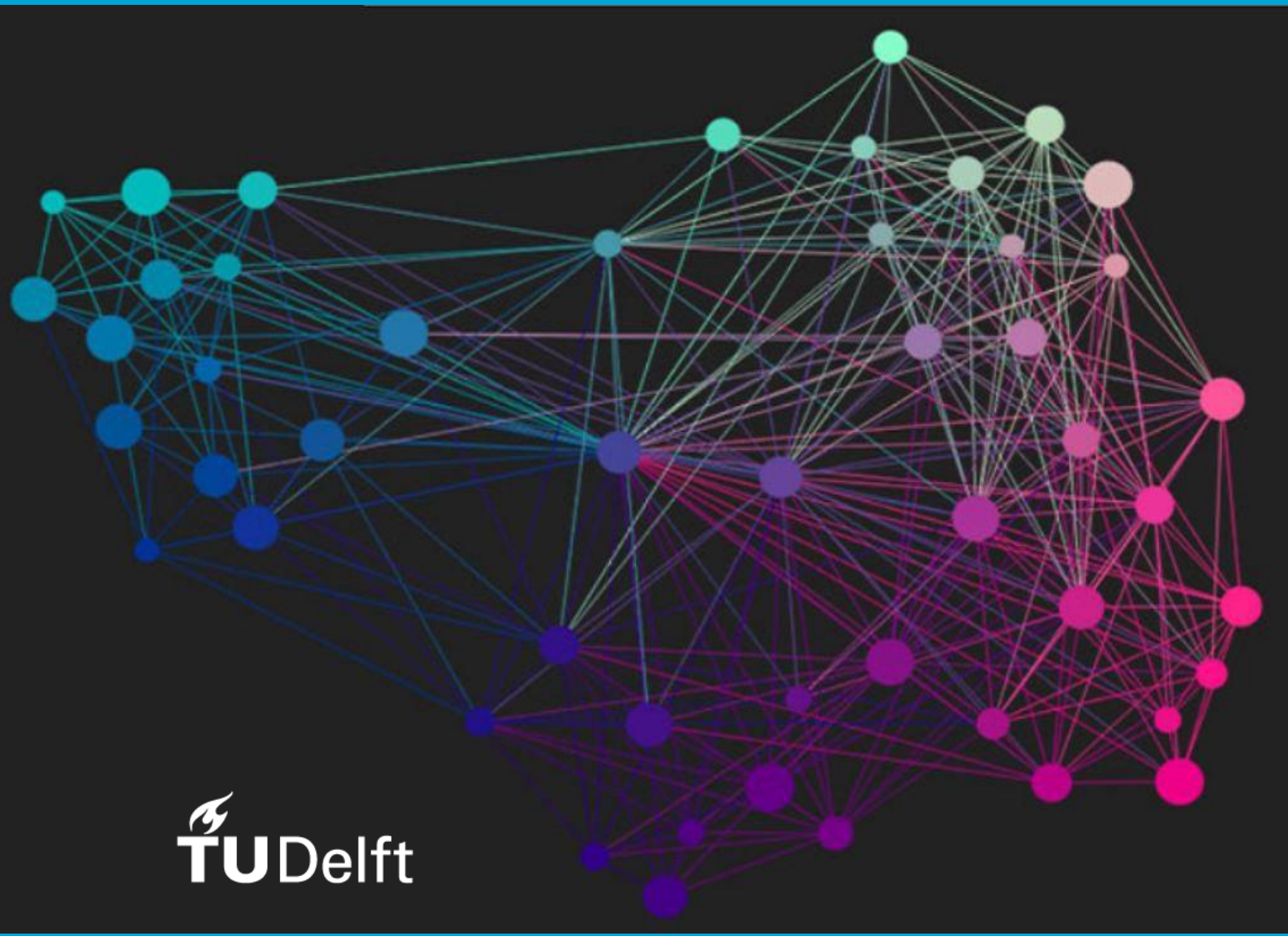
BSc Thesis in Applied Mathematics

Directed graph Laplacian applied on spectral clustering

Y.S. Şavlı

2021



TU Delft

DIRECTED GRAPH LAPLACIAN APPLIED ON SPECTRAL CLUSTERING

A thesis submitted to Delft University of Technology in partial fulfilment of the requirements for the degree of

Bachelor of Science in Applied Mathematics

by

Yasemin Simge Şavlı

August 2021

Student number:	4159403	
Thesis committee:	Dr. Y. van Gennip ,	TU Delft, supervisor
	Dr. H.N. Kekkonen,	TU Delft

TU Delft

PREFACE

There it is, it feels relieving to say: 'this is it, the final chapter of a journey I had during my studies in Delft'! It is both a final chapter, but also the beginning of a new chapter.

The report you are reading, was written for the fulfilment of the Bachelors degree in Applied Mathematics at Delft University of Technology, for the project named BEP (Bachelor End Project). The BEP is an overarching project, being the pinnacle of the Bachelor and as such the report comprises different subjects of the study Applied Mathematics.

With a global pandemic during the proceedings of this report, it was a difficult time for many of us. Not being able to study on-campus, to discuss with fellow students or with academic staff face-to-face and many more. However, with the tremendous help I had from my supervisor, Dr. Yves van Gennip this report became a reality. Therefore, first and foremost, I would like to thank him for his inspirational guidance and perseverance, making sure I always stayed on target.

Furthermore, I would like to thank my family and especially my parents and my husband for their support during my Bachelors. They have always been there for me and having them aside me gave me strength in pursuing further.
My gratitude also extends to my friends, with which I had an amazing time during my student life in Delft. All the nice experiences we have had together I will cherish with me.

Lastly, I would to like to thank all the academic staff at TU Delft, with a special notice for Nadia Palliser who has given me inspiration in the final stretch. Thank you for always believing in me and I am looking forward to the journey ahead!

Yasemin S. Şavlı
Delft, August 2021

ABSTRACT

The graph Laplacian is a tool which is commonly used in different applications, amongst which spectral clustering. This report contains a research in different definitions for the graph Laplacian applied on directed graphs, since it is rarely used in applications neither occurs often in literature. Two definitions are discussed into more detail.

The first definition considers a directed graph as a bipartite graph between the node set containing all nodes with outgoing edges and the node set containing all nodes with incoming edges. A Laplacian is defined on both sets distinctly and later on the two Laplacians are convexly combined to define a Laplacian on the whole directed graph.

The second definition considers a random walk on a directed graph. We will see why a random walk on a directed graph is equivalent to a finite Markov Chain with a corresponding transition probability matrix. This definition will be used for clustering.

The method for clustering as well as the definition for the directed graph Laplacian require a unique stationary distribution associated with the transition probabilities on the node set of the graph. The existence and uniqueness of the stationary distribution became an important part of this report, because the given data set has to meet up with these properties to be able to use this method.

Furthermore, the introduced definitions are used in a MATLAB based model, to get more insights about the spectrum of the directed graph Laplacian. The biggest conclusions to be drawn are that for strongly connected directed graphs, it seems that there is a correlation between the multiplicity of eigenvalues that are close to 1 and the numbers of subsets such that all nodes in the subset are adjacent.

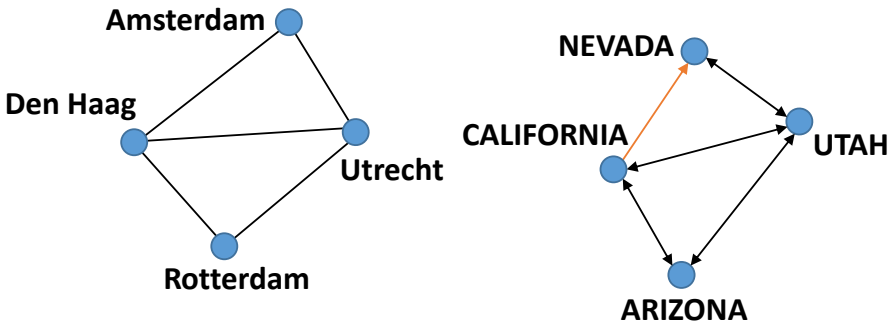
CONTENTS

1 INTRODUCTION

What do machine learning, the Poisson equation, heat transport phenomena, electrostatics and waves all have in common? They all share a common use of the Laplacian operator, or shortly the Laplacian, in partial differential equations. In discrete mathematics the graph Laplacian is usually a matrix representation of a graph, as often used in graph theory. If you are totally new to graph theory, you can see a brief description of the basic concepts used for this report in Chapter 3. In Chapter 2 the graph Laplacian will be clarified.

A graph is a way of visualising the dependencies in between different objects in the form of structures that are made up by a set of *nodes* and *edges*. Often the depiction is done with dots for nodes and lines for edges of a undirected graph and arcs for edges of a directed graph. A graph can be weighted and can be either directed or undirected. These concepts are explained in the following example.

To visualise this concept, we can use highway connections in between four states in the southwestern corner of the United States of America, being Arizona, California, Nevada and Utah, whereby the cities are nodes and the highway connections are represented by the edges of the graph, as depicted in Figure 1.1a.
The weight of the edge can be interpreted as the number of highway lanes connecting the states of the US, the higher the number of lanes, the higher the weight. The concept of undirected and directed graphs can also be made clear with this example, as the graph earlier referenced is undirected, i.e. drivers can go in both directions. If however, due to maintenance the highway from California to Utah would be closed, only a one-way high-way from California to Nevada would exist, as depicted by the orange edge in Figure 1.1b.



(a) Undirected graph. **(b)** Directed graph.

Figure 1.1: Example of high-ways between four states in the United States of America represented as a graphs

To see whether there are states that can be clustered into groups, we can use the graph Laplacian. Research into the graph Laplacian is limited to the undirected graph Laplacian (meaning the edges have the same weight in both ways), as proves sufficient in many applications.

The aim of this report is to research what kind of definitions are given in literature for the graph Laplacian on directed graphs. The main articles that are used for the literature studies on graph Laplacians are [1] where Luxburg gives an introduction to the undirected graph Laplacian and [2], [3] where Zhou, Schölfopf, Huang and Hofmann give different definitions for the directed graph Laplacian. Also the book 'Spectral graph theory' of Chung plays an important role in the understandings of the graph Laplacian and its spectrum.

The following research questions are to be answered:

What definitions can we give for a graph Laplacian on directed graphs? What are their properties? How can the graph Laplacian be used on spectral clustering methods?

The structure of this thesis report is as follows: first an introduction to the Laplace operator and spectral clustering will be introduced in Chapter 2, in Chapter 3 the basic mathematical concepts and definitions relevant for this work will be given, followed by the literature study regarding the types of graph Laplacians and its application in Chapter 4 and Chapter 5, respectively. Finally, Chapter 6 contains the conclusion and a discussion.

The reader is expected to have basic prior knowledge of linear algebra and probability theory.

2 THE LAPLACIAN & SPECTRAL CLUSTERING

This section includes an introduction about the (graph) Laplacian. Also clustering and spectral clustering are introduced in this chapter.

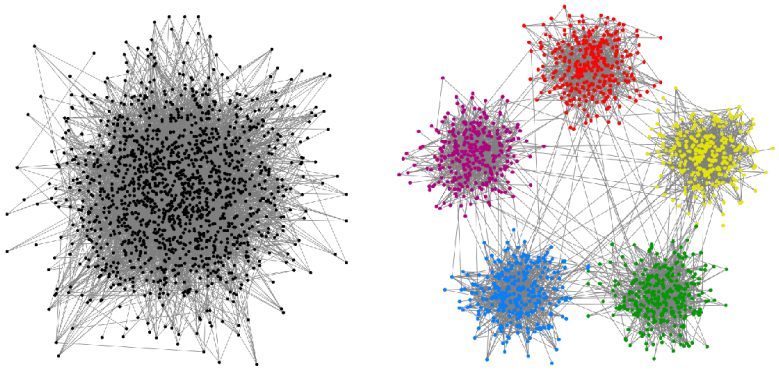


Figure 2.1: Clustering example of the same data set; left: unclustered - right: clustered [4]

2.1 THE LAPLACIAN

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a multivariate, continuous and differentiable function. Then, the *Laplacian* Δ of f is defined by

$$\Delta f(x_1, \ldots, x_n) := \nabla \cdot \nabla f = \sum_{k=1}^{n} \frac{\partial f^2}{\partial x_k^2}$$

In other words, the Laplacian of a function f is the divergence $\nabla\cdot$ of the gradient ∇ of f, which is the row-vector with partial derivatives of f as its entries. Specifically, the gradient is an operator on a multivariate function f at a specific point x, which shows the alteration of the function value at that point. So the divergence of the gradient will be big when the gradient at a point approaches a minimum and it will be small if the divergence approaches a maximum.This gives us an intuitive explanation for the Laplacian operator for continuous function.

