SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

PageRank as a Markov Chain: The Relationship Between Spectral Gap and Mixing Time

av

Celina Arnlund

2026 - No K10

PageRank as a Markov Chain: The Relationship Between Spectral Gap and Mixing Time

Celina Arnlund

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Daniel Ahlberg och Maria Deijfen

2026

Abstract

PageRank is a measure of importance that determines the relevance of a webpage by analyzing its incoming links. The algorithm for computing the PageRank, published in 1998, is a part of Google's search engine, and was developed by Larry Page and Sergey Brin. This paper provides a theoretical understanding of the algorithm from the perspective of Markov chains. By modeling the Web as a directed graph and introducing a Google matrix, the algorithm ensures the existence and the uniqueness of the stationary distribution called PageRank. Furthermore, the paper aims to examine the relationship between the damping factor, the spectral gap and the mixing time. The second largest eigenvalue of the Google matrix is examined in relation to the rate of convergence and mixing time respectively. Finally, the theory is applied to simple graph structures, the cycle and wheel graph, in order to exemplify the concepts.

Sammanfattning

PageRank är ett teoretiskt mått som bestämmer relevansen hos en webbsida genom att analysera dess ingående länkar. Algoritmen för att beräkna PageRank utvecklades av Larry Page och Sergey Brin och publicerades 1998. Det är än idag en del av Googles sökmotor. Denna uppsats syftar till att ge en teoretisk förståelse för algoritmen från perspektivet Markovkedjor. Genom att modellera internet som en riktad graf och introducera en Google matris, så garanterar algoritmen att den stationära fördelningen kallad PageRank existerar och är unik. Vidare undersöks relationen mellan dämpningsfaktorn och spektralgapet. Därefter studeras relationen mellan det näst största egenvärdet till Googlematrisen, konvergenshastigheten och blandningstiden. Slutligen appliceras teorin på enkla grafstrukturer, cykeln och hjulet, i syfte att exemplifiera koncepten.

Acknowledgments

I would like to express my gratitude to my supervisors, Professor Daniel Ahlberg and Professor Maria Deijfen, for their extensive support and guidance throughout this project. Thank you for your inspiration, feedback and patience. I would also like to thank Ewan Gordon for the discussions inspiring me to dive deep into the world of PageRank, and for constructively suggesting improvements after proofreading my work.

Generative AI Disclosure

This thesis acknowledges use of AI (the free version of Google Gemini and ChatGPT-5.5 Thinking) as a tool to support the writing and formatting process. More specifically, AI was used to debug by identifying and suggesting solutions to compilation errors related to the document class, bibliography, and packages setting up the layout. Moreover, the free version of Gemini was used to generate LaTeX code for the wheel and cycle graphs in Section 5. These were based on my own mathematical description of the directions of the edges, number of nodes, etc. When it comes to the analysis, I encountered problems understanding how relationships regarding the second eigenvalue, mixing time and convergence could be derived without the assumption that the Markov chain is reversible. A reversible Markov chain is not a realistic representation of reality in the Google application. However, after discussions with my advisors, we decided that this assumption was a suitable simplification to limit the thesis. As a result of not finding an answer in books and articles, I decided to ask AI how an analysis of the convergence could be done without the assumption of reversibility. However, this is not a path I continued exploring in this thesis, but something brought up as an improvement or further outlook in the discussion section. Lastly, AI tools were used to better understand the mathematical concepts of how left eigenvectors make an orthonormal basis with respect to an inner product. The use of AI was always limited to acting as a tutor and no AI was used to generate the text. All content has been reviewed and edited in a way that represents my own academic work.

Contents

1 Introduction

1.1 Description

The PageRank algorithm was developed by Sergey Brin and Lawrence Page in 1998. It aimed to address many of the problems of large-scale search engines that existed during the nineties, and to provide results of higher quality when searching the Web based on a "relevance score". Before the implementation of PageRank, the search engines relied heavily on the density of keywords. By using the link structure of the Web, a quality ranking called PageRank gets computed for each Web page and used to improve search results.

The intuition behind the theory of PageRank is that the Web can be seen as a directed graph where a surfer randomly clicks on links to visit page after page. The pages that are visited most frequently gain a higher importance score. PageRank can be interpreted as the stationary probability that a random surfer is found on a certain page. Thus, the number of links pointing to a page is important as it impacts how often a random surfer would visit the page. However, the quality of the website pointing to a page is also of importance as all links do not have equal impact or weight. In order to avoid that the random surfer gets stuck in loops or dead ends on the Web graph, it sometimes randomly jumps to an arbitrary page. Hence, the PageRank score is used to list search results in appropriate order based on the importance and amount of other pages linking to it.

1.2 Aim and Research Questions

The aim of this Bachelor Thesis is to describe and study the PageRank-algorithm as a finite Markov chain. In particular, the focus will be on stationary distributions, convergence and mixing time. Furthermore, the thesis aims to investigate how mixing time is effected by the damping factor, and how the the stationary distribution is impacted by different examples of underlying graph structures.

The following research questions will be examined:

- How does the rate of convergence of the PageRank algorithm depend on the spectral properties of the Google matrix?
- How does the damping factor α impact the mixing time for the PageRank-algorithm? Can this be explained by the spectral gap of the Google matrix?
- How does the convergence to the stationary distribution differ depending on underlying graph structures?

1.3 Literature Review

PageRank is a well studied topic with plenty of articles published. It was first introduced by Larry Page, Sergey Brin, Rajeev Motwani and Terry Winograd in 1998 in their report *The PageRank*

Citation Ranking: Bringing Order to the Web, with the aim of ranking Web pages according to human interest and attention devoted to them. They introduced the idea of modeling PageRank as a random Web surfer with the ability to jump to some other page. Later that year Page and Brin presented Google in their paper, *The Anatomy of a large-scale hypertextual Web search engine* as a prototype of a large scale search engine implementing the method of PageRank. These reports established PageRank both as an algorithm to be used in practice, as well as a mathematical problem to be further studied.

In 2003, Haveliwala and Kamvar published a report studying the second eigenvalue of the Google matrix and its implications for the convergence of the standard PageRank algorithm, a topic relevant to this thesis. Moreover, Amy Langville and Carl Meyer (2004, 2006) provide modern sources of the PageRank algorithm in their work, *Deeper Inside PageRank* and *Google's PageRank and Beyond: The Science of Search Engine Rankings* that, among other topics, discuss power iterations and how the second eigenvalue is related to the damping factor in the Google search engine.

To analyze the PageRank algorithm in terms of Markov chains and spectral properties of the Google matrix, a foundation of Markov theory and linear algebra is needed. Friedberg, Insel and Spence (2014) provide the theoretical framework needed to understand the linear algebra related to eigenvalues and eigenvectors, while Ross (2024) contribute with an introduction to probability models relevant to Markov chains, stationary distributions and transition matrices. To study the bounding of mixing times, total variation distance and Markov chains, the book *Markov Chains and Mixing Times* by Levin and Peres (2017) was useful to provide necessary theorems and definitions. Roch (2024) is a reliable source for discrete probability and was a good complement to relate spectral theory to Markov chains.

2 Mathematical/Theoretical Background

This section introduces the theoretical background required to analyze the PageRank algorithm from the perspective presented by the research questions. The material on Markov chains, stationary distributions and transition matrices is mainly based on the book *Introduction to Probability Models* by Ross (2024), while total variation distance and mixing time terminology follows *Markov Chains and Mixing Times* by Levin and Peres (2017). The theory related to linear algebra, specifically eigenvalues, eigenvectors, and transformations, is inspired by *Linear Algebra* by Friedberg et al. (2014), and *Modern Discrete Probability: An Essential Toolkit* by Roch (2024).

2.1 Markov Chains

In this thesis we will only consider finite state spaces as the Web is large but finite. Let $\mathcal{X} = \{1, \ldots, N\}$ denote the state space of a finite Markov chain. In the PageRank application, the states correspond to webpages. Thus, $X_n = i$ means that a random surfer is at webpage i at time n.

Definition 2.1. Let $\mathcal{X} = \{1, \ldots, N\}$ be a finite state space and let X_n denote the value of a process at time n. A discrete-time **Markov chain** is a stochastic process $\{X_n\}_{n \geq 0}$ such that the conditional distribution of X_{n+1} given the values of $X_n, X_{n-1}, ..., X_0$ depends only on the value of X_n. That is,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = P_{ij}$$

for all states $i, j, i_0, ..., i_{n-1}$ and all $n \geq 0$, where P_{ij} is the transition probability from state i to state j [11, p. 201].

The Markov chain is assumed to be time-homogeneous, meaning that the transition probabilities do not depend on time and can therefore be expressed as

$$P_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i).$$

The transition matrix containing the transition probabilities, $P = P_{ij}$, satisfies

$$P_{ij} \geq 0, \quad i, j \in \mathcal{X};$$
$$\sum_{j \in \mathcal{X}} P_{ij} = 1$$

for every $i \in \mathcal{X}$ and is therefore row stochastic [11, p. 201][9].

Evolution of the Distribution

Throughout this thesis, probability vectors are treated as column vectors and therefore the transpose $\pi^{(n)T}$ will be used when distributions are multiplied from the left by a row stochastic transition matrix. Let the initial distribution of a discrete-time Markov chain be the probability distribution

of the initial state X_0 denoted as $\pi^{(0)}$. Thus for each state i,

$$\pi_i^{(0)} = P(X_0 = i).$$

Similarly, let $\pi^{(n)}$ denote the distribution of X_n such that

$$\pi_j^{(n)} = P(X_n = j)$$

for each state j. Since the probability vectors are treated as column vectors, the distribution evolves by multiplying $\pi^{(n)T}$ from the left by the transition matrix P. Thus,

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

Iterating the process makes the distribution evolve according to

$$\pi^{(n)T} = \pi^{(0)T} P^n.$$

2.2 Irreducibility and Aperiodicity

In order for a finite Markov chain to converge to a unique stationary distribution regardless of initial distribution, two properties are required: irreducibility and aperiodicity. This enables the Markov chain to move between states and return to a state at any time. Since this thesis only concerns finite state spaces, irreducibility implies that the chain is positive recurrent, which is a condition otherwise needed in the Convergence theorem. Hence, recurrence and positive recurrence will not be defined or used.

Definition 2.2. A Markov chain is **irreducible** if all states communicate with each other, hence if there is only one class so that all states are accessible to each other [11, p. 214] That is, for all $i, j \in \mathcal{X}$, there exists some $n \geq 1$ such that

$$(P^n)_{ij} > 0.$$

Thus, if the random surfer starts in state i there is a positive probability of reaching state j in a finite number of steps.

