in I is defined to be $X : \Omega \to I$.

Definition A.0.9. Distribution.

$\mu = (\mu_i : i \in I)$ is a *measure* on I if $0 \leq \mu_i \leq \infty$ for all $i \in I$. In addition, if

$\sum_{i \in I} \mu_i = 1$, then we call μ a *distribution*. Suppose we set

$$\mu_i = \mathbb{P}(X = i) = \mathbb{P}(\{\omega : X(\omega) = i\}).$$

Then μ defines a distribution, the distribution of X.

Example A.0.10. Let us consider a simple example in which we toss two fair coins. The four possible outcomes $\{HH, HT, TH, TT\}$ are the entries in the sample space Ω.

$E_1 = \{HH\}$ is the event that both the coins are heads.

$E_2 = \{TT\}$ is the event that both the coins are tails.

$E_3 = \{HH, HT, TH\}$ is the event that at least one coin is head.

$E_4 = \{HT, TH, TT\}$ is the event that at least one coin is tail.

Suppose we define a random variable X as the number of heads in a toss. Then, X takes values $0, 1, 2$. The distributions μ_i of X are given by the probabilities

$$\mu_1 = \mathbb{P}(X = 0) = \mathbb{P}(\{X(TT)\}) = 1/4,$$
$$\mu_2 = \mathbb{P}(X = 1) = \mathbb{P}(\{X(HT), X(TH)\}) = 1/2,$$
$$\mu_3 = \mathbb{P}(X = 2) = \mathbb{P}(\{X(HH)\}) = 1/4.$$

Definition A.0.11. Independent and conditionally dependent events. Two events A and B are said to be *independent* if the outcome of one does not affect the outcome of the other and vice versa. We then say $P(A|B) = P(A)$ and $P(B|A) = P(B)$. If the events are not independent, then they are considered dependent. For example, let us consider tossing two fair dices, red and blue. The outcome of the toss of one dice has no impact on the outcome of the toss of the other. So, the two events, namely "A, tossing red dice" and "B, tossing blue dice" are independent. If two events are independent then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Two events A and B are in*dependent conditionally* on C if once we know that given C has occurred, A and B are independent. Then, we say that $\mathbb{P}(A|B \cap C) = \mathbb{P}(A|C)$ and $\mathbb{P}(B|A \cap C) = \mathbb{P}(B|C)$. Suppose we look at the same example, tossing two fair dices, as above. But introduce a condition C, namely the number rolled on both the dices is even. Condition C affects the outcomes of both the red and blue dice. But not the outcomes of each other. Knowing that the the blue dice rolled number four does not impact the results of the number rolled by the red dice. In such a case, $\mathbb{P}(A|B, C) = \mathbb{P}(A|C)$.

APPENDIX B

MATRIX THEORY

Definition B.0.12. Non-negative matrix.

A matrix $A = [a_{ij}]$ is *non-negative* if its entries are such that $a_{ij} \geq 0,$ for all i, j.

Definition B.0.13. Positive matrix.

A matrix $A = [a_{ij}]$ is *positive* if its entries are positive, i.e., $a_{ij} > 0,$ for all i, j.

Definition B.0.14. Reducible and irreducible matrix. A $n \times n$ matrix A is *reducible* if

i. $n = 1$ and $A = [0]$; or

ii. $n \geq 2$, there is a permutation matrix U and some integer r with $1 \leq r \leq n-1$ such that

$$U^T A U = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where matrix B has size $r \times r$ and matrix D has size $(n-r) \times (n-r)$.

If A is not reducible, then it is termed *irreducible.*

Definition B.0.15. Symmetric matrix.

A matrix A is *symmetric* if $A = A^T$.

Definition B.0.16. Primitive matrix.

A non-negative matrix A is *primitive* if for some positive integer L, A^L is a positive matrix.

Theorem B.0.17. *Let A be a symmetric matrix. Then its inverse matrix B is also symmetric.*

Proof. Suppose matrix A is symmetric and invertible. Then $A = A^T$. Let matrix B be the inverse of A.

$$AB = BA = I \rightarrow B^T A^T = (AB)^T = (BA)^T = A^T B^T = I.$$

Hence the inverse matrix B is also symmetric.