But what about the Laplacian defined on a graph or a set of points in $\mathbb{R}^n$? Let $G = (V, E)$ be an undirected graph with corresponding adjacency matrix A and degree matrix D (see Chapter 3 for explanation of these concepts). Then the graph Laplacian on G is the discretesation of the Laplace operator on continuous functions. The graph Laplacian of a function in a node i, is the sum of the value difference between its adjacent nodes. In other words the graph Laplacian is defined by

$$(\Delta f)(j) = \sum_{i,j \in V} |f(i) - f(j)|,$$

with $f : V \to \mathbb{R}$ a function defined on the node set of G. In Chapter 4 we will show how this results in the following matrix representation of the graph Laplacian, which is called the Laplacian matrix L:

$$L = D - A,$$

which is defined on undirected graphs. The spectrum of L has many interesting properties which are useful for spectral clustering. In the next section we will go into further details of spectral clustering.

2.2 CLUSTERING

Referred to as "the gold of the 21st century" [5], data are an integral part of the modern era. With the technological enhancements of smart devices, sensors, Internet-of-Things (IoT) and many more applications, the data influx is expanding rapidly. Data to be analysed expands in size (also known as big data [6]), making methods to analyse data of more relevance.

Data analysis starts by clustering overarching data and specifies clusters into more detail as the algorithm processes. Therefore, the first indicator of groups within the analysed data is generic. These groups help identify dependencies of the analysed data and can be a useful starting point for further analysis. For a graphic depiction of this concept, the reader is referred to Figure 2.1.

Clustering methods are applied in everyday life, from the simple clustering of a survey result to the definition of different species. However, clustering is not only limited to everyday life and its application ranges from unsupervised machine learning to page ranking [7]. Spectral clustering is a category of clustering methods, that is related to partitioning clustering, but makes use of the graph Laplacian. In the following sections, the method will be briefly introduced.

2.2.1 Spectral clustering

Etymological research into the word spectral leads to the algebraic use of spectrum, or in other words the set of eigenvalues. Spectral clustering offers benefits over partitioning clustering methods as such that it can be solved with standard linear algebra and can be implemented simply.

Spectral clustering is often seen in the application of machine learning, data analysis in statistics (named exploratory data analysis), speech analysis (to perform deep-learning applications of speech, named speech processing) and visual data analysis by computers, such as videos and images (named computer vision) [8].

Looking into the specifics of how spectral clustering works, it can be defined as follows: a graph representation of the to be analysed data is created and the eigenvectors of its Laplacian matrix are computed. The graph representation is named similarity graph, whereby the information is ordered based upon a relationship in between the data. Then the k-means method is applied to the first k-eigenvalues, which are ordered increasingly [1]. Whereas, the k-means method, is an algorithmic approach that chooses k random points and clusters all the points by looking at the distance to each following k.

The approach for spectral clustering on undirected graphs can be described as follows:

- Determine a similarity graph in between the data, with either of the two following methods:
  - k-neighbourhood graph, based on lines connecting to each other, with an integer number k defining the distance in between nearby data-points;
  - ϵ-neighbourhood graph, where instead of lines connecting to each other a sphere is drawn with a radius of ϵ to determine the distance in between nearby data-points.
- Determine the eigenvalues of the constructed similarity graph, with special emphasis on the first big difference in between the non-zero eigenvalues.
- The eigenvalues are ordered with respect to their magnitude and the k-means method is applied to determine the appropriate cluster:
  - Choose random points in the data set;
  - Determine the mean of the clustered points;
  - Define this as the new central point;
  - Repetition of these steps until a global optimum is reached.

For mathematical descriptions of the above introduced terms, such as graphs and spectrum, readers are referred to Chapter 3. In Chapter 5 another algorithm for spectral clustering will be discussed because the algorithm above focuses only on undirected graph constructions and for this report, we are interested in spectral clustering for directed graphs.

3 BASIC CONCEPTS

In this chapter the basic concepts and notation that will be used in this research will be introduced and explained.

3.1 GRAPHS AND MATRIX REPRESENTATIONS

A *graph* $G = (V, E)$ is a pair of nodes (or vertices) and edges. The set V is the set of nodes i and the set $E \subseteq V \times V$ is the set of edges $e_{ij} = (i, j)$ with $i, j = 1, \ldots, n$. For the number of vertices n we write $|V| := n$. Edges are an ordered pair of two nodes. So we have that $e_{ij} \in E$ if and only if there is an edge between node $i \in V$ and $j \in V$. We say that nodes i and j are *adjacent* if there is an edge between them. An *undirected graph* has edges with no directions, unlike a *directed graph*, which has directed edges. So for an undirected graph we have $(i, j) = (j, i)$. The edge (i, i) is called a *self-loop*, where there is an edge from the node to itself. An undirected graph without self-loops is referred to as *simple*.

In the case of a directed graph, the (directed) edge is an ordered pair (i, j) where i and j are nodes of the graph and there is a directed edge going from node i to node j. For directed graphs the node i is said to be *adjacent to* the node j and the node j is said to be *adjacent from* the node i for the edge (i, j). A directed edge from a node i to itself is also called a *self-loop*.

A *bipartite graph* $G = (V_1, V_2, E)$ is a graph such that its vertex set can be divided into two subsets V_1 and V_2 with $V_1 \cap V_2 = \varnothing$. All edges of a bipartite graph have one end in V_1 and the other end in V_2. Notice that for a *directed bipartite graph* the edges have directions from one subset to the other.

There are different ways of transforming given data points into graphs and vice versa. In this report we consider a matrix representation of a graph, which will be defined in the following section.

3.1.1 Adjacency matrix

The $n \times n$-matrix that will be used to show the adjacent nodes of a graph is called the *adjacency matrix* A, with $A_{ij} = 1$, if node i is adjacent to node j, and $A_{ij} = 0$ otherwise. Notice that the adjacency matrix is symmetric for an undirected graph, so $A_{ij} = A_{ji}$.

3.1.2 Weight matrix

When the edges of a graph have weights, it is called a *weighted graph*. This means that every edge (i, j) has a value w_{ij} with $i, j = 1, \ldots, n$. The weight of an edge is taken positive, so $w_{ij} \geq 0$. We write $w_{ij} = 0$ when node i and j are not adjacent.

It follows that for an undirected weighted graph w_{ij} is equal to w_{ji}. Putting these values in an $n \times n$-matrix gives us the following *weight matrix*:

$$W := \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$$

Notice that for an undirected graph, the weight matrix W is symmetric. This is not the case for directed graphs, since w_{ij} does not need to be equal to w_{ji}. Also notice that the adjacency matrix is a special case of the weight matrix, where $w_{ij} = 1$ if $e_{ij} \in E$ and $w_{ij} = 0$ if $e_{ij} \notin E$. Sometimes the weight matrix is called the *weighted adjacency matrix*. From now on we will assume that G is a weighted graph.
Remark: In this report both w_{ij} and $w(i,j)$ are used as notations for the weight of an edge from node i to node j.

3.1.3 Degree matrix

The *degree* $d(i)$ of vertex i is the sum of the weights of all edges (i,j) such that $(i,j) \in E$, so

$$d(i) := \sum_{j \in V} w_{ij}.$$

Notice that $d(i)$ is the sum of the elements of the i-th row of W. A diagonal matrix D with the degrees $d(i)$ on its diagonal, will be called the *degree matrix* of a graph, so

$$D := \begin{bmatrix} d(1) & & \\ & \ddots & \\ & & d(n) \end{bmatrix} \quad \text{with} \quad D_{ij} := \begin{cases} d(i), & i = j, \\ 0, & i \neq j. \end{cases}$$

Whereas, the *volume* of an undirected graph G is defined as

$$vol(G) := \sum_{i \in V} d(i).$$

Directed graph

For directed graphs there are two different kind of degrees of a node i, namely the *indegree* $d_{in}(i)$ and the *outdegree* $d_{out}(i)$. The indegree of a node i is defined as

$$d_{in}(i) := \sum_{j \in V} w_{ij}.$$

And the outdegree of a node i is defined as:

$$d_{out}(i) := \sum_{j \in V} w_{ji}.$$

We define diagonal matrices D_{in} and D_{out} as follows:

$$D_{in} := \begin{bmatrix} d_{in}(1) & & \\ & \ddots & \\ & & d_{in}(n) \end{bmatrix} \quad \text{with} \quad D_{in,ij} := \begin{cases} d_{in}(i), & i = j \\ 0, & i \neq j \end{cases}$$

$$D_{out} := \begin{bmatrix} d_{out}(1) & & \\ & \ddots & \\ & & d_{out}(n) \end{bmatrix} \quad \text{with} \quad D_{out,ij} := \begin{cases} d_{out}(i), & i = j \\ 0, & i \neq j \end{cases}$$

3.1.4 Spectrum

The *spectrum* of a matrix is the set which contains its eigenvalues. In this report the spectrum is an ordered set. So for the set of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, we have $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.
When we speak of the spectrum of a graph, the context needs to be taken into consideration. In this report the spectrum of a graph means the spectrum of the corresponding graph Laplacian matrix.

3.1.5 Connectivity of graphs

An undirected graph $G = (V, E)$ is called *connected*, if there is a path between every two vertices in V. There exists a *path* between node i and j in V, if there exist $i_1, \ldots, i_k \in V$ for some $k > 0$ such that:

$$w(i, i_1)w(i_1, i_2) \cdots w(i_k, j) \neq 0,$$

with $i, j \in V$. The i and j are called *connected nodes*. A subset $S \subseteq V$ is called a *connected component* of G, if S is connected and for all $u \in S$ and $v \in S^c$ there is no path between u and v, with $S^c := V \backslash S$. Notice that for an undirected graph all paths are reversible, which means that if there is a path from i to j, there is also a path from j to i.

For a directed graph $G' = (V', E')$ we use the term *strongly connected*, if there is a reversible path between every pair of nodes. We say node i and j are *strongly connected* if there exists a path from i to j and vice versa.

3.2 RANDOM WALKS AND MARKOV CHAINS

Let $G = (V, E)$ be a directed graph. In Section 4.2.2 a definition for the directed graph Laplacian will be given, which makes use of the stationary distribution of a *random walk* on a directed graph G. To clarify the concept of a random walk on a graph, assume we are at an arbitrary node of a graph and choose to walk in a random direction on an edge. When the graph is unweighted, the node visited in the next time step will be chosen uniformly from among the neighbouring nodes [7, 9]. In the case of a weighted graph, the probability of node i visiting node j is proportionally related to the weight of the edge going from node i to node j [9]. So a random walk between two strongly connected nodes i and j is a path where the edges in between the nodes are chosen by considering the weighted edge going from i to j and the outdegree of i.

We will define the *transition probability function* $p(i, j) : V \times V \to \mathbb{R}^+$ for a random walk by

$$p(i, j) := \begin{cases} \frac{w(i,j)}{d_{out}(i)}, & \text{if } (i, j) \in E \\ 0, & \text{otherwise,} \end{cases}$$

which gives the probability of visiting node j, knowing that you are in node i. Notice that $p(i, j) \geq 0$ for all $i, j \in V$.

3.2.1 Markov Chains

The random walk on a graph can be considered as a *finite Markov Chain* as used in the mathematical field of probability theory. A Markov Chain is a stochastic model which consists of a set of states and a stochastic matrix[10]. The *states* $i = 1, \dots n$ of a Markov Chain are the possible positions of a random walk, where the position of the next state depends only on the state you are in. The set of states $\mathcal{X} = \{1, \dots, n\}$ will be called the *state space*. A Markov Chain is finite, when the state space is finite. A *stochastic matrix* P is a non-negative matrix where its columns sum up to 1. In a non-negative matrix, all elements are greater than or equal to zero. We will define the *transition probability matrix* P by the matrix such that $P(i, j)$ is the probability that when a Markov Chain is in state i the next state will be j. This is similar to the transition probability function $p(i, j)$ described above.