Definition 2.3. A chain that can only return to a state in a multiple of $d > 1$ steps is called **periodic**. If the chain does not return to a state in fixed regular intervals, the chain is called **aperiodic**. That is, if the period $d(i) = \gcd\{n \geq 1 : (P^n)_{ii} > 0\} = 1$ then state i is aperiodic [11, p. 240-241].

Thus, aperiodicity guarantees that each state can return to itself in one step as the probability of staying in the same state is $P_{ii} > 0$ for all i. Since a finite irreducible Markov chain is positive recurrent, a finite Markov chain is called **ergodic** when it is irreducible and aperiodic [11, p. 241].

2.3 Random Walks on Graphs

Definition 2.4. A **directed graph** consists of a set V of nodes or elements and a set $A \subseteq V \times V$ of ordered pairs of elements of V, called arcs. Thus the arcs have directions [11, p. 152].

For a graph with state space $\mathcal{X}$, let A denote the adjacency matrix where

$$A_{ij} = \begin{cases} 1, & \text{if there is a directed edge from node } i \text{ to node } j \\ 0, & \text{otherwise.} \end{cases}$$

If matrix A is symmetric, it implies that the graph is undirected, as there is an edge from node i to node j if and only if there is an edge from node j to node i.

Assuming that each row in the adjacency matrix contains a non-zero element so that each node has an outgoing edge, we can define a random walk where the random surfer uniformly chooses an edge to follow at every step. Let C_i denote the row sum in adjacency matrix A. That is

$$C_i = \sum_{j=1}^{N} A_{ij}.$$

Then the entries of a transition matrix P for the random walk is given by

$$P_{ij} = \frac{A_{ij}}{C_i}.$$

To verify that P is a transition matrix satisfying the conditions for a row stochastic matrix, note that since $A_{ij} \geq 0$ and $C_i > 0$, it follows that $P_{ij} \geq 0$. Moreover, we show that the rows of matrix P sum up to 1:

$$\sum_{j=1}^{N} P_{ij} = \sum_{j=1}^{N} \frac{A_{ij}}{C_i} = \frac{C_i}{C_i} = 1.$$

In order for the described random walk to be irreducible, it has to be possible to reach any state by following directed edges. Equivalent to irreducibility of the Markov chain, is that the corresponding directed graph is strongly connected since a directed path exists from i to j for every ordered pair (i, j). If the chain can return to a state in an irregular amount of steps, (hence not restricted to a multiple of some integer $d > 1$), then it is aperiodic. In a graph this could be guaranteed by implementing self loops as it makes it possible to return to a state in one step. Without these properties the chain fails to converge to a unique stationary distribution. Another way to make sure these properties are guaranteed is by implementing a probability $\alpha > 0$ to take a step by following edges, and a probability $1 - \alpha$ to jump to another random node uniformly or according to a probability vector v. Moreover, aperiodicity can be guaranteed by using a lazy-chain where the random walker stays in state i with a probability $\beta > 0$. This discussion is a preview to be continued in the section about the PageRank Algorithm.

2.4 Stationary Distributions

In order for the PageRank ranking to be stable, it is important to know whether the distribution eventually stops changing after another step of the Markov chain is taken. The following definition of a stationary distribution is relevant to know how the chain stabilizes as time goes to infinity.

Definition 2.5. The **stationary distribution** π of a Markov chain with transition matrix P satisfies the equations:

$$\pi_j = \sum_i \pi_i P_{i,j}, \quad j \geq 1 \quad \text{Matrix form:} \quad \pi^T = \pi^T P$$
$$\sum_j \pi_j = 1 \quad \text{and} \quad \pi_j \geq 0.$$

[11, p. 226]

The following two theorems are central to this thesis.

Theorem 2.6. *Let P be the transition matrix of an irreducible finite Markov chain. Then there exists a unique probability distribution π satisfying $\pi^T = \pi^T P$ [9, p. 13].*

Although, Theorem 2.6 ensures the uniqueness of a stationary distribution, it does not guarantee convergence to the stationary distribution regardless of initial distribution. For convergence, aperiodicity is also needed.

Theorem 2.7. ***(The Convergence Theorem)*** *Let P be the transition matrix for a finite, irreducible and aperiodic Markov chain with stationary distribution π. Then there exists constants $a \in (0, 1)$ and $C > 0$ such that for all states in $i \in \mathcal{X}$ and all $n \geq 0$*

$$\sum_j |(P^n)_{ij} - \pi_j| \leq Ca^n, \quad n \geq 1.$$

In particular, this implies that for all $i, j \in \mathcal{X}$ the chain converges to its stationary distribution:

$$\lim_{n\to\infty} (P^n)_{ij} = \pi_j.$$

The proof of Theorem 2.6 and Theorem 2.7 can be found in Levin and Peres (2017) [9, p.13, pp. 52-53]. The convergence to an exponential expression will be a reoccurring topic in this thesis, which is why it was important to establish the definitions for irreducibility and aperiodicity before studying the long term behavior of finite Markov chains.

2.5 Total Variation Distance and Mixing Time

Total variation distance is a measure of distance between the current distribution of a Markov chain and the stationary distribution. This concept will be useful in this thesis to examine how many iterations of the power method are required in order for the ranking of webpages to be considered good enough.

Definition 2.8. The **total variation distance** between two probability distributions μ and ν on a finite state space $\mathcal{X}$ is defined by

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)|$$

and represents the maximum difference between the probabilities assigned to a single event by the two distributions [9, p. 47].

By Proposition 4.2 provided in Levin and Peres [9, p. 48], the total variation distance is reduced to a simple sum over state space. For two probability distributions μ and ν on a finite state space $\mathcal{X}$, this useful characterization can be expressed

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|. \tag{1}$$

Mixing time measures the minimal time (or number of steps) required by a Markov chain to guarantee that the distance to the stationary distribution is less than the tolerance level ε, regardless of initial distribution. The tolerance level ε is often chosen to be $1/4$. More generally, any fixed value $\varepsilon \in (0, 1/2)$ can be used to give comparable expressions for mixing time so that the distance is forced to decrease rapidly. It is relevant in the PageRank setting to determine the number of iterations necessary for a stable ranking. To measure the worst case distance to stationary, we define the maximal distance over all initial distributions to stationary after t steps as

$$d(t) := \sup_{\mu \in \mathcal{P}(\mathcal{X})} \|\mu^T P^t - \pi^T\|_{TV} \tag{2}$$

where $\mathcal{P}(\mathcal{X})$ denotes the collection of all probability distributions on the finite state space $\mathcal{X}$, P is the transition matrix, μ is the initial probability distribution, and π is the stationary distribution. This leads to the formal definition of mixing time:

Definition 2.9. For $\varepsilon > 0$, the **mixing time** is defined by

$$t_{\text{mix}}(\varepsilon) = \min\{t \geq 0 : d(t) \leq \varepsilon\}.$$

2.6 Eigenvalues and Spectral Gap

Definition 2.10. Let $A \in \mathbb{R}^{n \times n}$. A nonzero column vector $v \in \mathbb{C}^n$ is called a right **eigenvector** of A if there exists some scalar $\lambda \in \mathbb{C}$ such that

$$Av = \lambda v$$

where λ is called the corresponding **eigenvalue** [5, p. 246]. A nonzero row vector u^T is called a left eigenvector of A with eigenvalue λ if

$$u^T A = \lambda u^T.$$

Throughout this thesis vectors are treated as column vectors. Let $e \in \mathbb{R}^n$ denote the column

vector of all ones. Since the transition matrix P is row stochastic, the i-th component of Pe can be expressed

$$(Pe)_i = \sum_{j=1}^{n} P_{ij} \cdot 1 = \sum_{j=1}^{n} P_{ij} = 1.$$

Hence, every component of Pe equals 1, which shows that e is a right eigenvector of P with corresponding eigenvalue 1 since

$$Pe = 1 \cdot e.$$

However, since the stationary distribution satisfies

$$\pi^T P = 1 \cdot \pi^T$$

π^T is a left eigenvector of P corresponding to the eigenvalue of 1. This distinction between right and left eigenvectors will be important moving forward.

Theorem 2.11. *Every eigenvalue λ of a stochastic matrix P satisfies*

$$| \lambda | \leq 1$$

[8, p. 41].

Proof. For a stochastic matrix P, all elements are nonnegative and the row sum for every row i is 1. Hence,

$$P_{ij} \geq 0 \quad \text{and} \quad \sum_{j=1}^{n} P_{ij} = 1, \quad \forall i.$$

Let λ be an eigenvalue of P and let $v \neq 0$ be a corresponding right eigenvector so that $Pv = \lambda v$. Now, the component with the greatest absolute value of the vector v and index i can be expressed as

$$| v_i | = \max_{1 \leq k \leq n} | v_k |$$

and since $v \neq 0$, it can be concluded that $| v_i | > 0$. The i-th row can be extracted from the eigenvalue problem $Pv = \lambda v$:

$$\sum_{j=1}^{n} P_{ij} v_j = \lambda v_i.$$

According to the triangle inequality $| a + b | \leq | a | + | b |$. After taking the absolute value on both sides and applying the triangle inequality gives

$$
\begin{aligned}
|\lambda| \cdot |v_i| &= \left| \sum_{j=1}^{n} P_{ij} v_j \right| \leq \sum_{j=1}^{n} |P_{ij} v_j| \\
|\lambda| \cdot |v_i| &\leq \sum_{j=1}^{n} P_{ij} |v_j|, && \text{since } P_{ij} \geq 0 \\
|\lambda| \cdot |v_i| &\leq \sum_{j=1}^{n} P_{ij} |v_i| = |v_i| \left(\sum_{j=1}^{n} P_{ij} \right), && \text{since } |v_j| \leq |v_i| \\
|\lambda| \cdot |v_i| &\leq |v_i| \cdot 1, && \text{since } \sum_{j=1}^{n} P_{ij} = 1
\end{aligned}
$$

Dividing both sides of the last row of equations by $|v_i| > 0$ results in $|\lambda| \leq 1$. □

Note that the theorem holds when we consider left eigenvectors as well, since the right and left eigenvalues of a square matrix $A \in \mathbb{R}^{n \times n}$ are the same. This can be shown by considering the right eigenvector v satisfying $Av = \lambda v$ and the left eigenvector u^T satisfying $u^T A = \lambda u^T$. Transposing the left eigenvector equation gives us $A^T u = \lambda u$. Since A and A^T have the same characteristic equation $\det(A^T - \lambda I) = \det((A - \lambda I)^T) = \det(A - \lambda I)$ it follows that they have the same roots or eigenvalues. Therefore, since the transition matrix is a square matrix, it is not necessary to specify if the eigenvalues come from right or left eigenvectors, although it is important to distinguish between the left and right eigenvectors themselves as they are usually different. In general, this thesis will concern the right spectrum, but for probability distributions, left eigenvectors will be used as they are multiplied from the left.

