BSc report APPLIED MATHEMATICS

"The electric network approach to the study of Markov Chains."

(Dutch title: "De elektrische netwerk methode voor het onderzoeken van Markov ketens.")

Mirte van Loenen

Delft University of Technology

Supervisor

Dr. Elena Pulvirenti

Other committee members

Prof. Arnold Heemink

Dr. Matthias Möller

June 25th, 2021

Delft

Abstract

Markov chains are used to describe random processes in discrete time, which have the property of being memoryless. This report covers Markov chains on a finite space that are homogeneous in time and mainly follows the structure of "Markov Chains and Mixing Times" [1]. Markov chains exhibit a strong connection with electric networks. We exploit such a connection and apply the laws of physics to answer several probabilistic questions about random walks, which are certain types of Markov chains. This connection is then called the electric network approach and is built on translating the random walk into an electric network by relating the transition probability to the so-called conductance. The electric network approach provides problem simplification tools, such as the Series/Parallel Law, and powerful inequalities, such as the Nash-Williams inequality. These tools are based on physics and are often more intuitive than their probabilistic counterpart.

We simulate a two-dimensional random walk that starts at the center of a square and "escapes" if it reaches the perimeter of the square before returning to the center. We then compare this escape probability to an upper bound, which results from using the Nash-Williams inequality. The sharpness of the upper bound depends on the choice of edge-cutsets. We find that choosing edge-cutsets with a minimal amount of edges gives a sharper upper bound, than choosing edge-cutsets that contain all the edges of the square. The relation between the upper bound and the escape probability seems to be independent of the size of the square. Furthermore, we provide proofs that are not explicit in [1] and [5] and add to the contents of [1] by studying random walks from a graph theory perspective.

Layman abstract

In this report we study two-dimensional random walks, which are an example of more general probabilistic processes called Markov chains. Random walks are random processes where you go from a point to a neighbouring point, chosen uniformly at random, during each step in time. We are interested in computing the "escape probability" of the random walk starting from the center of a square, i.e. the probability of reaching the perimeter of the square before returning to the center again. We address this problem first by using the so-called electric network approach, which uses physics laws to study random walks. Then we focus on the Nash-Williams inequality, which provides an upper bound of the escape probability. By simulating the random walk, we test how sharp this upper bound is for different sizes of squares.

Introduction

Markov chains are used to describe random processes in discrete time, which have the property of being memoryless. This report covers Markov chains on a finite space that are homogeneous in time and mainly follows the structure of "Markov Chains and Mixing Times" [1]. Markov chains exhibit a strong connection with electric networks. We will exploit such a connection and apply the laws of physics to answer several probabilistic questions. This report then has the following research question:

When do we use the electric network approach and in what way does this approach contribute to the understanding of Markov chains?

Markov chains were first researched by Russian mathematician Andrej Markov, who in 1907 published "Investigation of a remarkable example of dependent trials" (translated) [3], [2]. Markov chains have become an integral part of modern probability theory. An example is the random walk, which is a Markov chain on a graph going from a vertex to one of its neighbouring vertices by choosing it uniformly at random. In 1984 Peter Doyle and Laurie Snell first introduced the relation between random walks and electric networks in "Random walks and electric networks" [4]. The use of electric networks to study random walks is called the electric network approach. We will study and use this approach throughout the report.

The report is divided into two sections. Section 1 starts with introducing and defining Markov chains. In particular, in Section 1.1 we introduce the main properties of Markov chains: irreducibility, aperiodicity and the stationary distribution. In Section 1.2 we introduce expectations and harmonic functions. In Section 1.3 we give the definition of hitting time. In Section 1.4 we combine the results of the previous sections to study the uniqueness and convergence of the stationary distribution as well as to introduce reversible Markov chains. In Section 1.5 we relate Markov chains to graphs by introducing the random walk and we use graph theory to determine the properties of random walks. We also relate reversibility to random walks and give an example of a random walk on a network, on which we apply the results from 1.4.

The first section provides all background information needed in the second section. In the second section we introduce and explore the electric network approach, while focussing only on random walks. In Section 2.1 we introduce the electric network approach. In Section 2.2 we introduce the effective resistance and study some of its applications. In Section 2.3 we give three physics laws and give an example of their applications. In Section 2.4 we study the Nash-Williams inequality and test how sharp this bound is, via a simulation of a random walk. Although the electric network interpretation is based on physics, we will solely discuss the mathematical aspects.

Our contribution consists of original examples, detailed proofs that are not explicit in [1] and [5] and studying the Nash-Williams inequality by simulating a random walk on a grid.

Contents

Abstract i

Introduction iii

1 Markov Chains 1

1.1 Irreducibility, periodicity and stationary distributions 2

1.2 Expectations and harmonic functions 4

1.3 Hitting times 5

1.4 Uniqueness and convergence of the stationary distribution 5

1.5 Random walks on graphs 7

2 Electric Networks 13

2.1 Current flows 14

2.2 Effective resistance and results 17

2.3 Network reduction laws 18

2.4 The Nash-Williams inequality 21

Conclusion and discussion 25

Bibliography 27

Appendix A 29

1

Markov Chains

A Markov chain is a probabilistic process in discrete time, that moves among elements of a state space Ω and does so independently of its history. For a **finite** Markov chain, the state space Ω is finite and the next step from $x \in \Omega$ is chosen according to a probability distribution $P(x,\cdot)$. If this probability distribution is independent of time, i.e. is the same for every time t, we say that the Markov chain possesses the **time homogeneity property**, also called a homogeneous Markov chain [6]. In this project we always refer to finite homogeneous Markov chains, when talking about Markov chains.

This section of the project closely follows the definitions and the structure of the first section of the book "Markov Chains and Mixing Times" by David Levin, Yuval Peres and Elizabeth Wilmer [1]. Everything that is not proven here is proven in the book.

Let us start with the following key definition:

Definition 1.0.1. A sequence of random variables $(X_0, X_1, \dots)$ is a ***finite homogeneous Markov chain*** with state space Ω and transition matrix P, if $\forall x, y \in \Omega$, all $t \geq 1$ and all events $H_{t-1} = \cap_{s=0}^{t-1}\{X_s = x_s\}$ such that $\boldsymbol{P}(H_{t-1} \cap \{X_t = x\}) > 0$, then

$$\boldsymbol{P}\{X_{t+1} = y \,|\, H_{t-1} \cap \{X_t = x\}\} = \boldsymbol{P}\{X_{t+1} = u \,|\, X_t = x\} = P(x,y). \tag{1.1}$$

Equation (1.1) is called the ***Markov property***.

The next step of the chain is only dependent on the current state in Ω and the transition matrix, which stores all the transition probabilities of the Markov chain.

Definition 1.0.2. The ***transition matrix*** P is a stochastic $|\Omega|$ x $|\Omega|$ matrix, where the x-th row is the distribution $P(x,\cdot)$, all entries are non-negative and

$$\sum_{y \in \Omega} P(x,y) = 1, \; \forall x \in \Omega.$$

Where $P(x,y)$ is the probability of going from state x to y in a single step, $P(x,y) = \boldsymbol{P}\{X_{t+1} = y \,|\, X_t = x\}$.

Let us take a look at a Markov chain and construct its transition matrix.

Example 1.0.3. Alice (a) and Bob (b) find a coin. Alice starts tossing the coin and gives it to Bob if the coin lands on tails, but keeps tossing otherwise. When Bob receives the coin, he does the same, giving the coin back to Alice when he tosses tails.

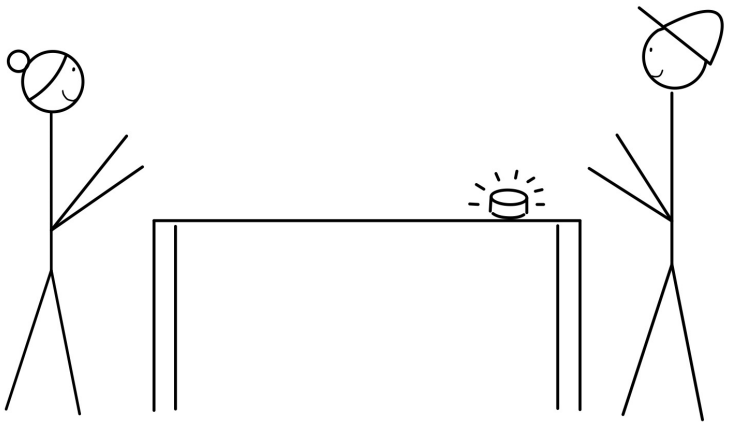

Figure 1.1: Alice (left) and Bob (right) are tossing a coin. When the person with the coin tosses tails, the coin goes to the other person and the game continues.