We see that a finite Markov Chain with transition probability matrix P and states $i, j \in \mathcal{X}$ is equivalent to a random walk on a directed graph with weight matrix W and nodes $i, j \in V$. A path on a directed graph is equivalent with a *walk* on a Markov Chain. The remaining of this section will contain concepts of Markov Chains that are equivalent to the concepts from graph theory.

Let $\mathcal{X} = \{1, \dots, n\}$ be a finite state space.

3.2.2 Stationary distribution

The probability distribution π which satisfies:

$$\pi^T = \pi^T P$$

with P the transition probability matrix as described above, is called the *stationary distribution*. Since it is a probability distribution, the following must hold for all $i \in \mathcal{X}$:

$$\pi(i) \geq 0 \text{ and } \sum_{i \in \mathcal{X}} \pi(i) = 1.$$

Notice that π is the eigenvector of P associated with the eigenvalue 1. We assume that π is a column vector in $\mathbb{R}^n$ with its entries $\pi(i)$ with $i \in \mathcal{X}$. More insights into the existence and uniqueness of the stationary distribution (and thus of the eigenvalue 1 and its associated eigenvector) can be found in Section 4.2.2.

3.2.3 Communicating states

In Section 3.1.5 we have seen that node i and j are strongly connected if there exists a path from i to j and vice versa. If we consider a Markov Chain with transition probability matrix P, we say that two states *communicate* if the probability of a walk starting from i visits j and vice versa is greater than zero. In other words, if for some $k > 0$ we have that

$$P^k(i, j) > 0,$$

then i and j are *communicating states*. The value $P^k(i, j)$ is equal to the element in the i^{th} row and j^{th} column of the matrix P^k. In other words, $P^k(i, j)$ is equal to the probability of the next visit being j while in state i after k time steps. Notice that communicating is a transitive relation. So if state i communicates with state j, then state j communicates with state i as well. The state space $\mathcal{X}$ can be split into *communicating classes*. A communicating class is a subset of $\mathcal{X}$ such that all states communicate.

Remark: Note that communicating states in a Markov Chain correspond to strongly connected nodes in a directed graph.

3.2.4 Irreducible chain

When a Markov Chain consists of just one communicating class, it is called an *irreducible chain*. In other words, if each state can be reached from every other state in the chain, the Markov Chain is *irreducible*. The corresponding transition probability matrix of a irreducible chain is also called irreducible.

Remark: Note that an irreducible chain corresponds to a strongly connected directed graph.

3.2.5 Transient and (positive) recurrent states

Define $\tau_{ij}^{(k)}$ as

$$\tau_{ij}^{(k)} := \mathbb{P}(X_1 \neq j, \ldots, X_{k-1} \neq j, X_k = j),$$

with $X_1, X_2, \ldots$ random variables which describe the state of a Markov Chain at a time step. Then $\tau_{ij}^{(k)}$ is equal to the probability that a path which starts at state i visits state j for the first time at some time step k. Define τ_{ij}^{*} as

$$\tau_{ij}^{*} := \sum_{k=1}^{\infty} \tau_{ij}^{(k)},$$

which describes the probability that a walk which starts at state i will visit state j at some later time step in the future. It is clear that if $\tau_{ij}^{*} > 0$ and $\tau_{ji}^{*} > 0$, then i and j are communicating states.

If we have $\tau_{ii}^{*} = 1$, then we call i a *recurrent* state. In other words, if the probability that the chain will return to state i when starting at state i is equal to one, the state i is called recurrent. If $\tau_{ii}^{*} < 1$, then the state i will be called *transient*. In other words, the probability to not come back to that state is bigger than zero.
Let t_{ij} be the time-step of the first visit to j when starting at i. Now define μ_{ij} as

$$\mu_{ij} := \mathbb{E}[t_{ij}].$$

This is the expected time for a walk from state i to j. The μ_{ii} can be defined as the expected return time. Then if $\mu_{ii} < \infty$, we will call state i a positive recurrent state. In other words, a recurrent state with a finite expected return time is called *positive recurrent*.

3.2.6 Periodic and aperiodic

Let i be a state such that $P^k(i,i) > 0$ for some $k > 0$, with P^k defined as in Section 3.2.3. Then the *period* q of state i is equal to the greatest common divisor of all $k > 0$ such that $P^k(i,i) > 0$. When $q = 1$, the state is called *aperiodic*. If all states in a finite Markov Chain are aperiodic, the chain is called *aperiodic*.

Remark: The periodicity of state i can be considered as the greatest common divisor of the lengths of all paths which start at node i and return to node i. The *length* of a path is the number of edges in that path, without considering the weights of the edges.

4 TYPES OF GRAPH LAPLACIANS

The graph Laplacian is a tool that is used for clustering methods, pageranking, chemistry, etc. In Chapter 2 background information over the Laplace operator is given. This chapter will focus on the types of graph Laplacians for undirected and directed graphs. There are multiple ways of defining a graph Laplacian. Furthermore we will see some of their most interesting properties.
The graph Laplacian defined on undirected graphs will be called the *undirected graph Laplacian*. The graph Laplacian defined on directed graphs will be called the *directed graph Laplacian*.

Notice that there is no such thing as a unique definition of the directed graph Laplacian.

4.1 UNDIRECTED GRAPH LAPLACIANS

Let us start by looking at an undirected graph $G = (V, E)$ which is weighted and has a weight matrix W as in Section 3.1.2 and a degree matrix D as in Section 3.1.3. Let $|V| = n$. Let $f : V \to \mathbb{R}$ be a function that is defined on the vertex set V of G. Then $f(i)$ is the value of f at the vertex i. We can denote $f = (f_1, \dots, f_n)^T$ as a column vector in $\mathbb{R}^n$. Let $\mathcal{V} = \{f : V \to \mathbb{R}\}$ be the set of all functions defined on V. Notice that $\mathcal{V} \cong \mathbb{R}^n$. The *graph Laplacian* is a linear transformation $\Delta : \mathcal{V} \to \mathcal{V}$, which is defined as

$$(\Delta f)(i) := \sum_{j \in V} w_{ij}(f(i) - f(j)), \tag{4.1}$$

for all $i \in V$.

4.1.1 Unnormalised graph Laplacian matrix

To find a matrix representation for the graph Laplacian we are going to write out the definition and for that we use the definitions from Section 3.1.2 and Section 3.1.3. For all $i \in V$ it holds that

$$\begin{aligned}
(\Delta f)(i) &= \sum_{j \in V} w_{ij}(f(i) - f(j)) \\
&= \sum_{j \in V} w_{ij} f(i) - \sum_{j \in V} w_{ij} f(j) \\
&= d(i) f(i) - \sum_{j \in V} w_{ij} f(j) \\
&= (Df)(i) - (Wf)(i) \\
&= ((D - W)f)(i).
\end{aligned}$$

So we find a definition for the *unnormalised graph Laplacian matrix* L for a weighted undirected graph G which is defined as

$$L := D - W \quad \text{with} \quad L_{ij} := \begin{cases} d(i) - w_{ii}, & \text{if } i = j, \\ -w_{ij}, & \text{if } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases}$$

Notice that for an unweighted undirected graph the definition of the graph Laplacian matrix reduces to $L = D - A$.
For spectral clustering a few properties of the graph Laplacian are important. The following proposition will sum these up:

Proposition 1 (Properties of unnormalised graph Laplacian matrix L) *Let L be the unnormalised graph Laplacian matrix defined on an undirected graph G with non-negative weights. Then the following hold:*

1. *Let $f : V \to \mathbb{R}$ be a function. Then we have:*

$$\langle f, Lf \rangle = f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f(i) - f(j))^2,$$

   $\langle \cdot, \cdot \rangle$ the inner product.

2. *The graph Laplacian matrix L is symmetric.*
3. *The graph Laplacian matrix L is positive semi-definite.*
4. *The graph Laplacian matrix L has 0 as its smallest eigenvalue with the constant one vector $\mathbb{1} := (1, 1, \ldots, 1)^T$ as its associated eigenvector.*
5. *The number of eigenvalues of L is n with $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.*

Proof.

1. Since we have $d(i) = \sum_{j=1}^{n} w_{i,j}$, it follows that for $f \in \mathbb{R}^n$:

$$\begin{aligned} f^T L f &= f^T D f - f^T W f = \sum_{i=1}^{n} d(i) f(i)^2 - \sum_{i,j=1}^{n} f(i) f(j) w_{ij} \\ &= \frac{1}{2} \left(\sum_{i=1}^{n} d(i) f(i)^2 - \sum_{i,j=1}^{n} f(i) f(j) w_{ij} + \sum_{j=1}^{n} d(j) f(j)^2 \right) \\ &= \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f(i) - f(j))^2. \end{aligned}$$

2. The symmetry of L follows from the symmetry of D and W. See Section 3.1.2 and Section 3.1.3.
3. From 1. it follows that for all f we have $f^T L f \geq 0$. This is exactly the definition for positive semi-definiteness of L.
4. We have that $L\mathbb{1} = 0$, so 4. is trivial.
5. This follows immediately from 1-4. □

In [11] and [12] many more properties can be found.

By inspecting the spectrum of the graph Laplacian L an important property for the undirected graph G follows. This property can be found in Proposition 2:

Proposition 2 (Number of connected components and the spectrum of L)
Let $G = (V, E)$ be an undirected weighted graph. Then the algebraic multiplicity k of the eigenvalue 0 of L is equal to the number of connected components $A_1, \ldots, A_k$ in the graph. So if $\lambda_1 = 0$ and $\lambda_2 > 0$ with λ_i as in Proposition 1.5, then G is a connected graph. The span of the indicator vectors $\mathbb{1}_{A_1}, \cdots, \mathbb{1}_{A_k}$ is the eigenspace of eigenvalue 0. The indicator vector is defined as $\mathbb{1}_{A_k}$ the constant one vector which is the eigenvector which belongs to the k^{th} component of G.

Proof.
Let $G = (V, E)$ be an undirected connected graph. Then G have one component, thus $k = 1$. Let f be an eigenvector with eigenvalue 0 of the graph Laplacian L. Then we have $Lf = 0$. Thus from Proposition 1.1 follows that

$$\sum_{i,j=1}^{n} w_{ij}(f(i) - (fj))^2 = f^T L f = f^T \cdot 0 = 0.$$

From the definition of a weighted graph we have that $w_{ij} \geq 0$. If the nodes i and j are adjacent, then $w_{ij} > 0$, so $f(i)$ needs to equal $f(j)$ for this sum to be equal to 0. It follows that f is constant for all connected nodes. See Section 3.1.5 for the definition of connected nodes. If all nodes in a graph are connected, then f will be a constant vector. Then G has the constant one vector $\mathbb{1}$ as its eigenvector with eigenvalue 0 corresponding to the one connected component.
Now let $G = (V, E)$ be a graph with k connected components. Assume without loss of generality, that the nodes are ordered according to their corresponding connected component. When the nodes are ordered, W and thus L have a block diagonal form as follows:

$$L = \begin{bmatrix} L_1 & & & \\ & L_2 & & \\ & & \cdots & \\ & & & L_k \end{bmatrix}.$$

Notice that block L_i corresponds to the i^{th} connected component of the graph. For block diagonal matrices we know that the spectrum of L consists of the union of the spectrum of the L_i, with eigenvectors of L being the eigenvectors of L_i with 0 entries on the positions of all L_j such that $j \neq i$. Since each L_i is a graph Laplacian of a connected graph, we know from Proposition 1.4 that each L_i has an eigenvalue 0. From the first part of this proof we know that the corresponding algebraic multiplicity of the eigenvalue 0 is 1 with the eigenvector being the constant one vector associated with the i^{th} connected component. Thus, L has as many 0 eigenvalues as the number of connected components with its eigenvectors being the indicator vectors corresponding to the connected components. □

4.1.2 Normalised graph Laplacian matrices

The two definitions for normalised graph Laplacians in the book 'Spectral graph theory' [12] are as follows. The first definition is

$$L_{sym} := D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2},$$

with L defined as in Section 4.1.1. In other words:

$$L_{sym,ij} := \begin{cases} 1 - \frac{w_{ii}}{d(i)}, & \text{if } i = j \text{ and } d(i) \neq 0, \\ -\frac{w_{ij}}{\sqrt{d(i)d(j)}}, & \text{if } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases}.$$

Notice that L_{sym} is a symmetric matrix. A second definition for a normalised graph Laplacian is

$$L_{rw} := D^{-1} L = I - D^{-1} W.$$

In other words,

$$L_{rw,ij} := \begin{cases} 1 - \frac{w_{ii}}{d(i)}, & \text{if } i = j \text{ and } d(i) \neq 0, \\ -\frac{w_{ij}}{d(i)}, & \text{if } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases}$$

L_{rw} is not symmetric in general. Furthermore it is closely related to a random walk which is described in Section 3.2. The entries $L_{rw,ij}$ are defined as the weight of

the edge (i, j) normalised by the degree of node i. So it can be considered as the probability of walking to node j, knowing that you are at node i. In Section 4.2 a closer look at random walks will be taken.