Another important concept to define in order to explain what happens when a Markov chain approaches its stationary distribution, is the spectral gap. The eigenvalues of a transition matrix contain information about the long term behavior of the chain and will play an important role when we later study the convergence properties of the PageRank algorithm.

Definition 2.12. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of a stochastic matrix P, ordered by decreasing absolute value. That is $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Since 1 is an eigenvalue of P and every eigenvalue satisfies $|\lambda| \leq 1$, we have $|\lambda_1| = 1$. The **spectral gap** γ is defined as

$$\gamma = |\lambda_1| - |\lambda_2| = 1 - |\lambda_2|.$$

3 PageRank Algorithm

3.1 Description of the PageRank Algorithm

In 1998, Sergey Brin and Lawrence Page presented Google as a prototype of a large-scale search engine, relying on the link structure of the Web and a ranking called PageRank to improve search results [2]. PageRank measures the importance of a webpage based on the number and quality of links pointing to it. Let $\{1, 2, \ldots, N\}$ denote the set of webpages in a network so that the total number of webpages is given by N. Assume page i is linked to by pages in the set denoted as B_i and that C_j denotes the number of outgoing links from page j. Each page distributes its PageRank uniformly across its outgoing links. Then a first attempt at defining PageRank of page i can be expressed as the following recursive relation:

$$PR(i) = \sum_{j \in B_i} \frac{PR(j)}{C_j} \tag{3}$$

where $PR(i)$ denotes the PageRank value of page i [8, pp. 32-34]. Thus, the importance of a page i is determined by the importance of the pages $j \in B_i$ linking, to it divided by the number of outgoing links. It is important to note that this definition is different to a later definition of PageRank after implementation of the damping factor and teleportation properties. In this thesis, final PageRank will be the stationary distribution to the Google matrix, not the unadjusted hyperlink matrix. This will be further discussed in later sections.

Since the PageRanks of the pages in the set B_i linking to page i are unknown, Brin and Page assumed that every page in the Web could be assigned an initial value of $\frac{1}{N}$, where N is the number of pages, so that the sum of all PageRanks is 1. Then Equation (3) was used to iteratively compute and update the PageRank for each page according to

$$PR_{t+1}(i) = \sum_{j \in B_i} \frac{PR_t(j)}{C_j} \tag{4}$$

where $PR_{t+1}(i)$ is the PageRank of page i at time or iteration $t+1$ and the process was initiated with $PR_0(i) = \frac{1}{N}$ for all pages i.

The summation can be expressed as a matrix representation, where the PageRank values are represented by a $1 \times N$ row vector. Let $\tilde{H}$ define the adjacency matrix of the Web graph, that is

$$\tilde{H}_{ij} = \begin{cases} 1, & \text{if page } i \text{ links to page } j \\ 0, & \text{otherwise.} \end{cases}$$

Then the rows correspond to outgoing links so that $C_i = \sum_{j=1}^{N} \tilde{H}_{ij}$. To obtain the transition probabilities describing a random walk on the Web rather than just the link structure of the Web like in $\tilde{H}$, the matrix needs to be row normalized. The entries of the corresponding row normalized matrix can be defined as $\frac{\tilde{H}_{ij}}{C_i}$ where the outgoing links from page i satisfy $C_i > 0$. Equivalently,

we could write the relation as:

$$H_{ij} = \begin{cases} \frac{1}{C_i}, & \text{if page i links to page j,} \\ 0, & \text{otherwise.} \end{cases}$$

and therefore each row corresponding to a page with outgoing links contains nonnegative entries, representing probabilities that sum up to 1. Let $PR^{(t)}$ denote the column vector containing the PageRank values of all pages. Since probability vectors are multiplied from the left in this thesis, the PageRank equation (3) can be written in matrix form as

$$PR^T = PR^T H.$$

Then the iterative PageRank equation (4) can be written in matrix form as

$$PR^{(t+1)T} = PR^{(t)T} H.$$

In order to mimic user behavior, PageRank can be described as a Markov chain where a random surfer starts at a given page on the Web and clicks on links. If a row of $\tilde{H}$ consists only of zeros, then the corresponding page is a dangling node without outgoing links. As a result, the corresponding row in matrix H cannot be normalized. To avoid problems where the random surfer gets stuck at a dangling node such as a pdf file or an image, it is necessary to handle the zero rows in the matrix representation H of Equation (4). The stochasticity adjustment of replacing the 0^T rows by $\frac{1}{N}e^T$, where e is a column vector of all ones ($e = (1, 1, \ldots, 1)^T \in \mathbb{R}^N$), results in that a random surfer may move to any page uniformly at random when encountering a dangling node. Let's denote this stochastic matrix as P [8, p. 37]. Now the original PageRank vector can be interpreted as a stationary distribution of the random walk on graph corresponding to a row normalized hyperlink matrix without zero rows. That is

$$PR^T = PR^T P$$

where P is understood as the adjusted version of the row normalized hyperlink matrix H. Important to note is that this is not the final PageRank model, just the stationary distribution of a random walk on a hyperlink graph without dangling nodes. However, there is no guarantee that this is a unique stationary distribution since the hyperlink matrix may not be irreducible. Moreover, the power iteration may not converge to the stationary distribution from every initial distribution if the hyperlink matrix is not aperiodic. This leads to the implementation of the damping factor α and the teleportation vector v, and will be further discussed in the next section. Brin and Page imagined that the random surfer sometimes gets bored and "jumps" to a different Web page instead of following the links. The probability that the surfer keeps following links is denoted by the damping factor α, where $0 < \alpha < 1$. Hence, the probability of the surfer jumping to a new part of the Web is $1 - \alpha$. Typically, the value 0.85 is chosen as the damping factor[2]. Where on the Web the surfer jumps is decided by the teleportation vector v (column vector). In the most basic model the vector is assumed to be uniformly distributed, so $v = \frac{1}{N}e$. Including all of the described concepts helps us define the Google matrix for the general case with teleportation vector v, the adjusted hyperlink matrix P, and row vectors:

$$G = \alpha P + (1 - \alpha) e v^T. \tag{5}$$

In the special case of uniform teleportation where $v = \frac{1}{N} e$ the Google matrix is given by:

$$G = \alpha P + (1 - \alpha) \frac{1}{N} e e^T.$$

Finally, Google's PageRank can be computed by applying the power method to the Google matrix. Let $\pi^{(0)}$ be the initial probability distribution. Iteration is given by

$$\pi^{(t+1)T} = \pi^{(t)T} G. \tag{6}$$

We now denote the stationary distribution as the row vector π^T. After convergence, the following relation is satisfied [8]:

$$\pi^T = \pi^T G. \tag{7}$$

3.2 Properties of the Google Matrix

Having defined the Google matrix, we now study the properties stochasticity, aperiodicity and irreducibility in order to confirm that the Google matrix is suitable for the Markov chain interpretation of PageRank. Suppose N is the total number of pages in the network and that the damping factor satisfies $0 < \alpha < 1$. We also suppose the teleportation vector $v \in \mathbb{R}^N$ satisfies

$$v_i > 0 \quad \text{for all } i, \quad \sum_{i=1}^{N} v_i = 1$$

and that P is the matrix obtained from the hyperlink matrix after handling the problem with dangling nodes. The Google matrix is then defined in Equation (5), where $e = (1, 1, ..., 1)^T \in \mathbb{R}^N$. The iterations evolve as described in section "Evolution of the distribution":

$$\pi^{(t+1)T} = \pi^{(t)T} G$$

where $\pi^{(t)T}$ is a row vector of probabilities. By definition, a matrix is row stochastic if all entries are nonnegative, and each row sums to 1. This is necessary for a Markov chain transition matrix and hence we want to verify if handling the problem with dangling nodes ensures stochasticity for matrix P.

Proposition 3.1. *Assume matrix P has been adjusted to avoid dangling nodes by replacing zero rows. Matrix P is row stochastic.*

Proof. By definition, for a non-dangling node i that has C_i outgoing links, we have

$$\sum_{j=1}^{N} P_{ij} = C_i \cdot \frac{1}{C_i} = 1.$$

For a dangling node i without any outgoing links, suppose we replace the whole row by $\frac{1}{N}e^T$ to make all elements nonnegative. We then obtain the row sum

$$\sum_{j=1}^{N} \frac{1}{N} = N \cdot \frac{1}{N} = 1.$$

Hence, each row of matrix P is nonnegative and sums to 1. Therefore P is row stochastic. □

Since the Google matrix is constructed from the transition matrix P and the teleportation matrix ev^T according to (5), we now want to verify if matrix G is also a transition matrix.

Proposition 3.2. *Assume P is row stochastic, damping factor $0 < \alpha < 1$, and that the teleportation vector $v_i > 0$ for all pages i. Then the Google matrix*

$$G = \alpha P + (1 - \alpha)ev^T$$

is row stochastic.

Proof. Assuming that P is row stochastic gives us that $P_{ij} \geq 0$ for all i, j. Since $v_i > 0$ for all i we get the elements $(ev^T)_{ij} = v_j > 0$ for all j. Thus, all elements of ev^T are also nonnegative and therefore it follows that

$$G_{ij} = \alpha P_{ij} + (1 - \alpha)(ev^T)_{ij} \geq 0.$$

By the definition of a row stochastic matrix it suffices to prove that the rows sum up to 1 in order to prove that a matrix is stochastic. The row sum for the Google matrix is given by:

$$\begin{aligned} \sum_{j=1}^{N} G_{ij} &= \sum_{j=1}^{N} (\alpha P_{ij} + (1 - \alpha)v_j) \\ &= \alpha \sum_{j=1}^{N} P_{ij} + (1 - \alpha) \sum_{j=1}^{N} v_j \quad \text{(by linearity).} \end{aligned}$$

Since P is stochastic and v is a probability vector we have

$$\sum_{j=1}^{N} P_{ij} = 1 \quad \text{and} \quad \sum_{j=1}^{N} v_j = 1.$$

Hence, by summing the elements in row i of the Google matrix we get

$$\sum_{j=1}^{N} G_{ij} = \alpha \cdot 1 + (1 - \alpha) \cdot 1 = \alpha + 1 - \alpha = 1.$$

Thus, the Google matrix G is row stochastic. □

To make sure that the chain converges to a unique stationary distribution, we need to make sure that the Markov chain can move between all states and does not get stuck due to periodicity. This is the purpose of the teleportation properties of the random surfer.