Theorem B.0.18. *A $n \times n$ matrix A is irreducible if and only if $(I + A)^{n-1}$ is positive.*

Theorem B.0.19. *A non-negative matrix A is primitive if it is irreducible and has only one eigenvalue of maximum modulus.*

Theorem B.0.20. *An irreducible matrix A is primitive if A has at least one positive diagonal element.*

Corollary B.0.21. *If matrix A is primitive and matrix B is non-negative, then matrix $A + B$ is primitive.*

Proof. Since A is primitive, $A^k > 0$ for some positive integer k. By Binomial theorem

$$(A + B)^k = A^k + \binom{k-1}{1} A^{k-1} B + \cdots + B^k.$$

Since $B^n \geq 0$ for all n, $(A + B)^k > 0$.

Theorem B.0.22. *Perron-Frobenius[P-F] Theorem*
Suppose A is non-negative, irreducible, and primitive. Then

i. $\rho(A) > 0$.

ii. $\rho(A)$ *is an algebraically simple eigenvalue of* A.

iii. A has left and right Perron vectors, x, y, such that $x, y > 0$ and $x^T A = \rho(A)x^T$, $Ay = \rho(A)y$, and $x^T y = 1$.

iv. There exists a positive matrix L such that

$$\lim_{m\to\infty} [\rho(A)^{-1}A]^m = L = yx^T.$$

Theorem B.0.23. *A real symmetric matrix A has a spectral decomposition given by*

$$A = \sum_{k=1}^{n} \lambda_k \nu_k \nu_k^T,$$

where λ_i are the eigenvalues of A and ν_i are the orthonormal eigenvectors of A.

Theorem B.0.24. *If A is an invertible matrix with an eigenvalue λ and corresponding eigenvector v, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} corresponding to the eigenvector v.*

Proof. Suppose A is an invertible square matrix with eigenpair (λ, v), we write

$$Av = \lambda v.$$

Multiplying by A^{-1} on both sides, we get

$$v = \lambda A^{-1} v,$$

yielding

$$A^{-1}v = \frac{1}{\lambda} v.$$

□

Theorem B.0.25. *Rayleigh-Ritz Theorem*

Let A be a $n \times n$ Hermitian matrix and let the eigenvalues of A be ordered such that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then

$$\lambda_1 x^* x \leq x^* A x \leq \lambda_n x^* x \text{ for all } x \in \mathbb{C}^n.$$

If A is real, then we choose $x \in \mathbb{R}^n$ and $x^* = x^T$. Moreover

$$\lambda_1 = \inf_{x \neq 0} \frac{x^T A x}{x^T x} \langle x, x \rangle$$

and

$$\lambda_n = \sup_{x \neq 0} \frac{x^T A x}{x^T x} \langle x, x \rangle.$$

APPENDIX C

GRAPH THEORY

Definition C.0.26. Graph.

A *graph* $\mathcal{G}$ is a triple consisting of a vertex set $V(\mathcal{G})$, $|V| = n$, an edge set $E(\mathcal{G}), |E| = m$, and a relation that associates with each edge two vertices (not necessarily distinct) called its endpoints.

Definition C.0.27. Loop.

A *loop* is an edge whose endpoints are equal (same vertex). *Multiple edges* are edges having the same endpoints. A *simple graph* is a graph having no loops or multiple edges. Any two distinct vertices $u, v \in V$ are *adjacent* or *neighbors* if they share an edge. We write this as $u \sim v$.

Definition C.0.28. Path.

A *path* is a simple graph whose vertices are ordered so that two vertices are adjacent if and only if they are consecutive in a list.

Definition C.0.29. Cycle.

A *cycle* is a simple graph with an equal number of vertices and edges, whose vertices are placed around in a circle so that two vertices are adjacent if and only if they are consecutive in a circle.

Definition C.0.30. Walk.

A *walk* is a list $v_0, e_1, v_1, e_2, \cdots, e_k, v_k$ of vertices and edges such that, for $1 \leq i \leq k$, the edge e_i has endpoints v_{i-1} and v_i.

Definition C.0.31. Subgraph.

A *subgraph* of a graph $\mathcal{G}$ is a graph $\mathcal{H}$ such that $V(\mathcal{H}) \subset V(\mathcal{G})$, $E(\mathcal{H}) \subset E(\mathcal{G})$, and the assignment of endpoints in $\mathcal{H}$ is the same as in $\mathcal{G}$. We then write $\mathcal{H} \subset \mathcal{G}$.

Definition C.0.32. Bipartite graph.