A few important properties of L_{sym} and L_{rw} are described in [13]. The most important properties for this project can be read in the following propositions.

Proposition 3 (Properties of L_{sym} and L_{rw}) Let L_{sym} and L_{rw} be normalised undirected graph Laplacians. Then the following hold:

1. *Let $f : V \to \mathbb{R}$ be a function. Then we have:*

$$\langle f, L_{sym} f \rangle = f^T L_{sym} f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \left(\frac{f(i)}{\sqrt{d(i)}} - \frac{f(j)}{\sqrt{d(j)}} \right)^2 .$$

2. *An eigenvalue λ of L_{rw} has corresponding eigenvector u if and only if L_{sym} has eigenvalue λ with corresponding eigenvector $v = D^{1/2}u$.*
3. *The normalised undirected graph Laplacian L_{rw} has eigenvalue λ with corresponding eigenvector u if and only if for λ and u we have that $Lu = \lambda Du$.*
4. *The normalised undirected graph Laplacians L_{rw} and L_{sym} both have eigenvalue 0, with corresponding eigenvectors the constant one vector $\mathbb{1}$ and $D^{1/2}\mathbb{1}$ respectively.*
5. *The normalised undirected graph Laplacian L_{sym} is positive semi-definite and has n eigenvalues such that $0 = \lambda_1 \leq \cdots \leq \lambda_n$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.*
6. *The normalised undirected graph Laplacian L_{rw} is positive semi-definite and has n eigenvalues such that $0 = \lambda_1 \leq \cdots \leq \lambda_n$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.*

Proof.

1. This can be proved in a similar way as in Proposition 1.1.
2. Let λ be an eigenvalue of both L_{rw} as L_{sym}. Then we have that $L_{sym}v = \lambda v$. Multiplying the definition of L_{sym} from the left with $D^{-1/2}$ gives:

$$D^{-1}LD^{-1/2}v = \lambda D^{-1/2}v.$$

   Substituting $u = D^{-1/2}v$ gives: $D^{-1}Lu = \lambda u$. So Ł$_{rw}u = \lambda u$, which shows 2. The other direction is analogous.
3. Same proof as 2. but multiply $L_{rw}u = \lambda u$ with D from the left side.
4. We have that $L_{rw}\mathbb{1} = 0$, so the constant one vector is indeed the corresponding eigenvector of the eigenvalue 0 of L_{rw}. The from 3.2 it follows that $D^{-1/2}\mathbb{1}$ is the corresponding eigenvector of eigenvalue 0 of L_{sym}.
5. From 1. it follows that $f^T L_{sym} f \geq 0$, so L_{sym} is positive semi-definite.
6. This follows from 2. en 5. □

Proposition 4 (Number of connected components and the spectrum of L_{sym}
Let $G = (V, E)$ be an undirected weighted graph. Then the algebraic multiplicity k of the eigenvalue 0 of L_{sym} and L_{rw} is equal to the number of connected components $A_1, \ldots, A_k$ in the graph. So if $\lambda_1 = 0$ and $\lambda_2 > 0$ with λ_i as in Proposition 3.5 and 3.6, then G is a connected graph. The span of the indicator vectors $\mathbb{1}_{A_1}, \cdots, \mathbb{1}_{A_k}$ is the eigenspace of eigenvalue 0 of L_{rw}. The span of the $\mathbb{1}_{A_1}, \cdots, \mathbb{1}_{A_k}$ is the eigenspace of eigenvalue 0 of L_{rw}. The span of the $D^{1/2}\mathbb{1}_{A_1}, \cdots, D^{1/2}\mathbb{1}_{A_k}$ is the eigenspace of eigenvalue 0 of L_{sym}.

Proof.
The proof of Proposition 4. goes in a similar way as the proof of Proposition 2.

4.2 DIRECTED GRAPH LAPLACIANS AND THEIR PROPERTIES

As seen in the previous section, the undirected graph Laplacians have interesting properties which can be used for spectral clustering. This leaves us thinking:

Do we have the same properties when we apply the graph Laplacian on directed graphs?

For the answer to this question we need dive into some definitions for the directed graph Laplacian. Literature into directed graph Laplacians is scarce, as much of research and applications mostly focuses on the undirected graph Laplacian. Therefore, in this research, two different definitions of directed graph Laplacians are introduced and compared with the undirected graph Laplacian.

4.2.1 Hub and authority sets

In [2], Zhou, Schölkopf and Hoffmann give a definition for a directed graph Laplacian. Notice that for this method no self-loops in directed graphs are allowed.

Let G be a weighted directed graph. In [2] it states that every directed graph G can be represented as a bipartite graph $\mathcal{G} = (V_{out}, V_{in}, \mathcal{E})$. Here V_{out} is a copy of the set of nodes with a positive outdegree and V_{in} is a copy of the set of nodes with a positive indegree. For node $i \in V$ with positive outdegree we write $i_{out} \in V_{out}$ and for $j \in V$ with positive indegree we write $j_{in} \in V_{in}$. In other words,

$$V_{out} := \{i_{out} : i \in V \text{ and } d_{out}(i) > 0\}$$

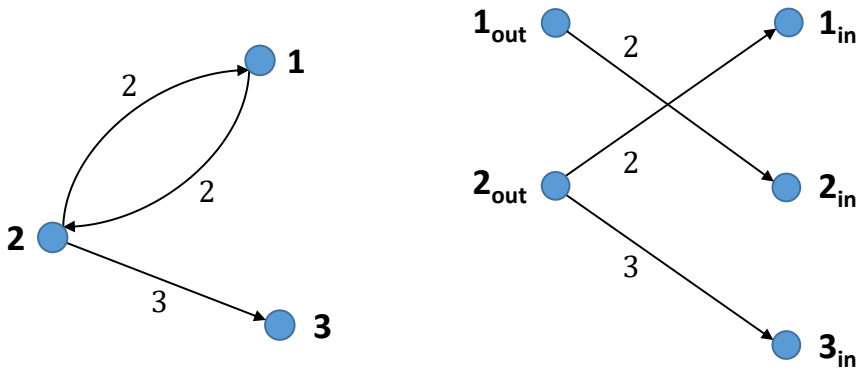
and

$$V_{in} := \{j_{in} : j \in V \text{ and } d_{in}(j) > 0\}.$$

Furthermore $\mathcal{E} \subseteq V_{out} \times V_{in}$ is the set of edges (i_{out}, j_{in}) from V_{out} to V_{in}. Note that this bipartite graph is still a directed graph which has directed edges from the set V_{out} to V_{in}.

The idea to construct a directed graph as a bipartite graph was proposed by Kleinberg in an article regarding authoritative hyperlink sources in the Journal of the Association for Computing Machinery [14]. In this article, Kleinberg introduced hub and authority web models, where authoritative pages have relevance to a specific topic, whereas hub pages do have a hyperlink to many related authorities. Note that a web page can both be an authoritative and hub page at the same time. This is equivalent to saying that a node in a directed graph can have both an outdegree as an indegree larger than zero. Furthermore, one of the essential findings of the paper is that a relevant authority page has a lot of hyperlinks from different relevant hub pages, whereas a relevant hub page has a lot of hyperlinks to relevant authority pages. Based on the proposal of Kleinberg, within this research V_{out} will be referred to as the hub set and V_{in} will be called the authority set. Notice that V_{out} and V_{in} are disjoint. If a node $i \in V$ has a out- and indegree greater than zero, we do not say i is in V_{out} and in V_{in}, since we put copies i_{out} and i_{in} of i in V_{out} and V_{in} respectively. There are no edges between the nodes in V_{out} and also no edges between the nodes in V_{in}.

So let us take a look at the directed graph of Figure 4.1a with $V = \{1, 2, 3\}$ and $w(1,2) = 2$, $w(2,1) = 2$ and $w(2,3) = 3$. So the corresponding directed bipartite graph can be shown as in Figure 4.1b. For this graph we have $V_{out} = \{1_{out}, 2_{out}\}$ and $V_{in} = \{1_{in}, 2_{in}, 3_{in}\}$ with the weighted edges $\mathcal{E} = \{(1_{out}, 2_{in}), (2_{out}, 1_{in}), (2_{out}, 3_{in})\}$.



(a) Directed graph $G = (V, E)$ with $V = \{1, 2, 3\}$, $E = \{(1, 2), (2, 1), (2, 3)\}$ and weights $w(1, 2) = 2$, $w(2, 1) = 2$ and $w(2, 3) = 3$.

(b) Bipartite graph $\mathcal{G} = (V_{out}, V_{in}, \mathcal{E})$ with on the left the hub set $V_{out} = \{1, 2\}$ and on the right the authority set $V_{in} = \{1, 2, 3\}$ with weights $w(1, 2) = 2$, $w(2, 1) = 2$ and $w(2, 3) = 3$.

Figure 4.1: Directed graph and bipartite graph.

To define a directed graph Laplacian, in [2] first Laplacians are defined on the hub set V_{out} and the authority set V_{in}. Then these two are combined to define a Laplacian on the whole vertex set V in the following way:

First we define the graph Laplacian on the authority set V_{in}. Let Q be a linear operator defined on $\mathcal{H}(V_{in})$, which is the space of functions $f : V_{in} \to \mathbb{R}$. Then $Q : \mathcal{H}(V_{in}) \to \mathcal{H}(V_{out})$ is defined as

$$(Qf)(i_{out}) := \sum_{i_{in} \in V_{in}} \frac{w(i_{out}, i_{in})}{\sqrt{d_{out}(i_{out}) d_{in}(i_{in})}} f(i_{in}).$$

Then the adjoint of this operator, $Q^* : \mathcal{H}(V_{out}) \to \mathcal{H}(V_{in})$, is:

$$(Q^* f)(i_{in}) = \sum_{i_{out} \in V_{out}} \frac{w(i_{out}, i_{in})}{\sqrt{d_{out}(i_{out}) d_{in}(i_{in})}} f(i_{out}).$$

Multiplying Q^* with Q gives a new operator $\mathcal{S}_{V_{in}} : \mathcal{H}(V_{in}) \to \mathcal{H}(V_{in})$:

$$\mathcal{S}_{V_{in}} := Q^* Q,$$

which finally gives us a definition for the graph Laplacian $\Delta_{V_{in}} : \mathcal{H}(V_{in}) \to \mathcal{H}(V_{in})$ on the authority set V_{in}:

$$\Delta_{V_{in}} := I - \mathcal{S}_{V_{in}}.$$

with I denoting the identity operator. In [15] the authors use $\mathcal{S}_{V_{in}}$ to determine a ranking for the authoritative pages, so for all nodes with an indegree bigger than zero. We will not go into the details of the application in page ranking.

A careful reader might have noticed the similarities between the directed graph Laplacian defined on the authority set and the definition of the normalised undirected graph Laplacian L_{sym} as defined in Section 4.1.