Proposition 3.3. *If $0 < \alpha < 1$ and $v_j > 0$ for all j, then the Google matrix G is irreducible and aperiodic.*

Proof. Assume P is a stochastic matrix, $0 < \alpha < 1$ is the damping factor and v is the teleportation vector such that $\sum_{j=1}^{N} v_j = 1$. We define the Google matrix in Equation (5) element wise:

$$G_{ij} = \alpha P_{ij} + (1 - \alpha)v_j.$$

By Definition 2.2, a finite Markov chain is irreducible if for all states for each pair of nodes (i, j) there is a finite path from i to j. Thus, there has to exist sone $n \geq 1$ such that $(G^n)_{ij} > 0$.

Since $P_{ij} \geq 0$, $0 < \alpha < 1$ and $v_j > 0$ for all j, we have $(1 - \alpha)v_j > 0$, and it follows

$$G_{ij} > 0 \quad \forall i, j \in \{1, ..., N\}.$$

Hence, there is a positive probability that it is possible to go from each state or page i to state j in one step. Therefore, G is irreducible.

Since $G_{ij} > 0 \quad \forall i, j$, the inequality is also valid for the case where $i = j$:

$$G_{ii} > 0 \quad \forall i \in \{1, ..., N\}.$$

This ensures that in the Markov chain defined by G there is a positive probability of returning to state i from itself in one step. Denote the probability of going from node i to node i in exactly k steps as $(G^k)_{ii}$. Then we have $(G^1)_{ii} = G_{ii} > 0$, meaning that the set $\{k \geq 1 : (G^k)_{ii} > 0\}$ contains the number 1. By Definition 2.3, aperiodicity depends on the the period $d(i) = gcd\{k \geq 1 : (G^k)_{ii} > 0\} = 1$ and since it is possible to return to node i in the Google matrix in one step, the greatest common divisor $d(i) = \gcd\{1, \dots\} = 1$ and we can conclude that the chain represented by the Google matrix is also aperiodic.

□

3.3 Existence and Uniqueness of PageRank

PageRank guarantees that each Web page receives a unique and stable ranking value. As previously shown, the Google matrix is row stochastic, irreducible and aperiodic. By using these properties, we can now study the long term behavior of the Markov chain used in the PageRank algorithm.

Theorem 3.4. *The Google matrix G has a unique stationary distribution*

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

and the stationary distribution is the PageRank vector.

Proof. Since the Google matrix represents a finite and irreducible chain by Proposition 3.3, it is positive recurrent. By Theorem 2.6, there exist a unique probability distribution π^T satisfying $\pi^T = \pi^T G$. The stationary distribution π^T is the PageRank vector. □

Theorem 3.5. *Assume $0 < \alpha < 1$ and $v_j > 0$ for all j and $\sum_{j=1}^{N} v_j = 1$. Then for any initial probability vector $\pi^{(0)T}$ the Google matrix converges through power iteration. That is*

$$\lim_{t\to\infty} \pi^{(0)T} G^t = \pi^T$$

where π^T is the unique stationary distribution denoted as the PageRank vector.

Proof. The Google matrix G is stochastic and can be described as the transition matrix for a finite Markov chain. In Theorem 3.4 the existence and uniqueness of the stationary distribution was stated as a result of the chain represented by the Google matrix G being irreducible and positive recurrent. By Proposition 3.3 the Google matrix is aperiodic, and therefore it follows that the power iteration converges to the unique stationary distribution (PageRank vector) independent of the initial probability vector $\pi^{(0)T}$ by the Theorem 2.7. □

Thus, the stationary distribution of the Google matrix G is what we call PageRank, and by Theorem 3.4 and Theorem 3.5 the stationary distribution is unique and the PageRank algorithm converges to it. However, note that since G depends on both α and v, the resulting PageRank vector also depends on both α and v. When $\alpha = 0$ the Google matrix is expressed $G = ev^T$. Since every row of the Google matrix then corresponds to the teleportation vector v^T, the random surfer just jumps around uniformly and hence the stationary distribution $\pi^T = v^T$. However, when $\alpha > 0$ the Google matrix is given by Equation (5) and then the stationary distribution is depending on both the link structure from matrix P and the teleportation from vector v. Thus, for every combination of α and v there exists a unique stationary distribution. Note also that this is an important distinction from the stationary distribution for the hyperlink matrix P, corresponding to the random walk without damping.

3.4 Construction of the Google Matrix Exemplified

The purpose of this section is to illustrate how the Google matrix is created, by considering the following example using small graph structures. Let $\tilde{H}$ denote the original hyperlink matrix reflecting the graph structure of the internet. Define $\tilde{H}$ as

$$\tilde{H} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The rows of matrix $\tilde{H}$ correspond to the outgoing links. For example, page $i = 2$ has two outgoing links to pages 3 and 4, and therefore the total number of outgoing links is $C_2 = \sum_{j=1}^{4} \tilde{H}_{ij} = 1 + 1 = 2$. To become a transition matrix the rows need to sum to 1 and therefore the hyperlink matrix needs to be row normalized. Let the modified matrix be denoted as H and define every

element as $H_{ij} = \frac{\tilde{H}_{ij}}{C_i}$. Then the following matrix is obtained:

$$H = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Since the matrix contains a zero row, it is by definition not stochastic. The zero row, representing a dangling node, is therefore replaced by a uniform distribution of all pages. Denote the obtained matrix as P. We get the following:

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

Now all rows of matrix P sum up to one and contain nonnegative entries, making it a valid transition matrix. The next step is to implement teleportation properties to ensure that the corresponding Markov chain is irreducible. Let the teleportation vector $v = \frac{1}{4}e$ be uniform and the damping factor satisfy $0 < \alpha < 1$. The Google matrix can then be defined according to Definition 3.1:

$$\begin{aligned} G &= \alpha P + (1-\alpha)\frac{1}{4}ee^T \\ &= \alpha \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} + (1-\alpha) \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}. \end{aligned}$$

Simplifying the expression gives us the following expression for the Google matrix:

$$G = \begin{pmatrix} \frac{1-\alpha}{4} & \alpha + \frac{1-\alpha}{4} & \frac{1-\alpha}{4} & \frac{1-\alpha}{4} \\ \frac{1-\alpha}{4} & \frac{1-\alpha}{4} & \frac{\alpha}{2} + \frac{1-\alpha}{4} & \frac{\alpha}{2} + \frac{1-\alpha}{4} \\ \frac{\alpha}{2} + \frac{1-\alpha}{4} & \frac{1-\alpha}{4} & \frac{1-\alpha}{4} & \frac{\alpha}{2} + \frac{1-\alpha}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

This example shows how to distinguish the Google matrix from the hyperlink matrix by modification such as row normalization, replacement of zero rows and implementation of the damping factor and teleportation vector.

4 Analysis of Convergence and Mixing Time

It has been concluded that the iterative algorithm $\pi^{(t+1)T} = \pi^{(t)T}G$ converges to a unique stationary distribution, the PageRank vector. The relevance of the second eigenvalue can be made concrete in the case where the underlying Markov chain is reversible. In general, this is not a property of the Google matrix since the underlying graph is directed. Yet, the absolute value of the second largest eigenvalue is often used to explain the rapid convergence of the PageRank algorithm. This section will be dedicated to examining the relationship between the damping factor α and the rate of convergence, mixing time, and the spectral gap under the simplifying assumption that the Markov chain is reversible. First, the relationship between the eigenvalues of the modified hyperlink matrix P and the Google matrix G will be studied.

4.1 Spectrum of the Google Matrix

In this section, we consider the right eigenvalues of matrix P denoted by

$$\lambda_1(P), \lambda_2(P), \ldots, \lambda_N(P).$$

Since P is stochastic and $Pe = e$ it follows that $\lambda_1(P) = 1$ is a right eigenvalue. In decreasing order of absolute value, the eigenvalues are ordered as

$$1 = \lambda_1(P) \geq \lambda_2(P) \geq \cdots \geq \lambda_N(P).$$

Theorem 4.1. *Let the Google matrix be defined as $G = \alpha P + (1-\alpha)ev^T$, where $P \in \mathbb{R}^{N\times N}$ is a row stochastic transition matrix, $\alpha \in (0,1)$ is the damping factor, $v \in \mathbb{R}^N$ is a teleportation column vector containing probabilities, and $e \in \mathbb{R}^N$ is a column vector of ones.*

Assume the stochastic matrix P has the set of eigenvalues $\{1, \lambda_2, \lambda_3, \ldots, \lambda_N\}$. Then the set of eigenvalues for the Google matrix is $\{1, \alpha\lambda_2, \ldots, \alpha\lambda_N\}$ as stated in [8, p. 41].

Proof. This proof is a detailed version inspired by the proof to Theorem 4.7.1 in [8, p. 46]. According to an example in [5, p. 261], assume that the matrix Q is a $n \times n$ matrix whose columns consist of eigenvectors of a diagonalizable matrix A. Then the transformation $Q^{-1}AQ$ would result in a diagonal matrix with the eigenvalues of A on the diagonal, corresponding to respective columns of Q. Thus, in such case $Q^{-1}AQ = \Lambda$ where Λ is diagonal. In our case, we consider the transition matrix P that is row stochastic and satisfies $Pe = e$. However, instead of assuming that P is diagonalizable and choosing a matrix Q consisting of eigenvectors, we choose Q so that the first column is e. Hence, instead we consider that two matrices A and B are similar if there exists an invertible matrix Q such that $B = Q^{-1}AQ$ and that similar matrices have the same eigenvalues even if the transformed matrix is not diagonal [4][7][8]. This can be proved by considering the characteristic equation of both matrices A and B and showing that they are the same. The characteristic equation for matrix A is given by $\det(A - \lambda I) = 0$. The characteristic equation of matrix B is given by $\det(B - \lambda I) = 0$. Developing the characteristic polynomial using

linear algebra framework from [5] gives us:

$$\begin{aligned}
\det(B - \lambda I) &= \det(Q^{-1}AQ - \lambda I) \\
&= \det(Q^{-1}AQ - \lambda Q^{-1}IQ) \\
&= \det(Q^{-1}AQ - Q^{-1}\lambda IQ) \\
&= \det(Q^{-1}(A - \lambda I)Q) \\
&= \det(Q^{-1})\det(A - \lambda I)\det(Q) \\
&= \frac{1}{\det(Q)}\det(A - \lambda I)\det(Q) \\
&= \det(A - \lambda I).
\end{aligned}$$

This shows that two similar matrices B and A have the same eigenvalues.