A graph is *bipartite* if the set of vertices $V(\mathcal{G})$ is the union of two disjoint sets called partite sets of $\mathcal{G}$ such that every edge connects a vertex in one set to a vertex in the other set. Simple example is a graph with two vertices connected by an edge.

Definition C.0.33. Complete graph.

A graph $\mathcal{G}$ is *complete* if there is an edge connecting each pair of vertices.

Definition C.0.34. Weighted graph.

Weight of an edge in a graph $\mathcal{G}$ is a measure assigned to the edge. A graph is *weighted* if there are two edges of the graph with different weights. If all the edges of $\mathcal{G}$ have the same weight, the graph is *unweighted*.

Definition C.0.35. Degree.

The *degree* of a vertex in a graph, written as $d(v), v \in V$ is the number of edges connected to it. A loop is counted as two edges. A graph is *regular* if all the vertices have the same degree. The order of a graph is the number of vertices and its size is the number of edges.

Theorem C.0.36. *Degree-Vertex formula.*

$$\sum_{v \in V(\mathcal{G})} d(v) = 2e(\mathcal{G}), e \text{ is the size of the graph.}$$

Definition C.0.37. Directed graph.

A *directed graph* or *digraph* is a triple $\mathcal{G}(V, E, f)$, where f is a function assigning each edge an ordered pair of vertices. The first vertex is the tail and the second is the head of the edge.

Theorem C.0.38. *A graph is bipartite if and only if it has no odd cycles.*

Proof. Let $\mathcal{G}$ be a graph. We assume that $\mathcal{G}$ is connected.

Suppose $\mathcal{G}$ is bipartite, we separate its vertices into two disjoint sets A, B, such that $A \cup B = V$ and $A \cap B = \emptyset$. Every edge $e_i \in E$ connects a vertex in A to a vertex in B or vice versa. Suppose we consider any cycle C in $\mathcal{G}$. The set of vertices $a_{i_1} \to b_{i_2} \to a_{i_3} \to \cdots a_{i_1}$ in C are written in cycle form as $(a_{i_1} b_{i_2} \cdots b_{i_k})$. $(a_{i_1} b_{i_2})$ is a cycle of length two and any extension of the same form $(a_{i_j} \cdots b_{i_k})$ is of even length. Hence the bipartite graph has no odd cycles.

We shall prove the reverse by contradiction. Suppose we select any vertex $v \in V$ at random and place it in set A. We divide the rest of the vertices based on their distance from v. If a vertex is odd steps from v, we add it to set A; if not we add it to set B. We have now divided the vertices such that $A \cup B = V$ and $A \cap B = \emptyset$. To prove our theorem, suppose we claim that two vertices x_1, x_2 in A (or B) are adjacent. Then the cycle $v \to \cdots x_1 \to x_2, \cdots w$ is of length $(v \to x_1) + (x_2 \to w) + 1$ is odd and $\mathcal{G}$ is not bipartite, since A(or B) has an adjacent pair of nodes.

□

APPENDIX D

NUMBER THEORY

Lemma D.0.39. *The gcd of any two positive integers a, b can be written as a linear combination of a and b.*

Proof. Consider the set L of all possible positive linear combinations of integers a and b, where $L = \{sa + tb\}$, for some integers s and t. By the well-ordering principle, L has a least element g. Clearly $g > 0$ and $g = ma + nb$ for some integers m, n. Using the division algorithm, we write the quotient of g and a as $a = qg + r,\ 0 \leq r < g$. If $r = 0$, then g divides a. Suppose $r > 0$. Then $r = a - qg$. Since $g = ma + nb$, we have $r = a - (ma + nb)q = (1 - mq)a + (-nq)b$, a linear combination of a and b and $r \in L$. Since g is the least element of L, $g < r$; a contradiction. So, $r = 0$. Similarly, g divides b. Then, g is a common divisor of both a and b.

We now have to show that g is the greatest common divisor. Suppose some positive integer $D \geq g$ divides a and b. Then D divides $ma + nb = g$. And $g \geq D$. Hence $D = g$.

□

Theorem D.0.40. *Let $F = \{f_1, f_2, \cdots, f_n\} \subset \mathbb{Z}^+$ be a finite ordered set with $gcd(F) = g_F$. Then, there exists a positive integer m_F such that for any integer $m \geq m_F$, we write mg_F as a linear combination of elements of F using non-negative*

integers.

$$mg_F = c_1 f_1 + c_2 f_2 + \cdots + c_n f_n,$$

where c_i are non-negative integers.