The coin has a probability p to land on heads and $1-p$ to land on tails. This process has the following transition matrix

$$P = \begin{pmatrix} P(a,a) & P(a,b) \\ P(b,a) & P(b,b) \end{pmatrix} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \tag{1.2}$$

The ***initial distribution*** μ_0 is a row vector of size $|\Omega|$. For Alice and Bob μ_0 has size two and $\mu_0 = (\mathrm{P}\{X_0 = a\}, \mathrm{P}\{X_0 = b\})$. Alice has the coin at the start, so we write $\mu_0 = [1,0]$.

The distribution after one coin flip is $\mu_1 = (p, 1-p)$, which is equal to $\mu_0 P$. After t coin flips, or at time t, the probability distribution is equal to $\mu_t = \mu_0 P^t$.

If a Markov chain starts in a single state, like it did for Alice and Bob, the initial distribution is written as

$$\delta_x(y) = \begin{cases} 1 \text{ if } y = x, \\ 0 \text{ if } y \neq x. \end{cases} \tag{1.3}$$

In this case the starting state is added as a subscript to the probability and expectation, $\mathbf{P}$ and $\mathbf{E}$, of the chain. For example, $\mathbf{P}_a\{X_t = b\}$ is the probability that the chain is in state b at time t if it starts in state a.

In general, we have that $P^t(x,y) = \mathbf{P}\{X_t = y \,|\, X_0 = x\}$, which is the probability of going from x to y in time t. The distribution at time t stores information on the current possible states, $\mu_t(x) = \mathbf{P}\{X_t = x\}$, $\forall x \in \Omega$.

1.1. Irreducibility, periodicity and stationary distributions

Because the definition of a Markov chain does not impose many restrictions on the state space or transition matrix, there is a substantial variety of different Markov chains. To further examine Markov chains we ask ourselves questions like: "Can this Markov chain go from this state to that state?" and "What will happen with the chain over time?". In this section of the report we introduce and discuss everything needed to answer both questions.

To answer the first question, we will introduce the notion of irreducible Markov chains.

Definition 1.1.1. A chain P is ***irreducible*** if

$$\forall x, y \in \Omega, \exists t : P^t(x, y) > 0. \tag{1.4}$$

If a Markov chain is irreducible, it can go from one element of the state space Ω to any other state in finitely many steps. Irreducible chains have very interesting properties, which we will discuss throughout the report.

It is interesting to examine when the chain is able to return to its starting state x, i.e. when $P^t(x, x) > 0$. For this we look at the period of a state.

Definition 1.1.2. Let $\mathcal{T}(x) := \{t \geq 1 : P^t(x, x) > 0\}$ be the set of times when it is possible for the chain to return to its starting state x. The greatest common divisor (gcd) of $\mathcal{T}(x)$ is called the ***period of*** x.

For large state spaces, it is inefficient to study the period of every state. However, this is no problem for irreducible chains. The period is related to irreducibility by the following lemma.

Lemma 1.1.3. If P is irreducible, then $\gcd\mathcal{T}(x) = \gcd\mathcal{T}(y)$, $\forall x, y \in \Omega$, i.e. all states have the same period.

As the period is constant for all states of an irreducible chain, we call this period the **period of the chain**. An irreducible chain is either periodic or aperiodic, depending on its period.

Definition 1.1.4. An irreducible Markov chain is ***aperiodic*** if its period is equal to 1, and ***periodic*** otherwise.

In Definition 1.1.1, the time t at which $P^t(x, y) > 0$ could depend on the states x and y. The next proposition combines aperiodicity and irreducibility.

Proposition 1.1.5. If P is aperiodic and irreducible, then there is an integer r such that $P^r(x, y) > 0$, $\forall x, y \in \Omega$.

Therefore, there exists a time r where it is possible to reach y if the chain started at x and this time is independent of state x and y.

Irreducibility and aperiodicity tell us something about the chain and its transition matrix. We will come back to this, but first we discuss probability distributions.

In Example 1.0.3 we introduced the notion of a row vector μ_t containing the distribution at time t. At time $t+1$ one has $\mu_{t+1} = \mu_t P$, which means that most distributions change over time. If a distribution does not change over time, it is called a stationary distribution.

Definition 1.1.6. The ***stationary distribution*** π is a distribution on Ω such that

$$\pi = \pi P. \tag{1.5}$$

In linear algebra, this means that for a transition matrix P, π^{T} is a normal eigenvector of P^{T} with eigenvalue 1.

If μ_0 is stationary, it will remain the same at every time t. The stationary distribution encodes information about a Markov chain. We will use this information to answer the second question (see beginning of 1.1), but before we examine this encoded information, let us look at a specific stationary distribution.

If the stationary distribution is uniform, the chain has the same probability of being in any state in Ω for every following step. An easy to spot case of a uniform stationary distribution is when the transition matrix is symmetric.

Definition 1.1.7. A transition matrix P is ***symmetric*** if $P(x,y) = P(y,x)$ for all $x, y \in \Omega$.

The proof of the following proposition is left as an exercise in [1]. Below we give the detailed proof.

Proposition 1.1.8. If P is symmetric, then the uniform distribution on Ω is stationary for P.

Proof. Let $\pi = \frac{1}{|\Omega|} = \frac{1}{u}$ be the uniform distribution. Take $x \in \Omega$ arbitrary, then using that P is symmetric (sym)

$$\pi P(x) = \sum_{y\in\Omega} \frac{1}{u} P(y,x) \overset{\text{sym}}{=} \sum_{y\in\Omega} \frac{1}{u} P(x,y) = \frac{1}{u} \sum_{y\in\Omega} P(x,y) = \frac{1}{u} \cdot 1 = \pi(x).$$

Since x was taken arbitrarily, $\pi P(x) = \pi(x)$ for every $x \in \Omega$, thus $\pi P = \pi$. □

Another way to check if a probability distribution is stationary is by using the **detailed balance equations**,

$$\pi(x)P(x,y) = \pi(y)P(y,x), \ \forall x, y \in \Omega. \tag{1.6}$$

Proposition 1.1.9. Let P be the transition matrix of a Markov chain with state space Ω. Any distribution π satisfying the detailed balance equations (1.6) is stationary for P.

The proof of Proposition 1.1.9 is similar to the proof of Proposition 1.1.8 and is also given in [1].

The stationary distribution is an integral part of the study of Markov chains. In Section 1.4, we will combine all the contents of this section with the next two sections to showcase what information the stationary distribution provides.

1.2. Expectations and harmonic functions

So far we have discussed the evolution of probability distributions by multiplying a row vector by a transition matrix on the right. This multiplication takes a probability distribution to the probability distribution at the next step. What happens if we multiply a column vector by the same transition matrix on the left?

Take a function on the state space $f : \Omega \to \mathbb{R}$. Then the x-th entry of Pf is

$$Pf(x) = \sum_{y\in\Omega} P(x,y) f(y) = \sum_{y\in\Omega} \mathbf{P}_x\{X_1 = y\} f(y) = \mathbf{E}_x(f(X_1)). \tag{1.7}$$

Therefore, multiplying a column vector by a transition matrix on the left takes a function on the state space to the expected value of the function at the following step.

We saw that if a distribution is invariant under right multiplication, it is called stationary. If a function is invariant under left multiplication, it is called harmonic.

Definition 1.2.1. A function $h : \Omega \to \mathbb{R}$ is ***harmonic at*** $x \in \Omega$ if

$$h(x) = \sum_{y\in\Omega} P(x,y) h(y) \tag{1.8}$$

and ***harmonic on*** $S \subset \Omega$ if h is harmonic at every element of S. If h is ***harmonic on*** Ω, then $Ph = h$.