Now, let Q be a matrix where the first column is e, since e is a right eigenvector of P corresponding to the eigenvalue 1. Choose the remaining columns of Q, denoted by X, so that the matrix is invertible. Thus, $Q = \begin{pmatrix} e & X \end{pmatrix}$. Denote the first row of the inverted matrix y^T and the remaining rows Y^T such that $Q^{-1} = \begin{pmatrix} y^T \\ Y^T \end{pmatrix}$. A property of invertible matrices is that multiplication of a matrix and its inverse results in the identity matrix. Thus,

$$Q^{-1}Q = \begin{pmatrix} y^T \\ Y^T \end{pmatrix}\begin{pmatrix} e & X \end{pmatrix} = \begin{pmatrix} y^Te & y^TX \\ Y^Te & Y^TX \end{pmatrix} = I.$$

From this it follows that $y^Te = 1$ and that $Y^Te = 0$ since the entries correspond to the first column of the identity matrix. A matrix denoted as M is similar to matrix P if there exists an invertible matrix Q such that $M = Q^{-1}PQ$ [8][5, p. 115]. We compute matrix M:

$$\begin{aligned}
M &= Q^{-1}PQ \\
&= \begin{pmatrix} y^T \\ Y^T \end{pmatrix} P \begin{pmatrix} e, X \end{pmatrix} \\
&= \begin{pmatrix} y^TPe & y^TPX \\ Y^TPe & Y^TPX \end{pmatrix}
\end{aligned}$$

We know $Pe = e$ since P is row stochastic, and that $y^Te = 1$ and $Y^Te = 0$. Applying these properties results in:

$$\begin{aligned}
M &= \begin{pmatrix} y^TPe & y^TPX \\ Y^TPe & Y^TPX \end{pmatrix} \\
&= \begin{pmatrix} y^Te & y^TPX \\ Y^Te & Y^TPX \end{pmatrix} \\
&= \begin{pmatrix} 1 & y^TPX \\ \mathbf{0} & Y^TPX \end{pmatrix}
\end{aligned}$$

This shows that matrix M is a block upper triangular matrix. The eigenvalues of M are found using the characteristic equation. Thus,

$$\det(M - \lambda I) = \det \begin{pmatrix} 1-\lambda & y^T PX \\ \mathbf{0} & Y^T PX - \lambda I_{N-1} \end{pmatrix} = (1-\lambda) \det \left(Y^T PX - \lambda I_{N-1} \right) = 0.$$

Thus, one solution is given by $(1-\lambda) = 0$, implying that the first eigenvalue $\lambda_1 = 1$, meanwhile $\lambda_2, ..., \lambda_N$ are given by the characteristic equation of the block $\det \left(Y^T PX - \lambda I_{N-1} \right) = 0$. Since P and M are similar they have the same eigenvalues, thus $\{1, \lambda_2, ..., \lambda_N\}$. Now the same transformation is applied to the Google matrix $G = \alpha P + (1-\alpha) e v^T$. By using that $v^T e = 1$ since v is a probability vector in the following derivation, we get:

$$\begin{aligned}
Q^{-1}GQ &= \alpha Q^{-1}PQ + (1-\alpha) Q^{-1} e v^T Q \\
&= \alpha \begin{pmatrix} 1 & y^T PX \\ \mathbf{0} & Y^T PX \end{pmatrix} + (1-\alpha) \begin{pmatrix} y^T \\ Y^T \end{pmatrix} e v^T \begin{pmatrix} e & X \end{pmatrix} \\
&= \alpha \begin{pmatrix} 1 & y^T PX \\ \mathbf{0} & Y^T PX \end{pmatrix} + (1-\alpha) \begin{pmatrix} y^T e \\ Y^T e \end{pmatrix} \begin{pmatrix} v^T e & v^T X \end{pmatrix} \\
&= \alpha \begin{pmatrix} 1 & y^T PX \\ \mathbf{0} & Y^T PX \end{pmatrix} + (1-\alpha) \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} 1 & v^T X \end{pmatrix} \\
&= \begin{pmatrix} \alpha & \alpha y^T PX \\ \mathbf{0} & \alpha Y^T PX \end{pmatrix} + \begin{pmatrix} 1-\alpha & (1-\alpha) v^T X \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\
&= \begin{pmatrix} \alpha + 1 - \alpha & \alpha y^T PX + (1-\alpha) v^T X \\ \mathbf{0} + \mathbf{0} & \alpha Y^T PX + \mathbf{0} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \alpha y^T PX + (1-\alpha) v^T X \\ \mathbf{0} & \alpha Y^T PX \end{pmatrix}
\end{aligned}$$

The derivation shows that we once again obtained a block upper triangular matrix. The characteristic equation is given by

$$\det(Q^{-1}GQ - \lambda I) = \det \begin{pmatrix} 1-\lambda & \alpha y^T PX + (1-\alpha) v^T X \\ \mathbf{0} & \alpha Y^T PX - \lambda I_{N-1} \end{pmatrix} = (1-\lambda) \det(\alpha Y^T PX - \lambda I_{N-1}) = 0.$$

Therefore the first eigenvalue is $\lambda_1 = 1$, meanwhile the remaining are given by $\det(\alpha Y^T PX - \lambda I_{N-1}) = 0$. Since G and $Q^{-1}GQ$ are similar, they have the same eigenvalues. Finally, the eigenvalues $\lambda_2 \dots \lambda_n$ of $Q^{-1}PQ$ are given by the block $Y^T PX$. Multiplying by the damping factor α results in the block containing the eigenvalues of $Q^{-1}GQ$. Therefore, if the eigenvalues of P are $\{1, \lambda_2, \dots, \lambda_n\}$, then the eigenvalues of G are given by $\{1, \alpha\lambda_2, \dots, \alpha\lambda_n\}$. □

By Theorem 2.11, all eigenvalues λ of a stochastic matrix are bounded by 1. Sorting the eigenvalues of matrix P gives us

$$1 = |\lambda_1(P)| \geq |\lambda_2(P)| \geq \cdots \geq |\lambda_N(P)|.$$

Theorem 4.1 states that $\lambda_i(G) = \alpha\lambda_i(P)$ for all $i \geq 2$. Since $\alpha \in (0, 1)$, the absolute value of the second largest eigenvalue is given by

$$|\lambda_2(G)| = |\alpha\lambda_2(P)| = \alpha|\lambda_2(P)|.$$

Now, since $|\lambda_2(P)| \leq 1$ we can derive the following

$$|\lambda_2(G)| = \alpha|\lambda_2(P)| \leq 1 \cdot \alpha = \alpha \quad \text{(multiply by the damping factor on both sides).} \tag{8}$$

Thus, the second largest eigenvalue of the Google matrix is at most α.

4.2 The Spectral Gap

By Definition 2.12, the spectral gap γ is defined as

$$\gamma = 1 - |\lambda_2(G)|.$$

Hence, since $|\lambda_2(G)| \leq \alpha$ by Equation (8) and [8, p. 42], the spectral gap can be expressed as

$$\gamma \geq 1 - \alpha. \tag{9}$$

Since probability distributions are treated as row vectors in this thesis, the power iteration is performed by multiplication from the left:

$$\pi^{(t+1)T} = \pi^{(t)T} G.$$

Therefore we need to use left eigenvectors of G when decomposing the error vector.

To understand the impact of the spectral gap and second largest eigenvalue λ_2 on the rate of convergence, we consider the distance between the true stationary distribution and the distribution after t power iterations. This is the total variation distance expressed in Equation (1): $\|\pi^{(t)T} - \pi^T\|_{TV} = \frac{1}{2}\sum_{i=1}^{N} |\pi_i^{(t)} - \pi_i|$.

Consider the special case and simplification where the Web is seen as a reversible Markov chain. Hence, assume that the Markov chain representing the walk of a random surfer described by matrix G satisfies

$$\pi_i G_{ij} = \pi_j G_{ji}$$

with respect to the stationary distribution. This is in general not the case for PageRank, but is a simplification that will allow us to find a relationship between the the spectral gap and the convergence. Then the following holds:

Theorem 4.2. *(Reversibility: spectral theorem [10]) Let A be the transition matrix of a finite reversible Markov chain with stationary distribution π. Then A has real eigenvalues $1 = \lambda_1, \lambda_2, ..., \lambda_N$. Furthermore, there exists an orthonormal basis of right eigenvectors $\{f_j\}_{j=1}^N$ with corresponding real eigenvalues $\{\lambda_j\}_{j=1}^N$, with respect to the weighted inner product $\langle f, g\rangle_\pi = \sum_{i=1}^N f(i)g(i)\pi_i$.*

Moreover, there also exist an orthonormal basis of left eigenvectors $\{x_i\}_{j=1}^N$ *with respect an weighted inner product* $\langle \cdot, \cdot \rangle_{\pi^{-1}}$*, with the same corresponding real eigenvalues* $\{\lambda_j\}_{j=1}^N$.

Proof. Let D be the diagonal matrix with the stationary distribution π on the diagonal. Thus $D_{ii} = \pi_i$. Define $M := D^{1/2}AD^{-1/2}$. The entries of matrix M can then be expressed:

$$M_{ij} = (D^{1/2}AD^{-1/2})_{ij} = \sqrt{\pi_i}A_{ij}\frac{1}{\sqrt{\pi_j}} = \sqrt{\frac{\pi_i}{\pi_j}}A_{ij}.$$

By reversibility

$$\begin{aligned} M_{ij} &= \sqrt{\frac{\pi_i}{\pi_j}}A_{ij} \\ &= \sqrt{\frac{\pi_j}{\pi_i}}A_{ji} \\ &= (D^{1/2}AD^{-1/2})_{ji} \\ &= M_{ji}. \end{aligned}$$

Thus, matrix M is symmetric. By the Spectral theorem (Theorem 5.1.1 in [10, p. 257]) a symmetric matrix $M \in \mathbb{R}^n$ with elements satisfying $M_{ij} = M_{ji}$, has d eigenvectors with corresponding real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. Hence, M has real eigenvectors $\{\phi\}_{j=1}^N$ forming an orthonormal basis of $\mathbb{R}^n$ with corresponding real eigenvalues $\{\lambda_j\}_{j=1}^N$. Now define the relation between the eigenvectors of matrix A and matrix M as $f_j = D^{-1/2}\phi_j$. Then,

$$\begin{aligned} Af_j &= AD^{-1/2}\phi_j \\ &= D^{-1/2}D^{1/2}AD^{-1/2}\phi_j \\ &= D^{-1/2}M\phi_j \\ &= \lambda_j D^{-1/2}\phi_j \\ &= \lambda_j f_j. \end{aligned}$$

This indicates that $\{f\}_{j=1}^N$ are the right eigenvectors of matrix A and that the real eigenvalues $\{\lambda_j\}_{j=1}^N$ are the same as for matrix M. Now to show that $\{f\}_{j=1}^N$ is an orthonormal basis for $\mathbb{R}^N$ with respect to the stationary distribution π, the weighted scalar product is considered:

$$\langle f_i, f_j \rangle_\pi = \sum_{k=1}^{N} f_i(k)f_j(k)\pi_k.$$

Now since the components can be expressed

$$f_j(k) = \frac{\phi_j(k)}{\sqrt{\pi_k}}$$

we get that

$$\begin{aligned}\langle f_i, f_j\rangle_\pi &= \sum_{k=1}^{N} f_i(k) f_j(k) \pi_k \\ &= \sum_{k=1}^{N} \left(\frac{\phi_i(k)}{\sqrt{\pi_k}}\right)\left(\frac{\phi_j(k)}{\sqrt{\pi_k}}\right)\pi_k \\ &= \sum_{k=1}^{N} \frac{\phi_i(k)\phi_j(k)}{\pi_k}\pi_k \\ &= \sum_{k=1}^{N} \phi_i(k)\phi_j(k) \\ &= \langle \phi_i, \phi_j\rangle.\end{aligned}$$

Since $\{\phi_j\}_{j=1}^N$ is an orthonormal basis of $\mathbb{R}^N$ we have that $\{f_j\}_{j=1}^N$ is an orthonormal basis in the weighted space $(\mathbb{R}^N, \langle\cdot,\cdot\rangle_\pi)$ [10, pp. 257-259, pp. 278-279].