Let us first compare the inner product $\langle f, \Delta_{V_{in}} f \rangle$ with the inner product $\langle f^*, L_{sym} f^* \rangle$, with f denoting a function on V_{in} and f^* denoting a function on V. The inner product on a function space $\mathcal{H}(V)$ is defined by: $\langle f, g \rangle = \sum_{i \in V} f(i) g(i)$. From Proposition 3 in Section 4.1 we know that:

$$\langle f, L_{sym} f \rangle = f^T L f = \frac{1}{2} \sum_{i,j \in V} w_{ij} \left(\frac{f(i)}{\sqrt{d(i)}} - \frac{f(j)}{\sqrt{d(j)}} \right)^2. \tag{4.2}$$

In [2] a functional is defined which measures the smoothness of a function f and approaches small variances in the function values for nodes in V_{in}.

$$\Omega_{V_{in}}(f) := \frac{1}{2} \sum_{i_{in}, j_{in} \in V_{in}} \sum_{i_{out} \in V_{out}} c_{i_{out}}(i_{in}, j_{in}) \left(\frac{f(i_{in})}{\sqrt{d_{in}(i_{in})}} - \frac{f(j_{in})}{\sqrt{d_{in}(j_{in})}} \right)^2, \tag{4.3}$$

with

$$c_{i_{out}}(i_{in}, j_{in}) = \frac{w(i_{out}, i_{in}) w(i_{out}, j_{in})}{d_{out}(i_{out})}$$

Let $i_{in}, j_{in} \in V_{in}$ and $i_{out} \in V_{out}$ such that there is an edge from i_{out} to both i_{in} as j_{in}. The value $c_{i_{out}}(i_{in}, j_{in})$ is referred to as the *co-bridge value*.

For example in Figure 4.1b node 1_{in} and 3_{in} both have an edge coming from 2_{out}. Their corresponding co-bridge value is: $c_{2_{out}}(1_{in}, 3_{in}) = \frac{6}{5}$.

Notice that Equation 4.3 looks like Equation 4.2. Writing out the right side of Equation 4.3, substituting $c_{i_{out}}(i_{in}, j_{in})$ and using that $\sum_{j_{in} \in V_{in}} \frac{w(i_{out}, j_{in})}{d_{out}(i_{out})} = 1$ and $\sum_{i_{out} \in V_{out}} \frac{w(i_{out}, j_{in})}{d_{in}(i_{in})} = 1$ gives:

$$\begin{aligned} \Omega_{V_{in}}(f) &= \sum_{i_{in} \in V_{in}} f^2(i_{in}) - \sum_{i_{in}, j_{in}} \sum_{i_{out}} f(i_{in}) \frac{w(i_{out}, i_{in})}{\sqrt{d_{out}(i_{out}) d_{in}(i_{in})}} \frac{w(i_{out}, j_{in})}{\sqrt{d_{out}(i_{out}) d_{in}(j_{in})}} f(j_{in}) \\ &= \langle f, \Delta_{V_{in}} f \rangle. \end{aligned} \tag{4.4}$$

We can define a new graph $G_{in} = (V_{in}, E_{in})$ with edge weights

$$\omega(i_{in}, j_{in}) := \sum_{i_{out} \in V_{out}} c_{i_{out}}(i_{in}, j_{in}),$$

with corresponding weight matrix W with $\omega(i_{in}, j_{in})$ as its entries on the i^{th} row and j^{th} column. For our example of Figure 4.1a the corresponding graph defined on the authority set V_{in} looks as follows:

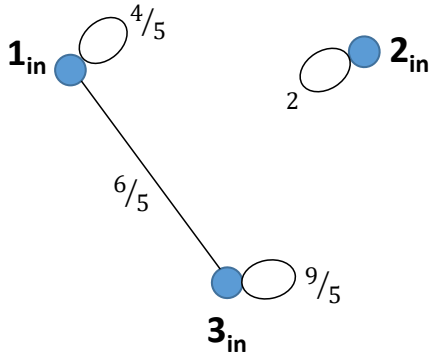


Figure 4.2: Graph $G_{in} = (V_{in}, E_{in})$ with edge weights $\omega(1_{in}, 1_{in}) = \frac{4}{5}$, $\omega(2_{in}, 2_{in}) = 2$, $\omega(3_{in}, 3_{in}) = \frac{9}{5}$ and $\omega(1_{in}, 3_{in}) = \omega(3_{in}, 1_{in}) = \frac{6}{5}$

The corresponding degree and weight matrix are:

$$D = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 3 \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} \frac{4}{5} & 0 & \frac{6}{5} \\ 0 & 2 & 0 \\ \frac{6}{5} & 0 & \frac{9}{5} \end{bmatrix}$$

Notice that $\omega(i_{in}, j_{in}) = \omega(j_{in}, i_{in})$, so G_{in} is an undirected graph. We also notice that the degree of a node j_{in} of graph G_{in} is equal to $d(j_{in}) = \sum_{i_{in} \in V_{in}} \omega(i_{in}, j_{in})$. We can show that this is equal to the indegree of j_{in} as follows:

Proof.

$$\begin{aligned}\sum_{i_{in} \in V_{in}} \omega(i_{in}, j_{in}) &= \sum_{i_{in} \in V_{in}} \sum_{i_{out} \in V_{out}} c_{i_{out}}(i_{in}, j_{in}) \\ &= \sum_{i_{in} \in V_{in}} \sum_{i_{out} \in V_{out}} \frac{w(i_{out}, i_{in}) w(i_{out}, j_{in})}{d_{out}(i_{out})} \\ &= \sum_{i_{out} \in V_{out}} w(i_{out}, j_{in}) \sum_{i_{in} \in V_{in}} \frac{w(i_{out}, i_{in})}{d_{out}(i_{out})}\end{aligned}$$

Since:

$$\sum_{i_{in} \in V_{in}} \frac{\omega(i_{out}, i_{in})}{d_{out}(i_{out})} = \frac{1}{d_{out}(i_{out})} \sum_{i_{in} \in V} \omega(i_{out}, i_{in}) = \frac{d_{out}(i_{out})}{d_{out}(i_{out})} = 1,$$

we have that

$$\begin{aligned}\sum_{i_{in} \in V_{in}} \omega(i_{in}, j_{in}) &= \sum_{i_{out} \in V_{out}} w(i_{out}, j_{in}) \\ &= d_{in}(j_{in}).\end{aligned}$$

□

We see that the definition of the graph Laplacian on the authority set of a directed graph is equivalent to the symmetric normalised graph Laplacian on an undirected graph.

So the graph Laplacian on the authority set can be written as:

$$\Delta_{V_{in}} = I - D^{-1/2} W D^{-1/2},$$

with D the degree matrix of the graph on the authority set and W the weight matrix defined on V_{in} with $\omega(i_{in}, j_{in})$ as its entries.

From Equation 4.4 we have that $\Delta_{V_{in}}$ is a positive semi-definite matrix. The eigenvalues of $\Delta_{V_{in}}$ are in $[0, 1]$. This can be shown in the same as in Proposition 1 and 3.

In the same way we can define a graph Laplacian on the hub set with:

$$\mathcal{S}_{V_{out}} = QQ^* \text{ and } \Delta_{V_{out}} = I - \mathcal{S}_{V_{out}},$$

with both operators defined on $\mathcal{H}(V_{in})$. The cobridge value can be defined as:

$$c_{i_{in}}(i_{out}, j_{out}) = \frac{w(i_{out}, i_{in}) w(j_{out}, i_{in})}{d_{in}(i_{in})},$$

where $i_{out}, j_{out} \in V_{out}$ both have an edge to the same node i_{in} in the authority set. We can define a graph G_{out} in the same way as for G_{in}. Furthermore another smoothness functional is defined as in Equation 4.3, but for the hub set V_{out}:

$$\Omega_{V_{out}}(f) = \frac{1}{2} \sum_{i_{out}, j_{out} \in V_{out}} \sum_{i_{in} \in V_{in}} c_{i_{in}}(i_{out}, j_{out}) \left(\frac{f(i)}{\sqrt{d_{out}(i)}} - \frac{f(j)}{\sqrt{d_{out}(j)}} \right)^2. \tag{4.5}$$

In the same way as for $\Omega_{V_{in}}$, we have that:

$$\Omega_{V_{out}}(f) = \langle f, \Delta_{V_{out}} f \rangle.$$

Thus the properties of the spectrum of $\Delta_{V_{out}}$ are the same.

As seen above we have defined a graph Laplacian for two subsets of the vertex set of a directed graph G. We have considered a new graph on V_{in} and V_{out} with edge weights $\omega(i_{in}, j_{in}) = \sum_{i_{out} \in V_{out}} c_{i_{out}}(i_{in}, j_{in})$ and $\omega(i_{out}, j_{out}) = \sum_{i_{in} \in V_{in}} c_{i_{in}}(i_{out}, j_{out})$. So we still do not have a proper definition of a graph Laplacian defined on all of a directed graph. The definition will be constructed as follows:

First let us start with the linear operator Q which is only defined on $\mathcal{H}(V_{in})$.The idea is to extend the domain of Q to $\mathcal{H}(V)$. We will do that by defining:

$$(Qf)(i) = \begin{cases} 0, & \text{if } i \in V_{in} \backslash V_{out} \\ \sum_{j \in V_{in}} \frac{w(i,j)}{\sqrt{d_{out}(i) d_{in}(j)}} f(j), & \text{if } i \in V_{out}. \end{cases}$$

Its corresponding adjoint Q^* with extended domain will give:

$$(Q * f)(i) = \begin{cases} 0, & \text{if } i \in V_{out} \backslash V_{in} \\ \sum_{j \in V_{out}} \frac{w(j,i)}{\sqrt{d_{out}(j) d_{in}(i)}} f(j), & \text{if } i \in V_{in}. \end{cases}$$

Then we can define $\mathcal{S}_\delta : \mathcal{H}(V) \to \mathcal{H}_V$ by:

$$\mathcal{S}_\delta = \delta \mathcal{S}_{V_{in}} + (1 - \delta) \mathcal{S}_{V_{out}} \text{ with } 0 \leq \delta \leq 1.$$

The parameter δ emphasises the importance of each $\mathcal{S}$. It will also be used to define the graph Laplacian $\Delta_\delta : \mathcal{H}(V) \to \mathcal{H}(V)$ for the whole node set V:

$$\Delta_\delta = I - \mathcal{S}_\delta = \delta \Delta_{V_{in}} + (1 - \delta) \Delta_{V_{out}}.$$

Linearly combining the two smoothness functionals by again using the parameter δ, gives a mapping defined on V:

$$\Omega_\delta(f) = \delta \Omega_{V_{in}}(f) + (1 - \delta) \Omega_{V_{out}}(f) = \delta \langle f, \Delta_{V_{in}} f \rangle + (1 - \delta) \langle f, \Delta_{V_{out}} f \rangle,$$

and by linearity of the inner product we find that:

$$\Omega_\delta(f) = \langle f, \Delta_\delta f \rangle.$$

Thus we have found a definition for a graph Laplacian defined on a directed graph, which is:

$$\begin{aligned} \Delta_\delta &= \delta \Delta_{V_{in}} + (1 - \delta) \Delta_{V_{out}} \\ &= \delta (I - D_{V_{in}}^{-1/2} W_{V_{in}} D_{V_{in}}^{-1/2}) + (1 - \delta)(I - D_{V_{out}}^{-1/2} W_{V_{out}} D_{V_{out}}^{-1/2}), \end{aligned}$$

with $D_{V_{in}}, W_{V_{in}}$ and $D_{V_{out}}, W_{V_{out}}$ the degree and weight matrix of the the graphs G_{in} and G_{out} respectively.

4.2.2 Transition Probability Matrix

The second definition for the directed graph Laplacian comes from the article of Zhou, Huang and Schölkopf [3]. Also Chung uses the same definition in her article [13]. It contains the transition probability matrix of a random walk on a directed graph, which can be considered as a finite Markov Chain. See Section 3.2 for an introduction to these concepts.

Let $G = (V, E)$ be a directed graph, with $|V| = n$. The definition that will be explained in this section, is applicable to graphs that are both strongly connected and aperiodic. Later we will clarify why aperiodicity is not necessary, but can be useful. For now, we assume that G is a strongly connected and aperiodic graph. A random walk defined on G with associated transition probability function $p : V \times V \to \mathbb{R}^+$ is

$$p(i,j) := \begin{cases} \frac{w_{ij}}{d_{out}(i)}, & \text{if } (i,j) \in E \\ 0, & \text{otherwise,} \end{cases}$$

with $i, j = 1, \ldots, n$. The matrix representation of the transition probabilities will be called the transition probability matrix P as in Section 3.2.
The next step in the definition for a directed graph Laplacian requires a unique stationary distribution π for the random walk, which is defined as in Section 3.2. The following comes to mind:

1. (*Existence*) Does a finite Markov Chain always have a stationary distribution?
2. (*Uniqueness*) If so, when is the corresponding stationary distribution unique?