This means that every eigenvector of a transition matrix P with eigenvalue 1 is harmonic.

For irreducible chains, harmonic functions have a specific form.

Lemma 1.2.2. Suppose that P is irreducible. If a function h is harmonic at every point of Ω, then h is constant.

Lemma 1.2.2 implies that the eigenspace of eigenvalue 1 of P has dimension 1. This is a powerful result and will be used to prove Proposition 1.4.1.

We have now introduced the building blocks of Markov chains needed to prove the two main results in Section 1.4. Before we discuss these results we first introduce the concept of hitting times.

1.3. Hitting times

One of the main goals while studying a Markov chain is to answer questions such as "When is it possible to reach this state?", or "Will this chain go back to its starting point before exiting from a certain domain?" We first introduce some notations. In Section 2 these notations are used to answer these questions.

Definition 1.3.1. The ***hitting time*** of $x \in \Omega$ is equal to

$$\tau_x := \min\{t \geq 0 : X_t = x\}.$$

i.e. the first time a Markov chain is in the state x. When only looking at positive time, we define

$$\tau_x^+ := \min\{t \geq 1 : X_t = x\}.$$

τ_x^+ is called the ***first return time*** if $X_0 = x$.

If a Markov chain returns to its starting state at time t_r, the process practically starts over, i.e. for $t \geq t_r$, $\mu_t = \mu_{t-t_r}$. The probability of reaching a certain state before the process starts over is called the escape probability.

Definition 1.3.2. The ***escape probability*** is the probability of escaping to z before returning to the starting state a, and is denoted by $\boldsymbol{P}_a\{\tau_z < \tau_a^+\}$.

Finding the escape probability often requires extensive calculations, especially on larger state spaces. In Section 2 we will apply the electric network approach to replace these calculations with simpler and more intuitive methods by using the electric network approach.

1.4. Uniqueness and convergence of the stationary distribution

We use the notation of the previous sections to introduce two key properties of the stationary distribution: uniqueness and convergence. We also look at the time reversal of an irreducible Markov chain.

First, the stationary distribution of an irreducible Markov chain is unique and is related to the expected first return time.

Proposition 1.4.1. Let P be the transition matrix of an irreducible Markov chain. Then

1. there exists a unique probability distribution π on Ω such that $\pi = \pi P$ and $\pi(x) > 0$ for all $x \in \Omega$, and moreover,
2. $\pi(x) = \frac{1}{\mathbf{E}_x(\tau_x^+)}$.

In [1], the uniqueness is shown by using Lemma 1.2.2 to get that the dimension of the eigenspace of eigenvalue 1 is equal to 1 and deriving that the dimension of the solution space of $\mu = P\mu$ is

also 1. This one-dimensional solution space only contains one vector whose entries sum up to 1, which makes the stationary distribution unique.

Most of the time, it is easy to check whether a chain is irreducible by studying the transition matrix. It is more complex, however, to check whether a stationary distribution is unique or to find the expected return times $\boldsymbol{E}_x(\tau_x^+)$, which makes Proposition 1.4.1 very practical.

We will now look at the time reversal and the convergence of irreducible chains using the uniqueness of the stationary distribution and Proposition 1.4.1.

In Section 1.1 we introduced the detailed balance equations (1.6). The following definition is from "Reversible Markov Chains and Random Walks on Graphs" [5].

Definition 1.4.2. Let (X_t) be a finite irreducible Markov chain with transition matrix P and stationary distribution π. Call the chain ***reversible*** if

$$\pi(x)P(x,y) = \pi(y)P(y,x),\ \forall x,y \in \Omega. \tag{1.9}$$

In other words, if the stationary distribution of an irreducible chain satisfies the detailed balance equations (1.6) it is reversible.

We use the term reversible, because a reversible chain has the same transition matrix for the time reversal. The time reversal of an irreducible chain is defined as follows.

Definition 1.4.3. The ***time reversal*** of an irreducible Markov chain with transition matrix P and stationary distribution π is the chain with transition matrix

$$\hat{P}(x,y) := \frac{\pi(y)P(y,x)}{\pi(x)}. \tag{1.10}$$

It is clear that if the chain is reversible $\hat{P} = P$. If a chain is reversible, its time reversal has the same distribution, i.e.

$$\mathbf{P}_\pi\{X_0 = x_0, X_1 = x_1, \cdots, X_t = x_t\} = \mathbf{P}_\pi\{X_0 = x_t, X_1 = x_{t-1}, \cdots, X_t = x_0\}. \tag{1.11}$$

Convergence of an irreducible Markov chain.

In Section 1.1 we asked the question: "What will happen with the chain over time?" For irreducible aperiodic chains, the answer to this question is actually very clear. To show this we will combine the properties of irreducibility, aperiodicity and the concept of the stationary distribution. In order to do this, we need to be able to compare probability distributions, which is done using the total variation distance (Section 4 in [1]).

Definition 1.4.4. The ***total variation distance*** between two probability distributions μ and ν on Ω is defined by

$$\|\mu - \nu\|_{TV} := \max_{A\subset\Omega} |\mu(A) - \nu(A)|. \tag{1.12}$$

Having a way to compare two different distributions is useful, but using subsets when comparing distributions is not ideal. To avoid using subsets we use the following proposition.

Proposition 1.4.5. Let μ and ν be two probability distributions on Ω. Then

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)|. \tag{1.13}$$

Now to show what will happen with an irreducible aperiodic Markov chain over time, we look at the convergence theorem. The following theorem and its proof are part of Section 4.3 of [1].

Theorem 1.4.6 (Convergence Theorem)**.** Suppose that P is irreducible and aperiodic, with stationary distribution π. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that

$$\max_{x\in\Omega} \left\| P^t(x,\cdot) - \pi \right\|_{TV} \leq C\alpha^t. \tag{1.14}$$

What the convergence theorem states is that, for an irreducible aperiodic chain, every row of P^t converges in total variation to a stationary distribution, as $C\alpha^t$ converges to zero. Therefore, for all $x, y \in \Omega$, $P^t(x,y)$ converges to $\pi(y)$ in total variation.

The convergence of P^t does not depend on the initial distribution, i.e. $\mu_t = \mu_0 P^t$ converges to π as time grows for any initial distribution μ_0. To show this, take an arbitrary initial distribution μ_0 and use $\sum_{x\in\Omega} \mu_0(x) = 1$ to get

$$\mu_t(x) = \mu_0 P^t(x) = \sum_{y\in\Omega} \mu_0(y) P^t(y,x) \overset{TV}{\to} \sum_{y\in\Omega} \mu_0(y)\pi(x) = \pi(x). \tag{1.15}$$

By Proposition 1.4.1, the π in (1.15) is the unique stationary distribution, which means that for aperiodic irreducible chains we know that for any initial distribution, the chain converges to a unique distribution π, with $\pi(x) = \frac{1}{\boldsymbol{E}_x(\tau_x^+)}$.

In this report, we only discuss transition matrices, but irreducible aperiodic matrices (called primitive matrices) are also a subject in linear algebra. For the linear algebra equivalent of the convergence theorem, which is for general primitive matrices, see [7].

1.5. Random walks on graphs

Finite Markov chains can be represented as graphs, where the vertices are elements of the state space and where there is an edge between x and y if and only if $P(x,y) > 0$. Let us start by giving a formal definition of a graph.

Definition 1.5.1. A ***graph*** is a pair $G = (V,E)$, where V is the ***vertex set*** and $E \subset V \times V$ is the ***edge set***. An edge in E is an unordered pair of vertices, i.e. $E \subset \{\{x,y\} : x,y \in V, x \neq y\}$. x and y are ***neighbours***, denoted by $x \sim y$, when $\{x,y\} \in E$. The number of neighbours of a vertex x is the ***degree*** of x, denoted by $\deg(x)$.

If it is possible to go from one vertex to another by walking along the edges of the graph, the vertices are connected. We often look at graphs, which are called connected graphs.

Definition 1.5.2. A graph G is ***connected*** if, for any two vertices x and y of G, there exists a sequence of vertices $x_0, x_1, \ldots, x_k$ such that $x_0 = x$ and $x_k = y$, and $x_i \sim x_{i+1}$ for $0 \leq i \leq k-1$.