To show the corresponding result for left eigenvalues of A, we express the left eigenvectors in terms of the right eigenvector, using the diagonal matrix D with the stationary distribution π on the diagonal. Define the row vector

$$x_j^T = f_j^T D \quad \text{(or equivalently: } x_j(i) = \pi_i f_j(i)\text{)}.$$

We now show that the vector x_j^T is a left eigenvector of A corresponding to the eigenvalue λ_j. Hence it needs to satisfy that $x_j^T A = \lambda_j x_j^T$:

$$x_j^T A = (f_j^T D)A = f_j^T (DA).$$

According to [1], a reversible Markov matrix the detailed balance $\pi_i A_{ij} = \pi_j A_{ji}$ can be written in matrix form as $DA = A^T D$. This yields:

$$x_j^T A = f_j^T (DA) = f_j^T (A^T D) = (Af_j)^T D.$$

Since f_j is a right eigenvector, it satisfies $Af_j = \lambda_j f_j$. Substitution gives us:

$$x_j^T A = (Af_j)^T D = (\lambda_j f_j)^T D = \lambda_j (f_j^T D) = \lambda_j x_j^T.$$

This shows that x_j^T is a left eigenvector with the same corresponding eigenvalue λ_j as the right eigenvector f_j [1, p. 2-3]. Finally, since the right eigenvectors f_j are orthonormal with respect to the stationary distribution π, it is possible to think that the left eigenvectors would be orthonormal with respect to the reciprocal π^{-1} based on the notation the representation of the inner product in [9, Section 12.1]. We express the inner product as:

$$\langle x_i, x_j\rangle_{\pi^{-1}} = \sum_{k=1}^{N} \frac{x_i(k) x_j(k)}{\pi_k}.$$

Then the following would hold when the left eigenvectors are defined as $x_j(i) = \pi_i f_j(i)$, making $\{x_j\}_{j=1}^N$ an orthonormal basis withe respect to π^{-1}:

$$\langle x_i, x_j \rangle_{\pi^{-1}} = \sum_{k=1}^{N} \frac{x_i(k)x_j(k)}{\pi_k} = \sum_{k=1}^{N} \frac{\pi_k f_i(k) \pi_k f_j(k)}{\pi_k} = \sum_{k=1}^{N} \pi_k f_i(k) f_j(k) = \langle f_i, f_j \rangle_{\pi}.$$

This completes the proof and shows that $\{x_j\}_{j=1}^N$ is an orthonormal basis with respect to $\langle \cdot, \cdot \rangle_{\pi^{-1}}$. □

This theorem will soon be applied with regards to the left eigenvectors to the special case using a reversible Markov chain corresponding to the Google matrix G. Now, to find an expression for the distance, we let $x_1^T, ..., x_N^T$ represent the left eigenvectors corresponding to the eigenvalues $\lambda_1, ..., \lambda_N$ of the Google matrix G. Thus,

$$x_j^T G = \lambda_j x_j^T.$$

By Theorem 3.4 the PageRank vector π^T is the unique stationary distribution satisfying $\pi^T = \pi^T G$. Hence, π^T is the left eigenvector of G corresponding to the eigenvalue 1. Moreover, since G is row stochastic, the column vector e satisfies

$$Ge = e.$$

Thus, e is a right eigenvector corresponding to the eigenvalue 1.

The difference between the distribution at iteration t and the true stationary distribution can be expressed

$$\pi^{(t+1)T} - \pi^T = \pi^{(t)T} G - \pi^T G = (\pi^{(t)T} - \pi^T) G.$$

By denoting the difference between the distributions $\pi^{(t)T} - \pi^T$ as $\delta^{(t)T}$, the following expression has been obtained: $\delta^{(t+1)T} = \delta^{(t)T} G$. By considering the case where PageRank is simplified to a reversible Markov chain, the initial error can be expanded in a basis formed by eigenvectors by Theorem 4.2. By Theorem 5.1.1 in [10, p. 257] we are also ensured that the Google matrix is diagonalizable with real eigenvalues under the simplification that it is reversible. Thus, by the definition of an left eigenbasis, there exists a basis of linearly independent left eigenvectors $x_1^T, ..., x_N^T$. We choose to work with a basis of left eigenvectors as the probability distributions are multiplied from the left in the PageRank algorithm. By the definition of a basis, every vector in $\mathbb{R}^N$ can be uniquely expressed as a linear combination of the vectors of the basis set [3][5, p. 43][12, Section 3.1]. Hence by assuming that G is diagonalizable and that $x_1^T, ..., x_N^T$ form a basis of left eigenvectors, we can express any initial distribution $\pi^{(0)T}$ as:

$$\pi^{(0)T} = c_1 x_1^T + c_2 x_2^T + ... + c_N x_N^T. \tag{10}$$

Let $x_1^T = \pi^T$, meaning that the first left eigenvector is the stationary vector. Since left eigenvectors are defined as $x_j(i) = \pi_i f_j(i)$ in the reversibility case, this corresponds to choosing the first right eigenvector as $f_1 = e$ [6]. By Theorem 4.2, left eigenvectors are orthonormal with respect to the inner product $\langle \cdot, \cdot \rangle_{\pi^{-1}}$ and not the usual inner product [9, p. 162]. Hence, based on [10,

Section 5.2, p. 276] the weighted $l_2 -$ norm would be

$$\|x\|^2_{\pi^{-1}} = \sum_{i=1}^{N} \frac{x_i^2}{\pi_i}.$$

For $x_1 = \pi$ we would then get

$$\|x_1\|^2_{\pi^{-1}} = \|\pi\|^2_{\pi^{-1}} = \sum_{i=1}^{N} \frac{\pi_i^2}{\pi_i} = \sum_{i=1}^{N} \pi_i = 1.$$

Hence, $x_1 = \pi$ is normalized in the weighted norm corresponding to the left eigenbasis. Since $\pi^{(0)T}$ and π^T are probability distributions, their entries sum to 1. Hence

$$\pi^{(0)T} e = 1 \quad \text{and} \quad x_1^T e = \pi^T e = 1.$$

For the eigenvectors $x_2^T, x_3^T, ..., x_N^T$, where $\lambda_j \neq 1$, we have that for $j \geq 2$

$$x_j^T G = \lambda_j x_j^T.$$

Multiplying both sides of the equation by e gives

$$x_j^T G e = \lambda_j x_j^T e.$$

Substitution of $Ge = e$ gives

$$x_j^T e = \lambda_j x_j^T e.$$

Rearranging gives

$$\begin{aligned} x_j^T e - \lambda_j x_j^T e &= 0 \\ (1 - \lambda_j) x_j^T e &= 0. \end{aligned}$$

Since $\lambda_j \neq 1$ for $j \geq 2$ it follows that $x_j^T e = 0$. Substitution gives us

$$\pi^{(0)T} e = c_1 x_1^T e + c_2 x_2^T e + ... + c_N x_N^T e.$$

Since $\pi^{(0)T} e = 1$, $x_1^T e = \pi^T e = 1$, and $x_j^T e = 0$ for $j \geq 2$ it follows that $c_1 = 1$ so that

$$\pi^{(0)T} = x_1^T + c_2 x_2^T + ... + c_N x_N^T.$$

After t iterations the following is obtained

$$\begin{aligned}
\pi^{(t)T} &= \pi^{(0)T} G^t \\
&= (x_1^T + c_2 x_2^T + ... + c_N x_N^T) G^t \\
&= (1 \cdot x_1^T + c_2 x_2^T + ... + c_N x_N^T) G^t \\
&= x_1^T G^t + c_2 x_2^T G^t + ... + c_N x_N^T G^t \\
&= 1^t x_1^T + c_2 \lambda_2^t x_2^T + ... + c_N \lambda_N^t x_N^T \quad (\text{since} \quad x_j^T G^t = \lambda_j^t x_j^T) \\
&= \pi^T + c_2 \lambda_2^t x_2^T + ... + c_N \lambda_N^t x_N^T \quad (\text{since} \quad x_1^T = \pi^T).
\end{aligned}$$

Now the difference (or error) is given by

$$\delta^{(t)T} = \pi^{(t)T} - \pi^T = c_2 \lambda_2^t x_2^T + ... + c_N \lambda_N^t x_N^T. \tag{11}$$

As $t \to \infty$, all $\lambda_i^t \to 0$ for $i \geq 2$ since $1 = |\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq ...$. However, since λ_2 is the eigenvalue of largest magnitude for all $i \geq 2$ it will have the slowest decrease and therefore $|\lambda_2|^t$ will dominate the impact on the rate of convergence. Finally, since $|\lambda_2| \leq \alpha$ by Equation (8), it follows that the error $\delta^{(t)T}$ is bounded by a constant times α^t.