Existence of stationary distribution π

First we will start by showing the *existence* of π. Assume that we have a *Markov Chain defined on a finite state space* $\mathcal{X} = \{1, 2, ..., n\}$.
Claim:

For every finite Markov Chain there exists a stationary distribution.

For the proof of this statement we will use the Krylov–Bogoliubov argument [16], which shows that every finite Markov Chain has a stationary distribution, even when not irreducible. Let $\mathcal{P}$ be the set of probability distributions that are defined on the finite state space $\mathcal{X}$:

$$\mathcal{P} := \{v \in \mathbb{R}^n \mid \text{ for all } i \in \mathcal{X} : v(i) \geq 0 \text{ and } \sum_{i \in \mathcal{X}} v(i) = 1\}.$$

We consider $v \in \mathcal{P}$ as column vectors. A probability distribution can be considered as a column vector in $\mathbb{R}^n$ with the i^{th} entry being the probability of being in that state, where its entries sum up to 1. The argument uses that $\mathcal{P}$ is compact. Notice that $\mathcal{X}$ is closed and bounded. Then from the Heine–Borel theorem [17] it follows that $\mathcal{X}$ is compact. The probability distributions are continuous and defined on a compact set. Then from Theorem 8.4 in [17] the probability distributions are compact and thus $\mathcal{P} \subset \mathbb{R}^n$ is compact. Now we have that $\mathcal{P}$ is compact we can use Bolzano–Weierstrass [17], which states that every sequence in a compact set $\mathcal{P}$ has a convergent subsequence. Then we show that this limit is a stationary distribution.

Choose $a \in \mathcal{P}$ arbitrary. Define a sequence (a_n^T) such that:

$$(a_n^T) := \frac{1}{n} \sum_{k=1}^{n} a^T P^k = \frac{a^T P + a^T P^2 + \cdots + a^T P^n}{n}.$$

The elements of this sequence are called a *Cesàro mean* [10]. Notice that taking an average of positive vectors with elements that sum up to one is again a vector with

the same properties. And thus for all $1 \leq m \leq n$ we have $a_m^T \in \mathcal{P}$. So a_n^T has a convergent subsequence $a_{n_l}^T$ with limit:

$$\lim_{l\to\infty} a_{n_l}^T := \pi^T.$$

It suffices to show that π is a stationary distribution, in other words: $\pi^T P = \pi^T$. By continuity of the multiplication with a matrix we have that:

$$\begin{aligned}
\pi^T P &= (\lim_{l\to\infty} a_{n_l}^T)P = \lim_{l\to\infty}(a_{n_l}^T P) \\
&= \lim_{l\to\infty} \frac{1}{n_l} \sum_{j=1}^{n_l} a^T P^j P \\
&= \lim_{l\to\infty} \frac{1}{n_l} \sum_{j=2}^{n_l+1} a^T P^j \\
&= \lim_{l\to\infty} \frac{1}{n_l} \left(a^T P^{n_l+1} - a^T P + \sum_{j=1}^{n_l} a^T P^j \right)
\end{aligned}$$

By continuity of the mapping of multiplying rowvector by P is continuous, we have that $\lim_{l\to\infty}(a^T P^{n_l+1} - a^T P) = 0$. Thus

$$\begin{aligned}
\pi^T P &= \lim_{l\to\infty} \frac{1}{n_l} \sum_{j=1}^{n_l} a^T P^j \\
&= \pi^T
\end{aligned}$$

Thus π^T is a stationary distribution.
So we have shown that for a Markov Chain defined on a finite state space, there always exists a stationary distribution.

Uniqueness of stationary distribution

After some research in the literature of stationary distributions, most of the found theorems regarding the uniqueness of the stationary distribution state that for *irreducible* Markov Chains aperiodicity is required to have a *unique* stationary distribution. See for instance Theorem 3.1 in [18], the article of Zhou, Huang en Schölkopf [3] and the book 'Spectral graph theory' from Chung[12]. It is true that if a chain is irreducible and aperiodic the stationary distribution is unique.

We can give a counterexample where a directed graph is strongly connected and not aperiodic, but has a unique stationary distribution. See Figure 4.3.

The associated transition probability matrix of the graph is:

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{with} \quad P^T = \begin{bmatrix} 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Solving $\ker(P^T - \lambda I)$ with $\lambda = 1$ gives us:

$$\begin{bmatrix} -1 & 0 & 1 & 0 \\ \frac{1}{2} & -1 & 0 & 0 \\ \frac{1}{2} & 0 & -1 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

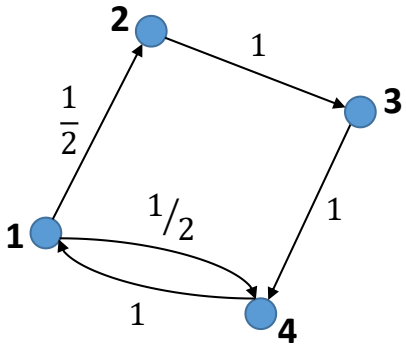


Figure 4.3: Strongly connected and periodic directed graph with period 2.

This yields in the following set of linear equations:

$$\begin{cases} x_1 - x_3 = 0 \\ x_2 - x_4 = 0 \\ x_3 - 2x_4 = 0 \\ x_4 = x_4 \end{cases}$$

Leading to the normalised eigenvector

$$\pi = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix},$$

which corresponds to eigenvalue 1 and is unique.

We assume from Theorem 7.4 from [10] that for an aperiodic irreducible Markov Chain with finite state space $\mathcal{X}$ and corresponding transition probability matrix P there exists a unique stationary distribution π described as in Section 3.2. Theorem 7.4 from [10] says that we also have:

$$\lim_{k\to\infty} P^k(i,j) \to \pi(j) \text{ for all } i,j \in \mathcal{X}. \tag{4.6}$$

So π is also the limit distribution of P. We will now show that if a Markov Chain is irreducible and is not aperiodic, the stationary distribution will still be unique.

Assume we have an irreducible Markov Chain on a finite state space $\mathcal{X}$ with period q. Then we know that a stationary distribution exists. First we will show that $\mathcal{X}$ can be partitioned into q nonempty disjoint subsets $X_0, \dots, X_{q-1}$.

Proof. Consider the Markov Chain which makes q steps at a time, so its corresponding transition probability matrix is P^q. Fix one of the subsets of $\mathcal{X}$ and name it X_0. Define X_ρ by

$$X_\rho := \{j \in \mathcal{X} | \text{ there exist } i \in X_0 \text{ and } k \in \{\rho, \rho+q, \rho+2q, \dots\} \text{ such that } P^k(i,j) > 0\},$$

with $1 \leq \rho \leq q-1$. Claim:

The finite state space $\mathcal{X}$ is the union of $X_0, \dots, X_{q-1}$ with $X_\rho \cap X_{\rho*} = \varnothing$.

Since the Markov Chain is irreducible, we know that every $i, j \in \mathcal{X}$ communicate. So it follows that $\mathcal{X} = X_0 \cup \cdots \cup X_{q-1}$. Choose ρ_1, ρ_2 and arbitrary $j_1 \in X_{\rho_1}, j_2 \in X_{\rho_2}$. From the definition of X_ρ and the irreducibility of the Markov Chain it follows that for some $i_1, i_2 \in X_0$ and k, k', k^*, k^{**} we have that the transition probabilities:

$$P^{\rho_1+kq}(i_1,j_1), P^{\rho+k'q}(i_2,j_2), P^{k^*q}(i_1,i_2), P^{k^{**}}(j_1,i_1) > 0$$

Suppose $j_1 = j_2$. Then from the transitions probabilities that the number of steps from i_1 to i_1 can be either $k_1 := \rho_1 + kq + k^{**}$ or $k_2 := k^8 q + \rho_2 + k'q + k^{**}$. The chain has period q, so state i_1 also has period q. Then from the definition of periodicity follows that k_1 and k_2 have q as their divisors. In particular, q is a divisor of $\rho_1 - \rho_2$. This is a contradiction and since we chose j_1 and j_2 arbitrary, we can conclude that $X_{\rho_1} \cap X_{\rho_2} = \varnothing$ for all $\rho_1, \rho_2 \in \{1, \ldots, q-1\}$. In the same way we can show that if you start at a state $i \in X_\rho$, you will stay in X_ρ when you make q steps. So it follows that under P^q each X_ρ is aperiodic.
Now it remains to show that the Markov Chain has a unique stationary distribution π.
Define

$$C := \lim_{k\to\infty} \frac{\sum_{i=1}^{k} P^i}{k}.$$

Notice that this is the Cesàro mean. Consider we are at $i, j \in X_\rho$. The Markov Chain on each X_ρ is irreducible and aperiodic with transition probability matrix P^q. So from Equation 4.6 we know that $\mathcal{X}_\rho$ has a limit distribution. With the continuity of the matrix multiplication we have that:

$$CP = \lim_{k\to\infty} \frac{(P^2 + \cdots + P^{k+1})}{k} = \lim_{k\to\infty} \frac{(P + \cdots + P^{k+1})}{k+1} = C,$$

with the stationary distribution π^T as its rows. So we find that $\pi^T P = \pi^T$. Notice that C is a stochastic matrix, so we have $\sum_{i=1}^{n} \pi(i) = 1$. If there is another probability distribution such that $\tilde{\pi}^T P = \tilde{\pi}$, then we have $\tilde{\pi} \frac{(P+\cdots+P^k)}{k} = \tilde{\pi}$. So $\tilde{\pi} C = \tilde{\pi}$. Thus $\tilde{\pi}^T = \pi^T$. So π^T is the unique stationary distribution of the irreducible chain with period q.

□

Now that we have showed that irreducibility is enough for a unique stationary distribution for a finite Markov Chain, we can continue to define the directed graph Laplacian. Notice that for a random walk directed graph, the graph needs to be strongly connected to have a uniwue stationary distribution. Let $G = (V, E)$ be a directed graph which is strongly connected. Let P be its associated transition probability matrix with unique stationary distribution π. Define a diagonal matrix Π with the stationary distribution on its diagonal. Then we can define a linear operator Θ on $\mathcal{H}(V)$ such that for $i, j \in V$:

$$(\Theta f)(j) = \frac{1}{2}\left(\sum_{i\in V} \frac{\pi(i)p(i,j)f(i)}{\sqrt{\pi(i)\pi(j)}} + \sum_{i\in V} \frac{\pi(j)p(j,i)f(i)}{\sqrt{\pi(j\pi(i)}} \right),$$

with associated matrix:

$$\Theta = \frac{\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}} + \Pi^{-\frac{1}{2}} P^T \Pi^{\frac{1}{2}}}{2}.$$

Then the graph Laplacian can be defined as:

$$\Delta_\Theta = I - \Theta,$$

where I is the identity matrix.
Let us take a look at the spectrum of Θ. Notice that Θ can be written as:

$$\Theta = \Pi^{\frac{1}{2}} \frac{(P + \Pi^{-1} P^T \Pi)}{2} \Pi^{\frac{1}{2}}.$$

Let Φ be $\Phi := \frac{(P+\Pi^{-1}P^T\Pi)}{2}$,then the eigenvalues of Θ are the same as the eigenvalues of Φ. Now assume that λ_ϕ is an eigenvalue of Φ with corresponding eigenvector ϕ, then $\Phi\phi = \lambda_\phi\phi$. Choose $j \in V$ such that: $|\phi(j)| = max_{i\in V}|\phi(i)|$. Then:

$$\begin{aligned}
|\lambda_\phi||\phi(j)| &= \left|\sum_{i\in V}\Phi(j,i)\phi(i)\right| \leq \sum_{i\in V}\Phi(j,i)|\phi(j)| \\
&= \frac{|\phi(j)|}{2}\left(\sum_{i\in V}p(j,i) + \sum_{i\in V}\frac{\pi(i)p(i,j)}{\pi(j)}\right) \\
&= |\phi(j)|.
\end{aligned}$$

Thus we find that $|\lambda_\phi||\phi(j)| \leq |\phi(j)|$ and thus $|\lambda_\phi| \leq 1$. So all eigenvalues of Θ lie in $[-1,1]$. Furthermore we can show that the corresponding eigenvector of the eigenvalue 1 is equal to $\sqrt{\pi}$:

$$\begin{aligned}
\Theta\sqrt{\pi} &= \frac{1}{2}\left(\sum_{i\in V}\frac{\pi(i)p(i,j)\sqrt{\pi(i)}}{\sqrt{\pi(i)\pi(j)}} + \sum_{i\in V}\frac{\pi(j)p(j,i)\sqrt{\pi(i)}}{\sqrt{\pi(j)\pi(i)}}\right) \\
&= \frac{1}{2}\left(\frac{1}{\sqrt{\pi(j)}}\sum_{i\in V}\pi(i)p(i,j) + \sqrt{\pi(j)}\sum_{i\in V}\pi(j)p(j,i)\right) \\
&= \sqrt{\pi(j)}.
\end{aligned}$$

So we have that $\Theta\sqrt{\pi} = 1\cdot\sqrt{\pi}$ and thus the eigenvector associating with eigenvalue 1 is: $\sqrt{\pi}$.