We have introduced connected graphs to study a special type of Markov chain: the random walk.

Definition 1.5.3. A ***simple random walk*** on a graph G is a Markov chain with state space $\Omega = V$ and with transition matrix

$$P(x,y) = \begin{cases} \frac{1}{\deg(x)}, & \text{if } x \sim y \\ 0, & \text{otherwise} \end{cases} \tag{1.16}$$

This means that when the state is in x, the chain picks uniformly at random one of the neighbouring states to jump to. As $\deg(x)$ is the amount of neighbours, it is also the amount of options to jump to. The probability of jumping to a specific neighbouring state is $\frac{1}{\deg(x)}$.

If the probability of going to neighbour y differs from the probability of going to neighbour z, the chain does not pick a neighbour to jump to uniformly at random anymore. This is called a weighted random walk, where the weight between x and y is denoted by $w(x,y)$.

Definition 1.5.4. Consider a pair $(G,\{w\})$, where $G=(V,E)$ is a graph and $\{w\}=\{w(x,y):x,y\in V,\{x,y\}\in E\}$. Then a ***weighted random walk*** on G is a Markov chain with state space $\Omega=V$ and with transition matrix

$$P(x,y)=\begin{cases}\frac{w(x,y)}{w(x)}, & \text{if } x\sim y\\ 0, & \text{otherwise}\end{cases}$$

where $w(x)=\sum_{y\sim x}w(x,y)$ is the sum of all the weights of x and its neighbours.

Note that a weighted random walk has stationary distribution

$$\pi(x)=\frac{w(x)}{W},\ W:=\sum_{x\in V}w(x). \tag{1.17}$$

We call W the total edge-weight, although every weight is counted twice as both $w(x,y)$ and $w(y,x)$ are counted.

The following proposition relates connectivity, a property of graphs, to the irreducible random walk. The proof of the following proposition is an exercise in [1]. Below we give a detailed proof.

Proposition 1.5.5. A random walk on G is irreducible if and only if G is connected.

Proof. "⇐": assume $G=(V,E)$ is connected and take two vertices $x,y\in V$. By definition, there exists a sequence of vertices $x_0,x_1,\dots,x_k$ such that $x_0=x$ and $x_k=y$, and $x_i\sim x_{i+1}$ for $0\le i\le k-1$. Because this is a random walk $x_i\sim x_{i+1}$ implies $P(x_i,x_{i+1})>0$. Then the product of all $P(x_i,x_{i+1})$'s is also strictly positive and

$$P^k(x,y)\ge\Pi_{i=0}^{k-1}P(x_i,x_{i+1})>0.$$

Therefore, by taking t equal to k, we have $P^t(x,y)>0$.

"⇒": assume you have an irreducible random walk on G. Take $x,y\in V$ arbitrary, then there exists a t such that $P^t(x,y)>0$. $P^t(x,y)$ is the sum of $\Pi_{i=0}^{t-1}P(x_i,x_{i+1})$ for every sequence $x_0,x_1,\dots,x_{t-1},x_t$ with $x_0=x$ and $x_t=y$, i.e.

$$P^t(x,y)=\sum_{\substack{x_0,\dots,x_t\\ x_0=x,x_t=y}}\Pi_{i=0}^{k-1}P(x_i,x_{i+1})>0.$$

Therefore there exists at least one sequence $x_0=x,x_1,\dots,x_{k-1},x_k=y$ such that $\Pi_{i=0}^{k-1}P(x_i,x_{i+1})>0$, and thus $P(x_i,x_{i+1})>0$ for all $0\le i\le t-1$. $P(x_i,x_{i+1})>0$ in a random walk implies $x_i\sim x_{i+1}$. Now take $k=t$ and we have constructed a sequence which connects x to y. □

In other words, by Proposition 1.5.5, if all vertices of the graph are connected, all vertices, or states, are reachable by a random walk on the same graph, regardless of the initial distribution. And if every state is reachable by a random walk on the graph with arbitrary initial distribution, the graph is connected. In Section 1 we showed the special properties of irreducible Markov chains. Checking whether a graph is connected is a simple way to see if the chain is irreducible. This demonstrates one of the many ways in which we are able to study a random walk by studying its graph.

Like irreducibility, aperiodicity is also linked to a graph property. The following proposition is mentioned but not proven in [5].

Proposition 1.5.6. An irreducible random walk on a graph G is aperiodic if and only if G is not bipartite.

To prove this we first give the definition of a bipartite graph and a proposition. Both are from [8] including the proof of the proposition.

Definition 1.5.7. A graph $G = (V, E)$ is called ***bipartite*** if V admits a partition into 2 classes such that every edge has its ends in different classes: vertices in the same partition class must not be neighbours.

Proposition 1.5.8. A graph is bipartite if and only if it contains no odd cycle.

A cycle in a graph is a path $x_0, x_1, \ldots, x_n$, where $x_0 = x_n$. For a chain on the graph, the period of a vertex v (see Definition 1.1.2) is equal to the greatest common divisor of the lengths of cycles containing v. Now we give a detailed proof of Proposition 1.5.6.

Proof. "⇒": Assume an irreducible random walk on a graph G is aperiodic, then by Proposition 1.5.5, G is connected. Then every vertex has at least one neighbour and therefore nonzero probability of going to that neighbour and returning directly after, i.e. for every x we have $2 \in \mathcal{T}(x) = \{t \geq 1 : P^t(x,x) > 0\}$. Because the chain is aperiodic, $\mathcal{T}(x)$ must also contain an odd number, which means x must be part of at least one odd cycle. Then by Proposition 1.5.8, the graph is not bipartite.

"⇐": Assume we have an irreducible random walk on a non-bipartite graph G. Again by the same argument as before, by Proposition 1.5.5 G is connected and for every x we have $2 \in \mathcal{T}(x) = \{t \geq 1 : P^t(x,x) > 0\}$. Because G is non-bipartite, by Proposition 1.5.8 G contains at least one odd cycle. Take vertex y on such an odd cycle. Then we have that the odd cycle length is contained in $\mathcal{T}(y)$. As 2 is also contained in $\mathcal{T}(y)$, the period of y is equal to 1 which means the chain is aperiodic. □

We now introduce one final result on weighted random walks. In “Reversible Markov Chains and Random Walks on Graphs” [5] D. Aldous and J. Fill state that weighted random walks on connected graphs are reversible and that every irreducible reversible Markov chain “can be regarded as as random walk on the weighted graph with edge-weights $w(x,y) := \pi(x)P(x,y)$.”

A weighted random walk on a connected graph is indeed reversible. By Proposition 1.5.5 the chain is irreducible. Then the chain is reversible (see Definition 1.4.2), because the detailed balance equations hold for all states.

$$\pi(x)P(x,y) = \frac{w(x)}{W}\frac{w(x,y)}{w(x)} = \frac{w(y)}{W}\frac{w(y,x)}{w(y)} = \pi(y)P(y,x). \tag{1.18}$$

Now we show that every irreducible reversible chain can be regarded as a weighted random walk with edge-weights $w(x,y) := \pi(x)P(x,y)$. Because the chain is reversible, the symmetry of the weights is conserved, i.e. $w(x,y) = \pi(x)P(x,y) = \pi(y)P(y,x) = w(y,x)$. Therefore, we can construct a weighted random walk (WRW) on a graph $G = (V,E)$ with edge-weights $w(x,y)$ for neighbouring x and y and transition matrix

$$P_{\mathrm{WRW}}(x,y) = \frac{w(x,y)}{\pi(x)},\ x,y \in V. \tag{1.19}$$

Where for row x the stationary distribution of the reversible chain is the normalisation constant

$$\sum_{y\in V} P_{\mathrm{WRW}}(x,y) = \sum_{y\in V}\frac{w(x,y)}{\pi(x)} = \sum_{y\in V} P(x,y) = 1. \tag{1.20}$$

We have introduced random walks on graphs. In practice, this type of Markov chain often occurs, as state spaces frequently represent places, like buildings in a city for example. Let us take a look at a different example of a random walk: a pigeon on a train network.

Example 1.5.9 (Random train rides)**.** Assume a pigeon starts taking random trains from Schiphol airport between the six stations in Figure 1.2. We would like to know at what station the pigeon is most likely to be after a certain time. To answer this question two things are needed: the initial distribution μ_0 and the transition matrix P.