4.3 Rate of Convergence

In the previous section, it was concluded that the difference between the distribution at iteration t and the true stationary distribution expressed in Equation (11) could be written as

$$\pi^{(t)T} - \pi^T = \sum_{j=2}^{n} c_j \lambda_j^t x_j. \tag{12}$$

To obtain an upper bound of the error at a specific node, the absolute value is applied to both sides of the expression. By the triangle inequality the following expression is obtained:

$$\begin{aligned}
|\pi^{(t)T} - \pi^T| &= |c_2 \lambda_2^t x_2 + c_3 \lambda_3^t x_3 + \cdots + c_n \lambda_n^t x_n| \\
&\leq |c_2 \lambda_2^t x_2| + |c_3 \lambda_3^t x_3| + \cdots + |c_n \lambda_n^t x_n|.
\end{aligned}$$

Since $|\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$ the upper bound is given by factoring out the term $|\lambda_2|^t$. Thus,

$$|\pi^{(t)T} - \pi^T| \leq |\lambda_2|^t \cdot \left(|c_2 x_2| + |c_3 x_3| \cdot \left| \frac{\lambda_3}{\lambda_2} \right|^t + \cdots + |c_n x_n| \cdot \left| \frac{\lambda_n}{\lambda_2} \right|^t \right)$$

where all the quotients $\left| \frac{\lambda_j}{\lambda_2} \right|^t \leq 1$ for $j = 3, ..., n$. To create an upper bound we consider the "worst case" where the quotients equal 1. This leads to the expression:

$$|\pi^{(t)T} - \pi^T| \leq |\lambda_2|^t \cdot \Big(|c_2 x_2| + |c_3 x_3| + \cdots + |c_n x_n| \Big). \tag{13}$$

Note that this compares well to the exponential bound given by the convergence Theorem 2.7:

$$|\pi^{(t)T} - \pi^T| \leq |\lambda_2|^t \cdot C$$

where $C = |c_2 x_2| + |c_3 x_3| + \cdots + |c_n x_n|$ is a constant. Hence, it is not necessary to know all eigenvalues for a graph of Google's magnitude in order to determine an upper bound for the convergence since the spectral gap $\gamma = 1 - |\lambda_2|^t$ determines how fast PageRank converges.

To analyze the performance for the whole network, let the total variation distance between the distribution at iteration t and the true stationary distribution be defined according to Definition 2.8 and Equation (2). That is

$$d(t) = \max_{\pi^{(0)T}} \|\pi^{(t)T} - \pi^T\|_{TV}$$

where the maximum is taken over all possible initial distributions $\pi^{(0)T}$. By Equation (1) this is equivalent to half of the L^1 norm of the difference vector (sum of the absolute values):

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_i |\mu_i - \nu_i|.$$

This vector is the difference between the distribution at iteration t and the stationary distribution. Thus,

$$\begin{aligned} \|\pi^{(t)T} - \pi^T\|_{TV} &= \frac{1}{2} \|\pi^{(t)T} - \pi^T\|_1 \\ &= \frac{1}{2} \Big\| \sum_{j=2}^{N} c_j \lambda_j^t x_j^T \Big\|_1 \\ &\leq \frac{1}{2} \sum_{j=2}^{N} |c_j| |\lambda_j|^t \|x_j^T\|_1. \end{aligned}$$

By factoring out $|\lambda_2|^t$ it is possible to see that just like before, the total variation distance is proportional to the sum of these absolute values. It follows that it is bounded by $C|\lambda_2|^t$ where $C = \frac{1}{2} \sum_{j=2}^{N} |c_j| |x_j^T|_1$. Hence, since the $l_1 -$ norms do not depend on t, it gets absorbed into the constant C and $|\lambda_2|^t$ determines the exponential rate. Finally, since $|\lambda_2| \leq \alpha$ the total variation distance is bounded by $C\alpha^t$. This highlights the relationship between the second largest eigenvalue, as the PageRank algorithm will converge to the stationary distribution at least as fast as the damping factor α.

4.4 The Bounds of the Mixing Time

In order to compute PageRank, it is important for Google to know that the algorithm converges within a reasonable amount of time, independently of the size N of the Web. Meanwhile the spectral gap tells us about the rate of convergence, the mixing time $t_{\text{mix}}(\varepsilon)$ can provide a theoretical measure of how many iterations are required before the distribution at iteration t is guaranteed to be within a distance ε of the stationary distribution.

We have established that the total variation distance is decreasing at a rate bounded by the

damping factor α^t. By Definition 2.9 mixing time is defined as the smallest t such that the distance to stationarity is less than or equal to the threshold ε. Hence, we can express the inequality

$$d(t) = \|\pi^{(t)T} - \pi^T\|_{TV} \leq C\alpha^t \leq \varepsilon.$$

Solving for the number of iterations t and satisfying the above inequality, gives us:

$$\begin{aligned} t\log(\alpha) &\leq \log(\frac{\varepsilon}{C}) \\ t &\geq \frac{\log(\varepsilon/C)}{\log(\alpha)}. \end{aligned}$$

Thus, at time t we are guaranteed to have arrived within ε of the stationary distribution. The actual mixing time is quicker or equal to this value. Hence,

$$t_{mix} \leq \frac{\log(\varepsilon/C)}{\log(\alpha)}.$$

This is an important result, as it shows that the mixing time depends on the damping factor/second eigenvalue. The constant C is independent of t but the exact value depends indirectly on the size of the internet, since we sum from 2 to N, and on the initial distribution $\pi^{(0)T}$. However, after a sufficient amount of iterations, the total variation distance is dominated by the decay of $|\lambda_2|^t \leq \alpha^t$.

To provide a more rigorous bound, accounting for both the spectral gap and the initial distribution, we refer to the proof of Theorem 12.4 in *Markov Chains and Mixing Time* by Levin and Peres [9, pp. 163-164]. This provides more intuition as we do not examine the details of what the constant C represents.

Theorem 4.3. *Let P be the transition matrix of a reversible, irreducible Markov chain with state space $\mathcal{X}$, and let $\pi_{\min} := \min_{x\in\mathcal{X}} \pi(x)$. Then the mixing time would be bound such that*

$$(t_{rel} - 1)\log(\frac{1}{2\varepsilon}) \leq t_{mix}(\varepsilon) \leq t_{rel}\log\left(\frac{1}{\varepsilon\pi_{\min}}\right)$$

where the relaxation time for a reversible Markov chain with spectral gap $\gamma = 1 - \lambda_2$ is defined

$$t_{rel} := \frac{1}{1-\lambda_2}.$$

Proof. Levin and Peres proves the theorems for the upper and lower bounds in *Markov Chains and Mixing Times* on pages 163-164, by decomposing the transition matrix, using the Cauchy-Schwarz inequality and the orthonormality of $\{x_j\}$. □

Finally, it is important to note that while the damping factor provides a worst case scenario for the upper bound, it is possible that the actual second eigenvalue is smaller than the damping factor. As a result, the rate of convergence and mixing time would be more rapid than the theoretical upper bound suggests. The only thing we know is that $\lambda_2(G) \leq \alpha$ due to the teleportation property.

5 Example Graphs

To illustrate how the mixing time of the power method PageRank algorithm behaves on different graph structures, we consider two simple examples: a cycle graph and a wheel graph. Generally, PageRank is computed for directed networks, and hence in reality each undirected edge in Figure 1 and Figure 2 would be interpreted as a pair of directed edges in opposite directions. While the Google matrix is not reversible in reality, we assume reversibility for the simple example graph structures to apply bounds of mixing time, where it is necessary to assume reversibility of the Markov chain. The purpose of these two examples is to highlight how the mixing time for PageRank is affected by the second largest eigenvalue for different link structures.

5.1 Cycle

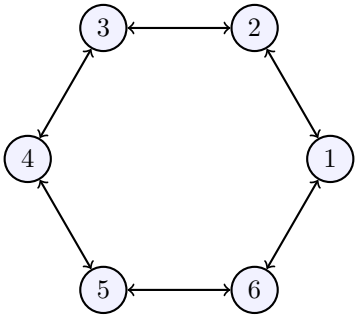

Figure 1: A cycle graph C consisting of six nodes connected by a pair of directed edges in opposite directions to ensure the reversibility of the Markov chain.

In Figure 1 the cycle graph is exemplified with 6 nodes in a ring. For the transition matrix P, let d_i denote the degree for node i and assume that $P_{ij} = \frac{1}{d_i}$. Then matrix P would be:

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

Assuming uniform teleportation vector v^T, the Google matrix has the form

$$G_C = \alpha P_C + (1-\alpha)\frac{1}{N} e e^T.$$

By Theorem 4.1 the set of eigenvalues for the Google matrix are the eigenvalues of matrix P scaled by the damping factor α. Thus, the second largest eigenvalue is given by $\lambda_2(G) = \alpha\lambda_2(P)$. According to [9, p. 166], the eigenvalues in the spectrum for an undirected n-cycle with transition

probability of 1/2 to each neighbor of a node, is given by

$$\lambda_j = \frac{\omega^j + \omega^{-j}}{2} = \cos\left(\frac{2\pi j}{N}\right) \quad j = 0, \dots N-1.$$

Therefore, the second largest eigenvalue (corresponding to $j = 1$) of matrix P can be expressed

$$\lambda_2(P_C) = \cos\left(\frac{2\pi}{N}\right)$$

and hence letting the damping factor take the value suggested by Page and Brin, $\alpha = 0.85$ results in the following second largest eigenvalue [2]:

$$\lambda_2(G_C) = \alpha\lambda_2(P_C) = \alpha\cos\left(\frac{2\pi}{N}\right) = 0.85\cos\left(\frac{2\pi}{6}\right) = 0.85\cos\left(\frac{\pi}{3}\right) = 0.425.$$

Since the cycle graph is regular, thus all nodes have the same degree, $\pi_{\min} = \frac{1}{N}$. Let $\varepsilon = 0.01$. Then the upper bound for the mixing time would be given by

$$t_{mix}(G_C) \leq \frac{1}{1-\lambda_2(G_C)}\log(\frac{N}{\varepsilon}) = \frac{1}{1-0.85\cos\left(\frac{\pi}{3}\right)}\log(\frac{6}{0.01}) \approx 4.83.$$

Hence, it takes maximum 4.83 iterations before the distribution is within a distance of $\varepsilon = 0.01$ from the stationary distribution.

5.2 Wheel

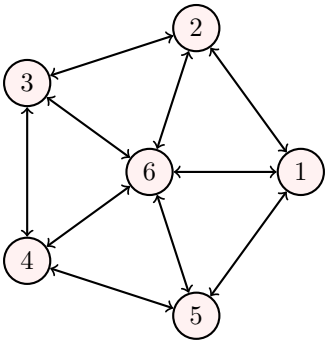

Figure 2: The plot shows a wheel graph W consisting of an outer cycle and a central hub node, connected by a pair of directed edges in opposite directions to ensure the reversibility of the Markov chain.