5 APPLICATIONS OF DIRECTED GRAPH LAPLACIANS

In this chapter we will apply directed spectral clustering using the definition of the directed graph Laplacian defined in Section 4.2.2.

5.1 DIRECTED SPECTRAL CLUSTERING

In [3] an algorithm for directed spectral clustering is given. The idea is to partition the vertices of a directed graph $G = (V, E)$ into k parts which have similarities between them. Note that the input is a weighted adjacency matrix, which will be visualised as a graph. Furthermore, the data we use is self-constructed unless stated otherwise.

5.1.1 Directed spectral clustering in 2 clusters

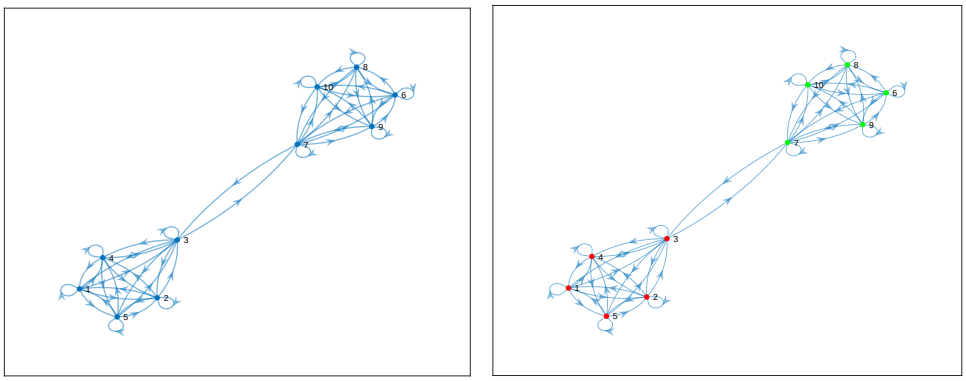
The algorithm to cluster a strongly connected directed graph $G = (V, E)$ into two clusters consists of 3 steps:

1. First a random walk over G is defined with a transition probability matrix P, such that it has a unique stationary distribution π.
2. The next step in the algorithm is computing Θ as described in Section 4.2.2.
3. Compute the eigenvector θ of Θ corresponding to the second largest eigenvalue λ_2 and then partition V of G into two parts such that: $S = \{i \in V \mid \theta(i) \geq 0\}$ and $S^c = \{i \in V \mid \theta(i) < 0\}$

From Section 4.2.2 we know that the eigenvalues of Θ are scattered in $[-1, 1]$.

Directed graph with 10 *nodes*

First we will start with a small graph with 10 nodes. See Figure 5.1.

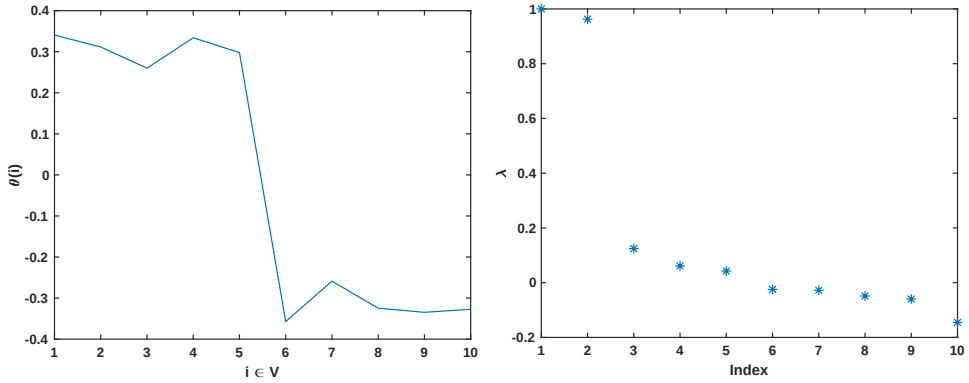


(a) Directed graph with 10 nodes

(b) Subsets S (red) and S^c (green)

Figure 5.1: Directed graph with $n = 10$ nodes containing edge weights $w_{ij} \in \{5, 6, 7, 8, 9, 10\}$ with $i, j = 1, \ldots, 10$

On the left in Figure 5.1a we see the constructed strongly connected directed graph with edge weights equal to randomly chosen integers between 5 and 10. On the right in Figure 5.1b we see the subsets S and S^c, as described in the algorithm, in red and green respectively.



(a) Eigenvector $\theta(i)$, with $i = \{1, \ldots, 10\}$.

(b) Eigenvalues $\lambda_1, \ldots, \lambda_{10}$ of Θ.

Figure 5.2: Eigenvectors and eigenvalues

In Figure 5.2b we see that there are two eigenvalues close to 1. Notice that we have plotted the sorted eigenvalues in descending form. In Figure 5.2a we see that for nodes $1, \ldots, 5$ it follows that $\theta \geq 0$. For $6, \ldots, 10$ we have that $\theta < 0$. We can also see a slight peak at node 3 and 7, this is because there are edges between them even though they belong to different clusters.

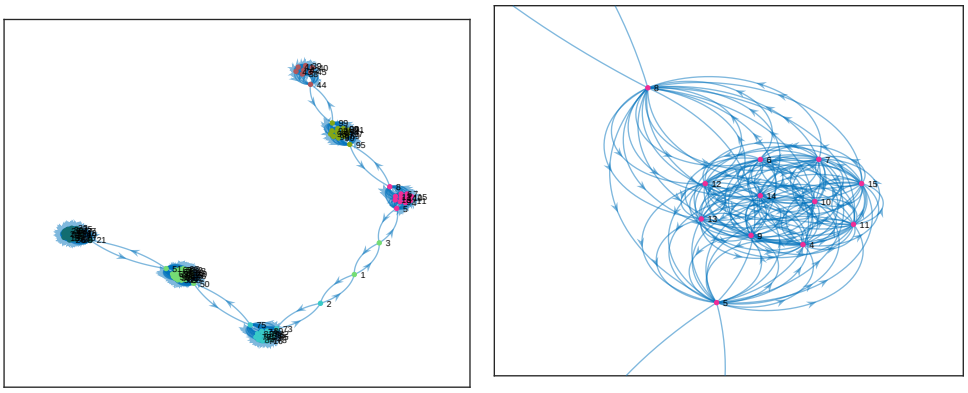
5.1.2 Directed spectral clustering in k clusters

The algorithm to cluster a directed graph $G = (V, E)$ into k clusters has almost the same structure:

1. First a random walk over G is defined with a transition probability matrix P, such that it has a unique stationary distribution π.
2. The next step in the algorithm is computing Θ as described in Section 4.2.2.
3. Compute the k eigenvectors corresponding to the k largest eigenvalues λ_i of Θ and create $K \in \mathbb{R}^{n \times k}$ such that its columns are the eigenvectors are the first k eigenvectors of Θ.
4. Use *k-means* algorithm to partition V of G into clusters $V_1 \cup \cdots \cup V_k$.

Directed graph with 6 clusters

Take a look at the strongly connected directed graph in Figure 5.3a. The graph contains 6 subsets of nodes such that they have weighted directed edges between every two nodes in that subset. In Figure 5.3b we zoomed in on one of the subsets. Between each of those subsets we define edges such that the whole graph is strongly connected. The nodes which belong to the same cluster have the same color, as can be seen in Figure 5.3a

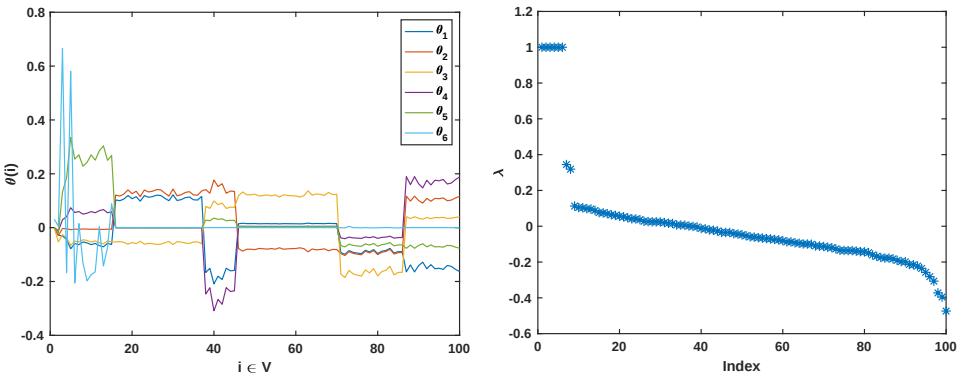


(a) Directed graph with 6 clusters.

(b) Zoomed in on one of the 6 subsets of the directed graph on the left.

Figure 5.3: Strongly connected directed graph with 6 clusters.

In Figure 5.4b we see that there are 6 eigenvalues very close to 1. Applying the algorithm by choosing $k = 6$, gives us 6 clusters, which are highlighted in different colors in Figure 5.3a.



(a) Eigenvectors for the nodes $i \in V$

(b) Eigenvalues of Θ

Figure 5.4: The first 6 eigenvectors and the eigenvalues of Θ.

In Figure 5.4a we have plotted the first 6 eigenvectors of Θ. We see that each eigenvector has a turning point when defined on nodes from a different subset.

It seems that if we make strongly connected graphs with subsets such that they are even stronger connected in between the nodes of that subset, the number of eigenvalues close to 1 are equivalent to the number of clusters in that graph.

If we combine two of the subsets by defining an edge between all of the nodes of the two subsets, in particular, the nodes $\{16, \ldots, 45\}$, we get Figure 5.5, with $k = 5$. We see that now we have 5 eigenvalues which are almost equal to 1, see Figure 5.6b. Also the first 5 eigenvectors show that the V could be divided in 5 subsets.

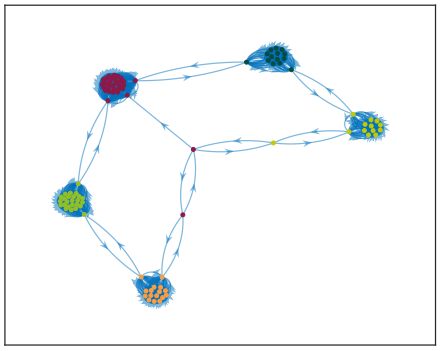
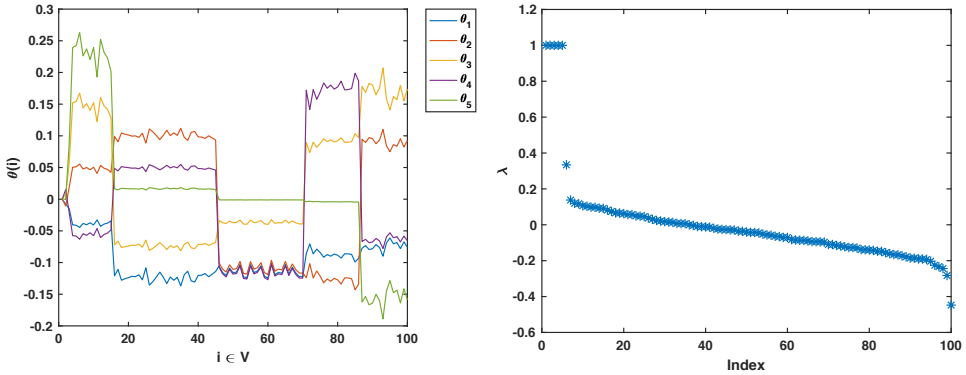


Figure 5.5: Directed graph with 5 clusters



(a) Eigenvectors for the nodes $i \in V$

(b) Eigenvalues of Θ

Figure 5.6: The first 6 eigenvectors and the eigenvalues of Θ.