Proof. First, note that $gcd(F)$ divides every element of F and hence every linear combination of elements of F. Second, there exists a subset $S \subseteq F$ such that $gcd(S) = gcd(F)$. We use an inductive approach on subsets of F.

Step 1. We start with the smallest subset S_1 of F, the first two distinct elements. $S_1 = \{f_1, f_2\}$.

Step 2. Set $g_{S_1} = gcd(S_1)$.

Step 3. Set $m_{S_1} = \frac{f_1 f_2 - f_1 + f_2}{g_{S_1}} + 1$.

Step 4. Clearly, m_{S_1} is greater than all the elements of S_1. By Lemma, D.0.39, g_{S_1} is written as a linear combination of f_1, f_2. Hence,

$$m_{S_1} g_{S_1} = f_1 f_2 - f_1 + f_2 + g_{S_1} = c f_1 + b f_2,$$

for some integers c, b.

Claim D.0.41. c and b are positive integers.

We write $f_1 f_2 - f_1 + f_2 = f_1(f_2 - 1) + f_2$. Now $0 < f_1 < f_2$. Suppose we require $0 < c < f_2$. Since $m_{S_1} g_{S_1} > f_1 f_2 - f_1 + f_2$, we must have $c f_1 + b f_2 > f_1 f_2 - f_1 + f_2$. Hence

$$b > \frac{f_1(f_2 - c - 1)}{f_2} + 1 > 0.$$

Step 5. For the inductive step, let $S_i = \{x\} \cup S_{i-1}$, where $x \in F$ is such that $gcd(S_i) = gcd(F)$. Since F is ordered, $g_{S_i} < x$. Let $X = \{g_{S_{i-1}}, x\}$. We now apply Step 3 to X. Hence

$$m_F = m_{S_i} = \frac{xg_{S_{i-1}} - g_{S_{i-1}} + x}{g_F} + 1.$$

By Step 4,

$$m_F g_F = c_1 g_{S_{i-1}} + c_2 x = \sum b_i f_i,$$

where b_i are positive integers. This is possible since $g_{S_{i-1}}$ is expressed as a linear combination of elements of S_{i-1} and $x \in F$.

Step 6. Finally, suppose $m > m_F$ is a positive integer. Let $m - m_F = a$. Then

$$mg_F = (m_F + a)g_F = \left(\frac{xg_{S_{i-1}} - g_{S_{i-1}} + x}{g_F} + 1 + a\right)g_F = \sum c_i f_i,$$

where c_i are positive integers.

□

Corollary D.0.42. *Suppose $S \subset \mathbb{Z}^+$ is non-empty, closed under addition and $gcd(S) = 1$.*

i. There exists a finite subset $F \subset S$ such that $gcd(F) = 1$.

ii. There exists $m_S \in \mathbb{Z}^+$ such that for any positive integer $m > m_s$, $m \in S$.

iii. In particular $\{m, m+1, m+2, \cdots\} \subset S$.

Proof. S is ordered. Since S is closed under addition, it is infinite.

i. First note that the gcd of finite subsets of S decreases only finite number of times. Consider finite subsets of S such that

$$F_1 \subset F_2 \subset F_3 \cdots \subset S.$$

Clearly, $gcd(F_1) \geq gcd(F_2) \geq \cdots$. Furthermore, $gcd(F_2) \mid gcd(F_1),\ gcd(F_3) \mid gcd(F_2)$ and so forth. Thus $gcd(F_i), i > 1$ is a factor of $gcd(F_1)$. Since the factors are finite, for some $k \in \mathbb{Z}^+$, $gcd(F_k) = gcd(F_{k+1}) = gcd(F_{k+2}) = \cdots$ Consequently, $gcd(F_k) = 1$. Hence $F = F_k$.

ii. This is a direct consequence of Lemma D.0.40.

iii. Since S is closed under addition, $\{m, m+1, m+2, \cdots\} \subseteq S$.

□

Example D.0.43. $S = \{3, 6, 7, 11, \cdots\}$ $gcd(S) = 1$.