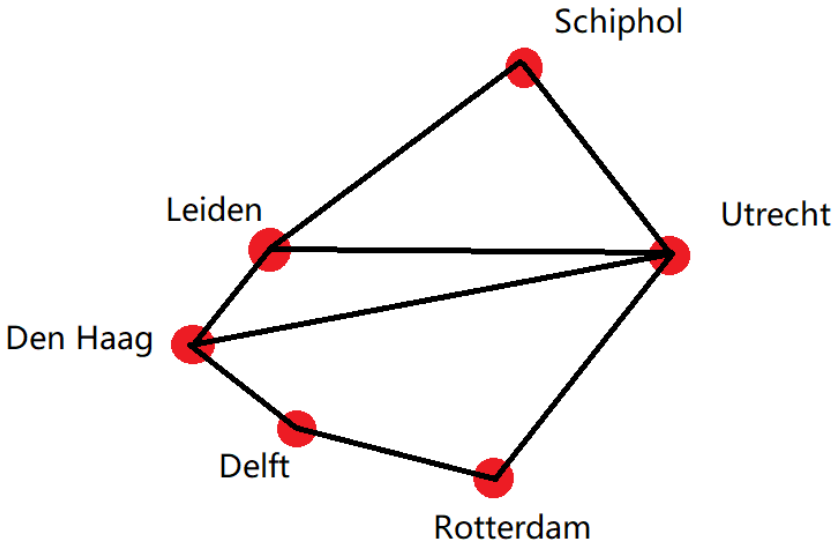

Figure 1.2: A train graph G with stations as vertices and train tracks as edges. Train connections from [9].

Since the pigeon always chooses the next train uniformly at random among the ones that leave from the station, the Markov chain is a simple random walk and the transition matrix is the same as in Equation 1.16. Ordering the cities in a clockwise manner starting at Schiphol (s), the pigeon has initial distribution $\mu_0 = (1,0,\ldots,0)$ and the transition matrix is:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \end{pmatrix}. \tag{1.21}$$

We now look if the chain is irreducible and what the period is for each station. From Figure 1.2 it is clear that G is a connected graph. Therefore by Proposition 1.5.5, the chain is irreducible.

Because the chain is irreducible, by Lemma 1.1.3 the period is the same for every station. We look at the period of Schiphol to find the period of the chain. It is possible to return to Schiphol in 2 steps by taking the train to Leiden and back, and in 3 steps by first taking the train to Leiden, then to Utrecht and then back to Schiphol. As gcd(2,3) = 1, by Definition 1.1.2 the period of Schiphol is 1. Therefore the chain is aperiodic.

To find the probability of finding the pigeon at station x at time t, we solve $\mu_t(x) = \mu_0 P^t(s,x) = \boldsymbol{P}_s\{X_t = x\}$. In Figure 1.3 the odds of finding the pigeon at a station are displayed for the first twenty steps.

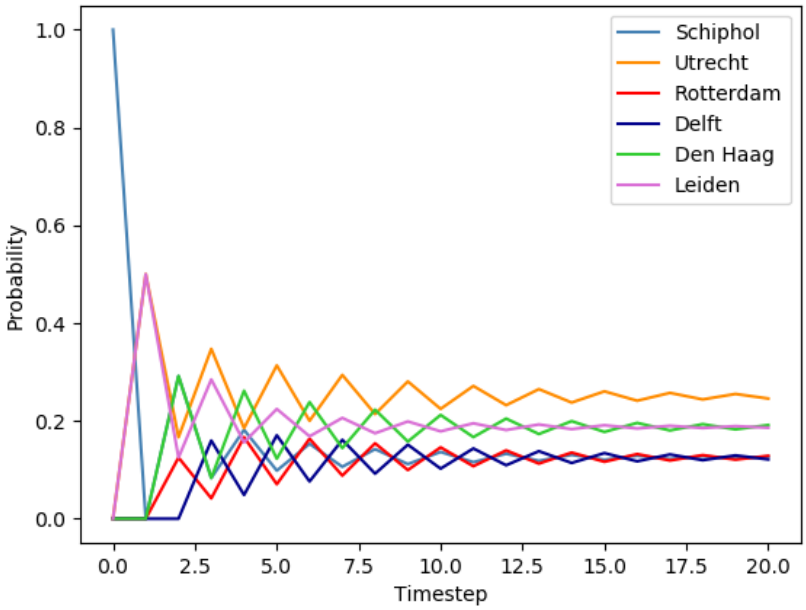

Figure 1.3: Where is the pigeon? The probability $\mathbf{P}_s\{X_t = x\}$ for every station x at time $0 \leq t \leq 20$.

It seems that the probability distribution of the pigeon converges to some distribution over time. As the chain is aperiodic and irreducible, this convergence is true by the Convergence Theorem 1.4.6. Therefore, the probability distribution of the pigeon position converges to a unique stationary distribution π, where

$$\pi(x) = \frac{\deg(x)}{2|E|}, \quad \text{and}$$
$$\pi = \left(\frac{2}{16}, \frac{4}{16}, \frac{2}{16}, \frac{2}{16}, \frac{3}{16}, \frac{3}{16}\right).$$

By Proposition 1.4.1, $\pi(x)$ is also equal to the inverse of the expected return time $\boldsymbol{E}_x(\tau_x^+)$. Then by looking at π, we see that the pigeon is expected to return in 8 steps if it starts at Schiphol, Rotterdam or Delft, expected to return in $\frac{16}{3} \approx 5.3$ steps if it starts at Leiden or Den Haag and expected to return in 4 steps if it starts in Utrecht.

Then by the convergence theorem we know that for large t, the probability of finding the pigeon in Utrecht are the highest, namely around 25%. We also found that this is the case for $t \geq 9$ (see Figure 1.3). We thus conclude it is best to look for the pigeon in Utrecht after some time has passed.

2

Electric Networks

In the previous section we introduced Markov chains and illustrated some of their properties. Here we will introduce electric networks and show that they are closely related to Markov chains, in particular random walks. The structure of this section follows partially Section 9 of "Markov Chains and Mixing Times" [1], while the most classic reference for the connection between electric networks and random walks is the work of P. Doyle and J. Snell (1984), "Random walks and electric networks" [4].

Let us start with the definition of an electric network.

Definition 2.0.1. An ***electric network*** is a pair $(G, \{c(e)\})$, where $G = (V, E)$ is a finite undirected connected graph, and where $\{c(e)\}$ are non-negative numbers that are associated to the edges of G. These numbers are called ***conductances*** and are symmetric, i.e. $c(x, y) = c(y, x)$, $\forall \{x, y\} \in E$. The ***total conductance*** of a vertex is the sum of the conductances with its neighbours: $c(x) = \sum_{x \sim y} c(x, y)$. The reciprocal of the conductance is the ***resistance***, $r(x, y) = \frac{1}{c(x,y)}$. The vertices in an electric network are also called ***nodes***.

Electric networks are also characterized by **voltages** $W(x)$ and **currents** $I(x, y)$, which are explained in detail in Section 2.1.

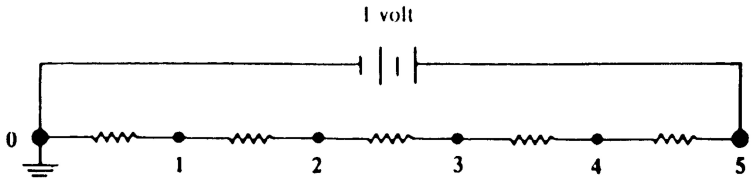

Figure 2.1: An example of an electric network. Taken from Random walks and electric networks by Doyle and Snell [4].