Migration in the United States of America

We will apply the algorithm on a data set which shows the migration between states in the USA, see Figure 5.7. The data set has the form of a weighted adjacency matrix. The graph representation of the data set is a strongly connected directed graph. The eigenvalues of the graph can be seen in Figure 5.8b.

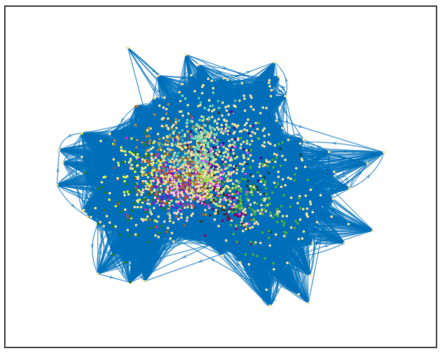
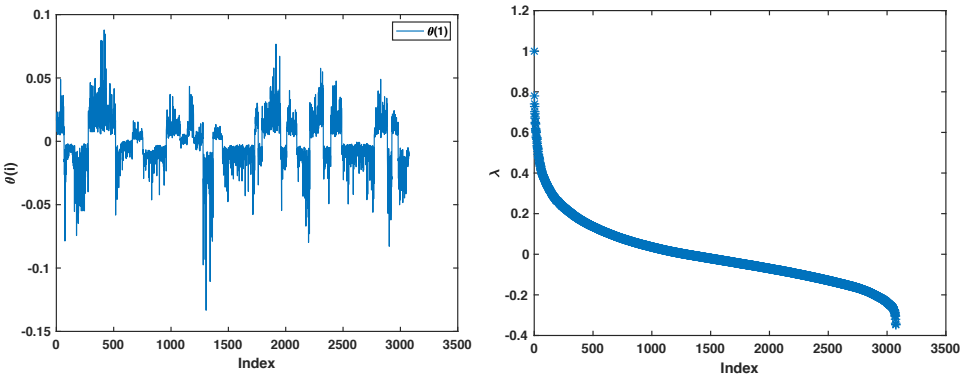


Figure 5.7: States of the USA clustered by the migration data.

We see that there is just one eigenvalue close to 1 which is 1 itself. From the eigenvector corresponding to the second largest eigenvalue, it seems like there are 26 clusters. So we apply the algorithm by using $k = 26$. There is no structure in the eigenvalues nor in the figure for which we can take conclusions. This could have something to do with that there is not enough difference in the migrations between the states.



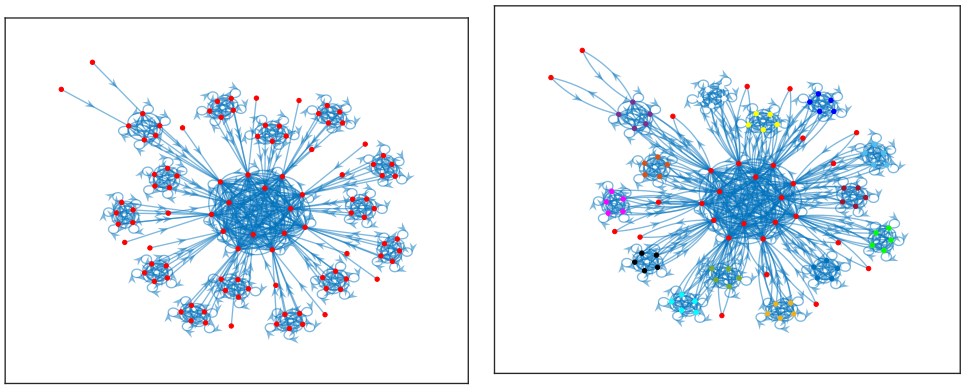
(a) Eigenvectors for the nodes $i \in V$

(b) Eigenvalues of Θ

Figure 5.8: The first 6 eigenvectors and the eigenvalues of Θ.

Directed graph which is not strongly connected

What if a graph is not strongly connected? Take a look at Figure 5.9a. This graph shows the websites which have links to or from https://www.mathworks.com. The nodes are a representation of the websites and the edges are links between the websites. This is an existing data set in MATLAB. Notice that this graph is not strongly connected. So that means we cannot apply our algorithm since there is no unique stationary distribution associated with its transition probability matrix.

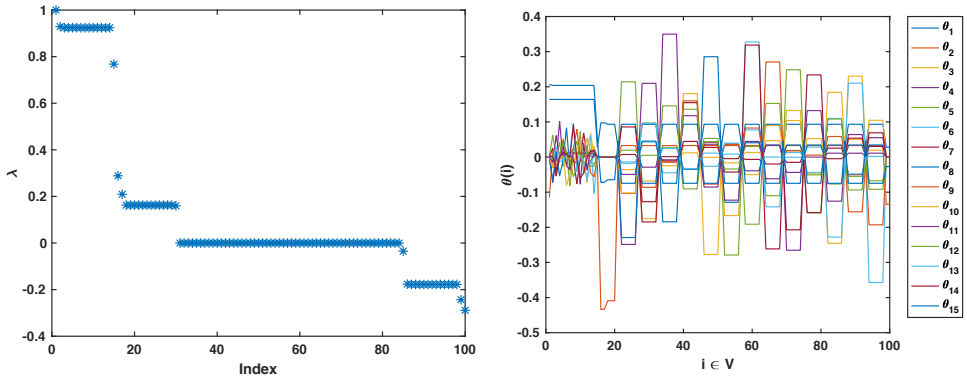


(a) Eigenvectors for the nodes $i \in V$

(b) Strongly connected directed graph corresponding to the graph on the left.

Figure 5.9: Graph representation of the websites which have links to https://www.mathworks.com

We made the graph symmetric, by adding its transpose to its weighted adjacency matrix and dividing by two. The graph is also strongly connected now. In Figure 5.9b you can see the 15 different clusters for the symmetric graph shown by different colored nodes. In Figure 5.10a we see that there are 15 eigenvalues which are almost equal to one, which again corresponds with the number of cluster such that all nodes in the clusters are adjacent.



(a) Eigenvectors for the nodes $i \in V$

(b) Eigenvalues of Θ

Figure 5.10: The first 6 eigenvectors and the eigenvalues of Θ.

6 CONCLUSIONS AND RECOMMENDATIONS

After our research on directed graph Laplacians, we can answer our research questions:

What definitions can we give for a graph Laplacian on directed graphs? What are their properties? How can the graph Laplacian be used on spectral clustering methods?

Let $G = (V, E)$ be a directed graph on V. The first definition that we can give for the directed graph Laplacian is:

$$\begin{aligned}\Delta_\delta &= \delta\Delta_{V_{in}} + (1-\delta)\Delta_{V_{out}} \\ &= \delta(I - D_{V_{in}}^{-1/2} W_{V_{in}} D_{V_{in}}^{-1/2}) + (1-\delta)(I - D_{V_{out}}^{-1/2} W_{V_{out}} D_{V_{out}}^{-1/2}),\end{aligned}$$

with $D_{V_{in}}$, $W_{V_{in}}$ and $D_{V_{out}}$, $W_{V_{out}}$ the degree and weight matrix of the the graphs G_{in} and G_{out} respectively. The matrices G_{in} and G_{out} are as described in Section 4.2.1. We have seen that the graph Laplacian on G_{in} and G_{out} is positive semi-definite and has eigenvalues between 0 and 1.

Let $G = (V, E)$ be a strongly connected directed graph. Then the second definition for the graph Laplacian will be defined as:

$$\Delta_\Theta = I - \Theta,$$

with

$$\Theta = \frac{\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}} + \Pi^{-\frac{1}{2}} P^T \Pi^{\frac{1}{2}}}{2},$$

with P the transition probability matrix and Π the diagonal matrix with the unique stationary distribution π of the random walk on a G. We have shown that there exist a unique stationary distribution for the random walk on a directed graph if the graph is strongly connected.
The matrix Θ has eigenvalues between -1 and 1 with the eigenvector which corresponds to the eigenvalue 1 equal to $\sqrt{(}\pi)$.

In Chapter 5 we have defined an algorithm for directed spectral clustering. The results seem to show that there is a correlation between the multiplicity of the eigenvalues really close to 1 and the number of subsets of V such that all the nodes in those subsets are adjacent. There is no proper proof given for this and yet there is not enough research done on these types of graphs to assume that this is true.

FURTHER RESEARCH

The directed graph Laplacian applied on spectral clustering is not a simple problem. There is still a lot of research that remains to be done on this topic. Therefore, we propose some ideas for further research. One statement which can be either proved or disproved is:

1. Let $G = (V, E)$ be a strongly connected directed graph. Then the number of eigenvalues, really close to or equal to 1, gives the number k of disjoint strongly connected clusters $A_1, \ldots, A_k$ such that for all $i, j \in A_l$ we have that (i, j) and $(j, i) \in E$, with $1 \leq l \leq k$.

Another idea for further research is:

2. Are there different spectral clustering algorithms for directed graphs which are not strongly connected?

BIBLIOGRAPHY

[1] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[2] Dengyong Zhou, Thomas Hofmann, and Bernhard Schölkopf. Semi-supervised learning on directed graphs. In *Advances in neural information processing systems*, pages 1633–1640, 2004.

[3] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *Proceedings of the 22nd international conference on Machine learning*, pages 1036–1043, 2005.

[4] Abbe, Emmanuel. Community detection and the stochastic block model. https://www.semanticscholar.org/paper/Community-detection-and-the-stochastic-block-model-Abbe/40922ef4df50dc1fea9d8eeb5e209325641857ab?sort=is-influential, 2016. Accessed 26 July 2021.

[5] Tallis, Shannon. Data is the 21st century's oil, says Siemens CEO Joe Kaeser. https://economictimes.indiatimes.com/magazines/panache/data-is-the-21st-centurys-oil-says-siemens-ceo-joe-kaeser/articleshow/64298125.cms?from=mdr. Accessed 22 July 2021.

[6] Brittanica. Computer software. https://www.britannica.com/topic/information-system/Computer-software#ref1242521. Accessed 20 July 2021.

[7] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Stanford digital library technologies project, Standford University, January 1998.

[8] Francis R. Bach and Micheal I. Jordan. Learning Spectral Clustering, With Application to Speech Seperation. *Journal of Machine Learning Research*, 7(71):1963–2001, September 2006.

[9] Daniel A. Spielman. Lecture notes in Random Walks on Graphs, October 2018. Yale University. https://www.cs.yale.edu/homes/spielman/561/lect10-18.pdf. Accessed 25 June 2021.

[10] Ehrhard Behrends. *Introduction to Markov chains*, volume 228. Springer, 2000.

[11] Bojan Mohar, Yousef Ed. Alavi, Gary Chartrand, and Ortrud R. Oellermann. The Laplacian spectrum of graphs. *Graph theory, combinatorics, and applications*, 2(871-898):12, 1991.

[12] Fan Chung. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[13] Fan Chung. Laplacians and the Cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005.

[14] Jon M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46:604–632, September 1999.

[15] Chris Ding, Xiaofeng He, Parry Husbands, Hongyuan Zha, and Horst Simon. PageRank, HITS and a unified framework for link analysis. In *Proceedings of the 2003 SIAM International Conference on Data Mining*, pages 249–253. SIAM, 2003.

[16] Nikolaĭ N. Bogoliubov and Jurij A. Mitropol'skii. *Asymptotic methods in the theory of non-linear oscillations*, volume 10. CRC Press, 1961.

[17] Neal L. Carothers. *Real analysis*. Cambridge University Press, 2000.

[18] Walter R. Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.