1. $F_1 = \{3, 6\};\ \ g_1 = gcd(3, 6) = 3$

2. $m_1 g_1 = 24 = 3 \cdot 2 + 6 \cdot 3$

3. $F_2 = \{3, 6, 7\};\ g_2 = gcd(3, 7) = 1.$

4. $m_2 = 26 + 24 = 50 = 3 \cdot 4 + 6 \cdot 4 + 7 \cdot 2.$

Since $gcd(F_2) = gcd(S)$, we stop here. For any $m \geq 50$, we express $m \cdot g$ as a linear

combination of elements of S using non-negative integers.

$$\begin{aligned} 51 &= 3 \cdot 2 + 6 \cdot 4 + 7 \cdot 3 \\ 52 &= 6 \cdot 5 + 11 \cdot 2 \\ 60 &= 3 \cdot 4 + 6 \cdot 2 + 7 \cdot 2 + 11 \cdot 2 \\ 1005 &= 3 \cdot 335 \\ 1006 &= 3 \cdot 333 + 7 \\ &\vdots \end{aligned}$$

APPENDIX E

POWER METHOD

In computational mathematics, a matrix-free method is an algorithm for solving a linear system of equations or an eigenvalue problem that does not store the coefficient matrix explicitly, but accesses the matrix by evaluating matrix-vector products. One such method is the *power method*. The algorithm for the power method is as follows:

Algorithm for power method

Let V be a vector space over $\mathbb{R}^n$, $A \in V$ be a diagonalizable matrix, and $q \in \mathbb{R}^n$ be any random vector. The computational mechanics is as follows:

for $k = 1, 2, \cdots$

$$z^{(k)} = Aq^{(k-1)} \quad \text{while } \left\|q^{(k)} - q^{(k-1)}\right\| < \epsilon$$

$$q^{(k)} = z^{(k)}/|z^{(k)}|$$

$$\lambda^{(k)} = [q^{(k)}]^T A q^{(k)}$$

end

Since A is diagonalizable, its eigenvectors$\{x_1, x_2, \cdots, x_n\}$ form a basis of $\mathbb{R}^n$. So we write $q^{(0)}$ as a linear combination of these eigenvectors. For scalars $a_1, a_2, \cdots, a_n \in \mathbb{R}$, we write $q^{(0)}$ as

$$q = q^{(0)} = a_1 x_1 + a_2 x_2 + \cdots a_n x_n. \tag{E.1}$$

We now multiply both sides of the above equation by A^k to get

$$\begin{aligned} A^k q^{(0)} &= A^k(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \\ &= a_1\lambda_1^k x_1 + a_2\lambda_2^k x_2 + \cdots + a_n\lambda_n^k x_n \\ &= a_1\lambda_1^k\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)^k x_j\right) \end{aligned}$$

Now, if $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$, then we say that λ_1 is a dominant eigenvalue of A. In such a case $\left(\frac{\lambda_j}{\lambda_1}\right)^k \to 0$ as $k \to \infty$. Therefore, if $a_1 \neq 0$, $A^k q^{(0)} \to a_1\lambda_1^k x_1$. Since the power method normalizes each iteration, this converges to x_1.

$$\begin{aligned} z^{(1)} = Aq^{(0)} &= a_1\lambda_1\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)x_j\right) \\ q^{(1)} &= \frac{a_1\lambda_1\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)x_j\right)}{\left\|a_1\lambda_1\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)x_j\right)\right\|} \\ &= \frac{\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)x_j\right)}{\left\|\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)x_j\right)\right\|} \end{aligned}$$

Extending this, we get

$$\begin{aligned} q^{(k)} = A^k q^{(0)} &= \frac{\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)^k x_j\right)}{\left\|\left(x_1 + \sum_{j=2}^{n}\frac{a_j}{a_1}\left(\frac{\lambda_j}{\lambda_1}\right)^k x_j\right)\right\|} \\ &= x_1 \end{aligned}$$

We get the last step by noting that $\left(\frac{\lambda_j}{\lambda_1}\right)^k \to 0$ and $\|x_1\| = 1$.

The power method converges if λ_1 is dominant and if $q^{(0)}$ has a component in the direction of the corresponding eigenvector x_1, i.e., $a_1 \neq 0$. Since $q^{(0)}$ is random, the probability of a_1 being zero is zero. In reality, the convergence rate depends on the ratio $\frac{|\lambda_2|}{|\lambda_1|}$.

If the power method has converged to the dominant eigenvector x_1 after k iterations, then $[q^{(k)}]^T A q^{(k)} \approx [q^{(k)}]^T \lambda q^{(k)} = \lambda$, since $q^{(k)}$ is normalized after each iteration.