A first step in order to find a connection with Markov chains, is to consider a weighted random walk on a finite connected graph G (see Definition 1.5.4) with transition matrix

$$P(x, y) = \frac{c(x, y)}{c(x)}, \tag{2.1}$$

where $\{c(e)\}$ are the weights and the total conductances $c(x)$, $x \in V$, are normalising factors, i.e.

$$\sum_{y \in V} P(x, y) = \sum_{y \sim x} \frac{c(x, y)}{c(x)} = \frac{1}{c(x)} \sum_{y \sim x} c(x, y) = 1. \tag{2.2}$$

Since we assumed G is connected, the weighted random walk is irreducible by Proposition 1.5.5 and therefore by Proposition 1.4.1 it has the unique stationary distribution

$$\pi(x) = \frac{c(x)}{C}, \quad x \in V, \tag{2.3}$$

where C is the sum of the total conductances $C = \sum_{x \in V} c(x)$ and acts as the normalisation constant [10].

In [10] it is stated that this chain is reversible (see Definition 1.4.2). Indeed, the symmetry of the conductances $c(x,y) = c(y,x)$ implies

$$\pi(x)P(x,y) = \frac{c(x)}{C}\frac{c(x,y)}{c(x)} = \frac{c(y)}{C}\frac{c(y,x)}{c(y)} = \pi(y)P(y,x), \ \forall x, y \in V. \tag{2.4}$$

In Section 1.5 we showed that the study of reversible Markov chains is equal to the study of weighted random walks. We now have shown that to study reversible chains we can use the electric network approach.

2.1. Current flows

The current and voltage are linked to electric networks by a current flow, which will be introduced in this section.

Consider a network $(G, \{c(e)\})$, where we distinguish two nodes, a and z, which are called the **source** and the **sink** respectively. The **voltage** W is a harmonic function (see Definition 1.2.1) on $V \backslash \{a, z\}$. An **oriented edge** $\overrightarrow{e} = \overrightarrow{xy}$ is an ordered pair of nodes (x, y). A **flow** θ is an antisymmetric function on the oriented edges, i.e. $\theta(\overrightarrow{xy}) = -\theta(\overrightarrow{yx})$.

Definition 2.1.1. The ***divergence*** of θ at node x is defined as

$$\mathrm{div}\theta(x) := \sum_{y:\, y \sim x} \theta(\overrightarrow{xy}). \tag{2.5}$$

Because a flow is antisymmetric, for any flow we have the following

$$\sum_{x \in V} \mathrm{div}\theta(x) = \sum_{x \in V} \sum_{y \sim x} \theta(\overrightarrow{xy}) = \sum_{\{x,y\} \in E} [\theta(\overrightarrow{xy}) + \theta(\overrightarrow{yx})] = 0. \tag{2.6}$$

For every inner node not equal to the source a or sink z, the flow going in is equal to the flow going out. This gives a divergence of 0 on these nodes and this is called Kirchhoff's node law. The source is the only node which is allowed to have a positive divergence, i.e. the flow going out of the source is larger than the flow going into the source. In (2.6) we showed that the sum of the divergences is equal to 0, which implies that the divergence of the sink must be equal to minus the divergence of the source, $\mathrm{div}\theta(z) = -\mathrm{div}\theta(a)$. In other words, the flow going into the sink is larger than the flow going out of the sink. Below we give the formal definition.

Definition 2.1.2. A ***flow*** θ ***from*** a ***to*** z is a flow satisfying

1. ***Kirchhoff's node law***:
$$\mathrm{div}\theta(x) = 0, \quad \forall x \notin \{a, z\}, \tag{2.7}$$
2. $\mathrm{div}\theta(a) \geq 0$.

The ***strength*** of a flow θ from a to z is $\|\theta\| := \mathrm{div}\theta(a)$. If the strength is equal to 1, we call θ a ***unit flow***.

On an electric network with a voltage W, which is a harmonic function on $V\backslash\{a,z\}$, the current flow is constructed in the following way.

Definition 2.1.3. Given a voltage W on the network, the ***current flow*** I is defined by

$$I(\overrightarrow{xy}) = \frac{W(x)-W(y)}{r(\overrightarrow{xy})} = c(x,y)[W(x)-W(y)]. \tag{2.8}$$

A current flow is a well known concept in physics and it satisfies **Ohm's Law**

$$r(\overrightarrow{xy})I(\overrightarrow{xy}) = W(x)-W(y), \tag{2.9}$$

and the **cycle law**

$$\sum_{i=1}^{m} r(\overrightarrow{e_i})I(\overrightarrow{e_i}) = 0. \tag{2.10}$$

Now we show that any flow satisfying the cycle law and with the same strength as the current flow, must be equal to the current flow.

This proof is alternative to the one provided by D. Levin, Y. Peres and E. Wilmer in [1]. We follow instead Exercise 9.4 in [1] and divide the proof into 2 parts. In the first one we introduce some function h and show it is harmonic at all nodes and in the second one we use the specific choice of h to complete the proof.

Proposition 2.1.4 (Node law/cycle law/strength)**.** If θ is a flow from a to z satisfying the ***cycle law***

$$\sum_{i=1}^{m} r(\overrightarrow{e_i})\theta(\overrightarrow{e_i}) = 0 \tag{2.11}$$

for any cycle $\overrightarrow{e_1},\dots,\overrightarrow{e_m}$ and if $\|\theta\| = \|I\|$, then $\theta = I$.

Proof. Step 1: Define h as

$$h(x) = \sum_{i=1}^{m} [\theta(\overrightarrow{e_i}) - I(\overrightarrow{e_i})]\, r(\overrightarrow{e_i}), \tag{2.12}$$

where $\overrightarrow{e_1},\dots,\overrightarrow{e_m}$ is an arbitrary path from a to x.

We want to show that h is harmonic (recall Definition 1.2.1), i.e. $\forall x \in \Omega$, $h(x) = \sum_{y\in\Omega} \mathrm{P}(x,y)h(y)$. Take $x \in \Omega$ arbitrary, substitute (2.12) in $h(y)$ and set $\mathrm{P}(x,y) = \frac{c(x,y)}{c(x)}$ to get

$$\begin{aligned}\sum_{y\in\Omega} \mathrm{P}(x,y)h(y) &= \sum_{y\in\Omega} \frac{c(x,y)}{c(x)} \left(\sum_{i=1}^{n} [\theta(\overrightarrow{e_i}) - I(\overrightarrow{e_i})]\, r(\overrightarrow{e_i}) \right) \\ &= \sum_{y\sim x} \frac{c(x,y)}{c(x)} \left(\sum_{i=1}^{n} [\theta(\overrightarrow{e_i}) - I(\overrightarrow{e_i})]\, r(\overrightarrow{e_i}) \right),\end{aligned}$$

because $c(x,y) = 0$ for non-neighbouring y. Take for all neighbours y_i of x, $i = 0,\dots,\deg(x)$, a path from a to y_i as a path from a to x (of length m) plus one edge $\overrightarrow{e_{m+1}} = \overrightarrow{xy_i}$. Note that the path from a to x contains a single neighbour of x, say y_0. Then the path from a to y_0 of length $m+1$ contains the edge between x and y_0 twice, as it contains both $\overrightarrow{y_0x} = \overrightarrow{e_m}$ and $\overrightarrow{xy_0} = \overrightarrow{e_{m+1}}$

(see Figure 2.2). Therefore, we have

$$\begin{aligned}\sum_{y\in\Omega} P(x,y)h(y) &= \sum_{y\sim x}\frac{c(x,y)}{c(x)}\left(\sum_{i=1}^{m}[\theta(\overrightarrow{e_i})-I(\overrightarrow{e_i})]\,r(\overrightarrow{e_i})+[\theta(\overrightarrow{e_{m+1}})-I(\overrightarrow{e_{m+1}})]r(x,y)\right)\\ &= \sum_{i=1}^{m}[\theta(\overrightarrow{e_i})-I(\overrightarrow{e_i})]\,r(\overrightarrow{e_i})\sum_{y\sim x}\frac{c(x,y)}{c(x)}+\sum_{y\sim x}\frac{c(x,y)}{c(x)}r(x,y)[\theta(\overrightarrow{e_{m+1}})-I(\overrightarrow{e_{m+1}})]\\ &= \sum_{i=1}^{m}[\theta(\overrightarrow{e_i})-I(\overrightarrow{e_i})]\,r(\overrightarrow{e_i})+\frac{1}{c(x)}[\mathrm{div}\theta(x)-\mathrm{div}I(x)],\end{aligned}$$

$$\text{thus}\ \sum_{y\in\Omega} P(x,y)h(y) = h(x)+\frac{1}{c(x)}[\mathrm{div}\theta(x)-\mathrm{div}I(x)]. \tag{2.13}$$