Define an undirected wheel graph consisting of an outer cycle of $N-1$ nodes and a central node in the middle, see Figure 2. For the graph where $N = 6$, the central node has degree $d_{center} = 5$, meanwhile the outer nodes have degree $d_i = 3 \quad i = 1, \dots, 5$. Assuming the transition probability is

uniformly distributed on the outgoing links, the following initial transition matrix P_W is obtained:

$$P_W = \begin{pmatrix} 0 & 1/3 & 0 & 0 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 0 & 0 & 1/3 & 0 & 1/3 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 0 \end{pmatrix}.$$

Since the node degree in the wheel graph is not uniform, the graph is non-regular. The stationary distribution is proportional to the degree and can be expressed similarly to [9, pp. 9-10]:

$$\pi_i = \frac{d_i}{\sum_k d_k}.$$

For the outer nodes the stationary distribution $\pi_i \in \{1, ..., 5\}$ would be $\pi_i = \frac{3}{3\cdot 5+5} = \frac{3}{20} < \frac{5}{3\cdot 5+5} = \frac{5}{20} = \pi_6$. Therefore, $\pi_{min} = \frac{3}{20}$ would correspond to one of the outer nodes as the central node has a higher degree. Furthermore, since

$$\pi_i \cdot P_{i,6} = \frac{3}{20} \cdot \frac{1}{3} = \frac{1}{20} = \frac{5}{20} \cdot \frac{1}{5} = \pi_6 \cdot P_{6,i}$$

the graph satisfies $\pi_i P_{ij} = \pi_j P_{ji}$, which makes P_W reversible despite not being symmetric.

Solving the characteristic equation for P_W to find the second largest eigenvalues $\lambda_2(G_W) = \alpha\lambda_2(P_W)$ is computationally tedious by hand. However, intuition suggests that the wheel graph is more connected as the central node enables a random surfer to get from one side of the graph to the other in fewer steps than in the cycle graph. For any initial state, it is possible to go to any other state in only two steps by passing the central node. It follows that $\lambda_2(G_W) < \lambda_2(G_C)$, resulting in that the spectral gap $1-\lambda_2(G_W) > 1-\lambda_2(G_C)$. Although $\pi_{min}(P_C) = 1/6 \approx 0.1667 > 0.150 = 3/20 = \pi_{min}(P_W)$, the increase in the logarithmic term $\log(\frac{1}{\varepsilon\pi_{min}})$ is neglected compared to the dominating inverse spectral gap. Consequently, the mixing time is shorter for the wheel graph than for the cycle, leading to the conclusion that the Markov chain converges quicker to the stationary distribution. Hence,

$$t_{mix}(G_W) \leq \frac{1}{1-\lambda_2(G_W)} \log(\frac{1}{\varepsilon\pi_{\min}}) < t_{mix}(G_C).$$

Finally, when considering the scalability of these small example graphs, we can conclude that since the second eigenvalue of the cycle is given by $\lambda_2(G_C) = \alpha \cos\left(\frac{2\pi}{N}\right)$, an increasing number of nodes N would imply that $\cos\left(\frac{2\pi}{N}\right) \to 1$. This means that the second eigenvalue becomes larger, leading to a smaller spectral gap and in turn slower mixing for the random walk. As a result, the damping factor would be of importance for large cycles as the second eigenvalue $|\lambda_2(G)| \leq \alpha$ due to the teleportation property. In comparison, the mixing time for the wheel would not be as affected on a large scale, as it already has a central node connecting the different parts of the graph. Thus as N grows, it already has a way for the random surfer to move quickly in the graph structure. As a result, we could expect that $t_{mix}(G_W) < t_{mix}(G_C)$. To be able to state this more rigorously, we

would need to compute or find a bound for the second eigenvalue of the wheel graph. However, this argument gives some intuition for what would happen if the graphs were scaled by N.

6 Discussion

6.1 Discussion of Results

The purpose of this thesis was to study the PageRank algorithm from the perspective of Markov chains, with focus on stationary distributions, spectral properties, convergence and mixing time. As examined, the Google matrix is an altered version of the link matrix where dangling nodes are handled and teleportation is implemented. It can be expressed

$$G = \alpha P + (1 - \alpha)ev^T$$

where $0 < \alpha < 1$ and $v_i > 0$. This results in an irreducible and aperiodic Markov chain ensuring the existence and uniqueness of the stationary distribution PageRank. The analysis shows that if the spectrum of the stochastic matrix P is given by $\{1, \lambda_2, \ldots, \lambda_n\}$, then the spectrum for the Google matrix is given by $\{1, \alpha\lambda_2, \ldots, \alpha\lambda_n\}$. To tie back to the initial research questions, based on the analysis conducted in this thesis, the conclusion can be drawn that there certainly is a relationship between the damping factor and the second largest eigenvalue of the Google matrix, such that $|\lambda_2| \leq \alpha$. In the worst case scenario, this results in that the difference between the distribution at power iteration t and the stationary distribution is governed by α^t, as it is the upper bound of $|\lambda_2|^t$ due to the teleportation properties.

Some questions that arise are "How do we know that PageRank actually works?" and "What happens if C from Equation (13) is so large that the algorithm in practice takes thousands of steps to converge, despite a beneficial spectral gap?". In such a scenario, the PageRank algorithm would be computationally expensive for search engines. However, it has been proved that $|\lambda_2| \leq \alpha$. Therefore, the "worst case scenario" is that the error decays proportionally to α^t. Page and Brin suggested in their initial paper that $\alpha = 0.85$ is a standard value for the damping factor[2]. For $t = 50$ iterations the error would be: $\alpha^{50} = 0.85^{50} \approx 0.0003$ which efficiently decreases the constant C. The PageRank vector represents a probability distribution where all the entries sum up to 1. The total error is spread across approximately $8 \cdot 10^9$ coordinates, implying that the error per website is small when applied to the internet. This allows us to conclude that although we have yet to find an upper bound for C in this report, we can intuitively explain that $t = 50$ iterations is considered enough as stated by Page and Brin[2]. The teleportation mechanism ensures scalability as the average error is dominated by α^t, so that the rate of convergence only depends on the initial distribution through logarithmic terms, as seen in Theorem 4.3.

After concluding that the power method reduces the error efficiently after a few t iterations, another question becomes relevant: "Why not increase the damping factor to 0.95 instead?". Such an increase would mean that the random surfer follows the link structure 95% of the time, meaning that the underlying graph structure is more significant, but it comes at a great computational cost. After 50 iterations the total error would be $0.95^{50} \approx 0.0769$. Hence, to obtain as high precision or lack of total error as when using $\alpha = 0.85$, it would require over $t = 150$ iterations as $0.95^{150} \approx 0.00045$. This explains why 0.85 is a realistic "sweet spot", and why $t = 50$ iterations provides good accuracy for a constant C of realistic magnitude.

As seen for the simple example graphs in Figure 1 and Figure 2, the convergence to the sta-

tionary distribution differs depending on the underlying graph structure. The wheel graph added a central hub node connecting opposite sides of the outer cycle, increasing the spectral gap by decreasing the second largest eigen value compared to the uniform cycle graph. Consequently, the wheel had lower mixing time as a result of higher connectivity. The central hub node can intuitively be compared to the teleportation properties of the PageRank algorithm as it connects different parts of the graph, highlighting the importance of the damping factor when reducing the number of iterations needed to converge to a stationary distribution. Thus, despite that the size N of the Web is large, the convergence remains rapid due to the spectral gap and the teleportation property of the random surfer. In reality, the mixing time could be shorter than the upper bound suggests as $|\lambda_2| \leq \alpha$.

6.2 Limitations and Improvements

Several simplifications were applied in order to provide a theoretical understanding for the PageRank algorithm. The greatest limitation of this report is the assumption that the Markov chain is reversible. It was a necessary assumption to applying well known theorems for the upper bound of the mixing time to the example graphs. This assumption does not reflect reality as the internet is asymmetric with directed edges. Hence, in reality there is seldom a link pointing back to the source linking to it. An example is that smaller websites might point to a big well known source without a link pointing back from the bigger source. Consequently, the Google matrix is in reality non-reversible. However, in order to simplify the computations leading to the conclusions about the spectral properties that could be linked to the damping factor, it was a necessary assumption. To improve the theoretical framework, it would have been interesting to further explore the case where the Google matrix is not considered reversible, as it better applies to reality.

6.3 Further Outlook

A further limitation is that the teleportation vector v^T is uniform so that the random surfer jumps to a new page with equal probability. In modern applications of PageRank, the teleportation vector is often chosen non-uniformly, as a surfer often is more likely to have preferences impacting where on the internet they jump. This is called personalized PageRank and would have been interesting to dive deeper into as an extension as it impacts the stationary distribution.

Furthermore, it would have been relevant to study the case where the graphs are non-reversible and directed as they are better representations of the internet. This could be done through applying theory to small datasets instead of simple symmetric graphs as performed in this report.

6.4 Conclusion

In conclusion, the damping factor α is the key factor impacting the rate of convergence and mixing time of the PageRank algorithm. This is because the teleportation property guarantees that the corresponding Markov chain is aperiodic and irreducible so that the algorithm converges to a unique PageRank vector. The second largest eigenvalue $|\lambda_2(G)|^t \leq \alpha^t$ controls the rate of convergence and mixing time. Moreover, the simple example graphs show that the underlying graph structure

also matters as higher connectivity leads to faster mixing, a concept that could be compared to making the internet more connected through teleportation.

References

[1] Ellen Baake, Michael Baake, and Jeremy Sumner. Embedding of reversible Markov matrices, 2025. URL: https://arxiv.org/abs/2511.22800, arXiv:2511.22800.

[2] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998. doi:10.1016/S0169-7552(98)00110-X.

[3] Stephen Cass. How Google finds your needle in the web's haystack. AMS Feature Column, January 2006. Accessed: 2026-03-10. URL: https://www.ams.org/publicoutreach/feature-column/fcarc-pagerank.

[4] Lars Eldén. A note on the eigenvalues of the Google matrix, 2004. URL: https://arxiv.org/abs/math/0401177, arXiv:math/0401177.

[5] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence. *Linear Algebra*. Pearson, 4th edition, 2014.

[6] Taher H. Haveliwala and Sepandar D. Kamvar. The second eigenvalue of the Google matrix. Technical Report 2003-20, Stanford InfoLab, Stanford University, 2003. URL: https://ilpubs.stanford.edu/582/1/2003-20.pdf.

[7] Amy N. Langville and Carl D. Meyer. Deeper inside PageRank. *Internet Mathematics*, 1(3):335–380, 2004. doi:10.1080/15427951.2004.10129091.

[8] Amy N. Langville and Carl D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006. E-book, accessed 2026-03-15. doi:10.2307/j.ctt7t8z9.

[9] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society (AMS), 2nd edition, 2017.

[10] Sébastien Roch. *Modern Discrete Probability: An Essential Toolkit*, chapter 5. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2024.

[11] Sheldon M. Ross. *Introduction to Probability Models*. Elsevier, 13th edition, 2024.

[12] Perla Sousi. Mixing times of Markov chains, 2020. Lecture notes, University of Cambridge. URL: https://www.math.ubc.ca/~jhermon/Mixing/mixing-notes.pdf.