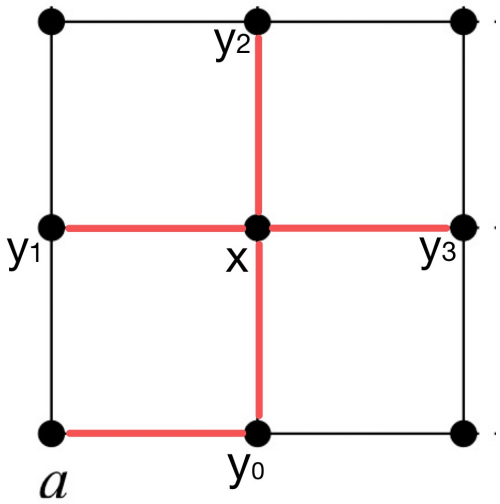

Figure 2.2: Path of length $m=2$ from a to x and neighbours $y_0, y_1, \ldots$. Since y_0 lies on the path from a to x, the constructed path of length $m+1$ from a to y_0 contains both $\overrightarrow{y_0x}$ and $\overrightarrow{xy_0}$.

We now show that $\mathrm{div}\theta(x)$ is equal to $\mathrm{div}I(x)$ for every $x\in\Omega$. Since θ is a flow from a to z, θ satisfies Kirchhof's node law at all nodes except a and z (1). θ also satisfies the cycle law (2) and $\|\theta\| = \|I\|$ (3). For $x=a$, $\mathrm{div}\theta(a) \overset{(1)}{=} \|\theta\| \overset{(3)}{=} \|I\| = \mathrm{div}I(a)$, so $\mathrm{div}\theta(x) = \mathrm{div}I(x)$ holds for $x=a$. For $x=z$, $\mathrm{div}\theta(z) \overset{(1)}{=} -\|\theta\| \overset{(3)}{=} -\|I\| = \mathrm{div}I(z)$. For all other nodes $\mathrm{div}\theta(x) \overset{(2)}{=} 0 = \mathrm{div}I(x)$. Therefore for every $x\in\Omega$, $[\mathrm{div}\theta(x)-\mathrm{div}I(x)]=0$ and

$$\sum_{y\in\Omega} P(x,y)h(y) = \sum_{i=1}^{m}[\theta(\overrightarrow{e_i})-I(\overrightarrow{e_i})]\,r(\overrightarrow{e_i}) = h(x).$$

Since x was taken arbitrarily, h is harmonic at every $x\in\Omega$, thus h is harmonic on Ω.

Step 2: By definition of electric networks, G is a connected graph. Then by Proposition 1.5.5 P is irreducible.

Since any path from a to a is an oriented cycle, by the cycle law we have $h(a)=0$. Therefore, by Lemma 1.2.2, $h=0$ at every node.

We now show that $\theta = I$, using the defined h and by induction on the path length m.

In the case $m = 1$, the path between x and a must be a single edge (with non-zero resistance), so x is a neighbour of a. Then $h(x) = [\theta(\overrightarrow{e_1}) - I(\overrightarrow{e_1})]\, r(\overrightarrow{e_1}) = 0$, which implies $\theta(\overrightarrow{e_1}) = I(\overrightarrow{e_1})$. Thus $\theta = I$ holds for $m = 1$ (and for all $\overrightarrow{e_1}$).

Suppose $\theta = I$ holds for any path starting at a of length $1 \le k < m$. Take x and a path of length m from a to x arbitrary, then

$$\begin{aligned} 0 = h(x) &= \sum_{i=1}^{m} [\theta(\overrightarrow{e_i}) - I(\overrightarrow{e_i})]\, r(\overrightarrow{e_i}) \\ &= \sum_{i=1}^{m-1} [\theta(\overrightarrow{e_i}) - I(\overrightarrow{e_i})]\, r(\overrightarrow{e_i}) + [\theta(\overrightarrow{e_m}) - I(\overrightarrow{e_m})]\, r(\overrightarrow{e_m}) \\ &= h(y) + [\theta(\overrightarrow{e_m}) - I(\overrightarrow{e_m})]\, r(\overrightarrow{e_m}) \\ &= 0 + [\theta(\overrightarrow{e_m}) - I(\overrightarrow{e_m})]\, r(\overrightarrow{e_m}). \end{aligned}$$

Where in the last step, we considered that y is a neighbour of x. This implies $\theta(\overrightarrow{e_m}) = I(\overrightarrow{e_m})$. We already have $\theta(\overrightarrow{e_k}) = I(\overrightarrow{e_k})$, $0 \le k < m$ as they form a path of length $m-1$ starting at a. Thus $\theta = I$ on this path of length m.

So by induction $\theta = I$ on any path of length $m \ge 1$. To show that $\theta = I$ on any edge, take an arbitrary edge. Then there exists a path of length ≥ 1 from a to a containing this edge, thus by the induction result $\theta = I$ on this edge.

□

2.2. Effective resistance and results

Now we introduce the concept of effective resistance and we will illustrate some of its implications.

Definition 2.2.1. The ***effective resistance*** between vertices a and z is

$$\mathscr{R}(a \longleftrightarrow z) = \frac{W(a) - W(z)}{||I||}, \tag{2.14}$$

where $\|I\|$ is the strength of the current flow.

The effective resistance would be the resistance between a and z, if the entire network would be replaced by a single edge between a and z.

Definition 2.2.2. The ***effective conductance*** between vertices a and z is the reciprocal of the effective resistance

$$\mathscr{C}(a \longleftrightarrow z) = \frac{1}{\mathscr{R}(a \longleftrightarrow z)} \tag{2.15}$$

The effective resistance brings along a useful equality, which is the first powerful connection between electric networks and Markov chains and it is given in the following proposition.

Proposition 2.2.3. For any x, a, $z \in \Omega$,

$$\boldsymbol{P}_a\{\tau_z < \tau_a^+\} = \frac{1}{c(a)\mathscr{R}(a \longleftrightarrow z)} = \frac{\mathscr{C}(a \longleftrightarrow z)}{c(a)}. \tag{2.16}$$

Relating the escape probability (1.3.2) to the electric network is a big advantage and is one of the main advantages of the electric network approach. It is often easier to calculate the effective resistance or the effective conductance, rather than calculating escape probabilities using

only probability theory. One reason for this is that the effective resistance can be computed or estimated using variational principles such as Thomson's Principle. Thomson's Principle relates the effective resistance (and thus indirectly the escape probability) to the energy in the network.

Definition 2.2.4. The ***energy*** of a flow θ is

$$\mathscr{E}(\theta) := \sum_e [\theta(e)]^2 r(e). \tag{2.17}$$

Theorem 2.2.5 (Thomson's Principle)**.** For any finite connected graph,

$$\mathscr{R}(a \longleftrightarrow z) = \inf\{\mathscr{E}(\theta) : \theta \text{ a unit flow from a to z}\}. \tag{2.18}$$

The unique minimizer in the inf above is the unit current flow.

By using the energy of a flow, Thomson's Principle relates the resistances of single edges to the effective resistance. It is then natural to assume that smaller resistances lead to a smaller effective resistance. This assumption is in fact correct according to Rayleigh's Monotonicity Law.

Theorem 2.2.6 (Rayleigh's Monotonicity Law)**.** If $\{r(e)\}$ and $\{r'(e)\}$ are sets of resistances on the edges of the same graph G and if $r(e) \le r'(e)$ for all e, then

$$\mathscr{R}(a \longleftrightarrow z; r) \le \mathscr{R}(a \longleftrightarrow z; r'). \tag{2.19}$$

2.3. Network reduction laws

Relating electric networks to probability theory means that we can also use physical laws of electric networks, like the following ones.

For the first two laws we provide a proof, which was lacking in [1].

Theorem 2.3.1 (Parallel Law)**.** Conductances in parallel add: suppose edges e_1 and e_2, with conductances c_1 and c_2, respectively, share vertices v_1 and v_2 as endpoints. Then both edges can be replaced with a single edge e of conductance $c_1 + c_2$ without affecting the rest of the network. All voltages and currents in $G\backslash\{e_1, e_2\}$ are unchanged and the current is given by $I(\overrightarrow{e}) = I(\overrightarrow{e_1}) + I(\overrightarrow{e_2})$.

Proof. We check whether Kirchhoff's node law and Ohm's law still hold when we replace e_1, e_2 with e.

Because $I(\overrightarrow{e}) = I(\overrightarrow{e_1}) + I(\overrightarrow{e_2})$, $\mathrm{div} I(v_1)$ and $\mathrm{div} I(v_2)$ are unchanged. Therefore Kirchhoffs node law (2.7) still holds.

Rewriting Ohm's law (2.9) gives $I(\overrightarrow{e_1}) = c(e_1)[W(v_1) - W(v_2)]$ and $I(\overrightarrow{e_2}) = c(e_2)[W(v_1) - W(v_2)]$. Adding them up we have

$$I(\overrightarrow{e}) = I(\overrightarrow{e_1}) + I(\overrightarrow{e_2}) = (c(e_1) + c(e_2))[W(v_1) - W(v_2)] = c(e)[W(v_1) - W(v_2)]. \tag{2.20}$$

Therefore

$$\frac{I(\overrightarrow{e})}{c(e)} = [W(v_1) - W(v_2)]. \tag{2.21}$$

□

Theorem 2.3.2 (Series Law)**.** Resistances in series add: if $v \in V\backslash\{a, z\}$ is a node of degree 2 with neighbours v_1 and v_2, the edges (v_1, v) and (v, v_2) can be replaced by a single edge (v_1, v_2) of resistance $r_{v_1 v} + r_{v v_2}$. All potentials and currents in $G\backslash\{v\}$ remain the same and the current that flows from v_1 to v_2 equals $I(\overrightarrow{v_1 v}) = I(\overrightarrow{v v_2})$.

Proof. We again check whether Kirchhoff's node law and Ohm's law still hold when we replace v_1v, vv_2 with v_1v_2.

Because $I(\overrightarrow{v_1v_2}) = I(\overrightarrow{v_1v})$ and $I(\overrightarrow{v_2v_1}) = I(\overrightarrow{vv_2})$, div$I(v_1)$ and div$I(v_2)$ are unchanged. Therefore Kirchhoffs node law (2.7) still holds.

Adding Ohm's law (2.9) for both edges

$$r_{v_1v}I(\overrightarrow{v_1v}) + r_{vv_2}I(\overrightarrow{vv_2}) = W(v_1) - W(v) + W(v) - W(v_2), \tag{2.22}$$

and we have

$$r_{v_1v_2}I(\overrightarrow{v_1v_2}) = [W(v_1) - W(v_2)]. \tag{2.23}$$

□

Theorem 2.3.3 (Gluing)**.** Vertices with the same voltage can be combined, while keeping all existing edges.

Proof. From physics we know that the current never flows between vertices with the same voltage, therefore the potentials and the currents are unchanged. □

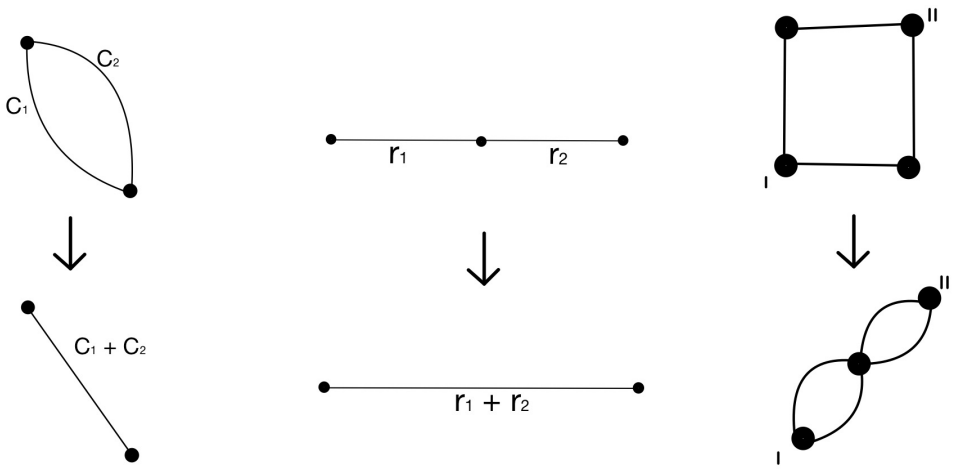

Figure 2.3: From left to right: parallel law, series law and gluing.

Let us use the network reduction laws to simplify a Markov chain.

The proof of the following proposition is left as an exercise in [1]. Below we give the detailed proof.

Proposition 2.3.4 (Gambler's ruin)**.** Assume that you are making fair unit bets on coin flips and will abandon the game when your fortune falls to 0 or rises to n. Start with $X_0 = x$ amount of fortune, where $0 \le x \le n$. Then

$$\boldsymbol{P}_x\{\tau_n < \tau_0\} = \frac{x}{n}.$$

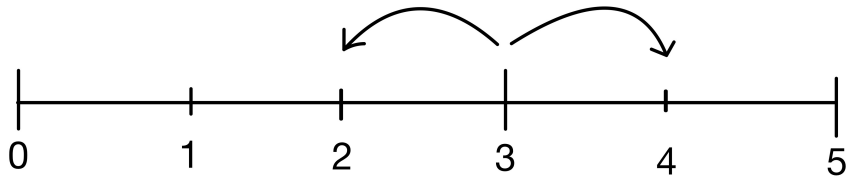

Figure 2.4: Simple random walk where $n = 5$.

Proof. If $x = 0$, then $\mathbf{P}_x\{\tau_n < \tau_0\} = 0 = \frac{x}{n}$.

If $x = n$, then $\mathbf{P}_x\{\tau_n < \tau_0\} = 1 = \frac{x}{n}$.

We now focus on the remaining case, that $x \in \{1, 2, \ldots, n-1\}$. Then the paths between x and 0 and between x and n must contain at least one edge.

Note that the simple random walk has unit conductance:

$$\mathrm{P}(x, y) = \frac{c(x,y)}{c(x)} = \begin{cases} \frac{1}{\deg(x)} & \text{, if } x \sim y. \\ 0 & \text{, elsewhere.} \end{cases}$$

Because every node between x and 0 and between x and n has degree 2, we apply the series law 2.3.2 on every node between 0 and n except for x until we are left with a network consisting of two edges and three nodes (see figure 2.5). The resistance of e_{0x} is $\sum_{i=0}^{x-1} r_{i,i+1} = x$ and the resistance of e_{xn} is $\sum_{i=x}^{n-1} r_{i,i+1} = n - x$.

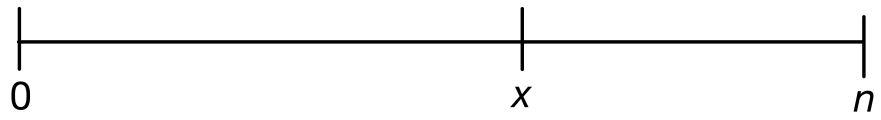

Figure 2.5: The reduced network after using the series law on all other nodes.

Because the reduced network only contains three nodes, $\mathbf{P}_x\{\tau_n < \tau_0\}$ is equal to the probability of going from x to n, $\mathrm{P}(x, n)$.

We then get the following

$$\begin{aligned} c(x,n) &= \frac{1}{r(x,n)} = \frac{1}{n-x} \\ c(0,x) &= \frac{1}{r(0,x)} = \frac{1}{x} \\ c(x) &= c(0,x) + c(x,n) = \frac{n}{x(n-x)} \\ \mathbf{P}_x\{\tau_n < \tau_0\} &= \mathrm{P}(x,n) = \frac{c(x,n)}{c(x)} = \frac{x(n-x)}{n(n-x)} = \frac{x}{n}. \end{aligned}$$

□

The proof of Proposition 2.3.4, if we were not to use the electric network approach, consists of solving a system of equations and calculating $\mathbf{P}\{\tau_n < \tau_0\}$ for every x simultaneously. By using the electric network approach, we do not have to take every node into account and simply reduce the network until we are left with easy calculations.

2.4. The Nash-Williams inequality

Another interesting result from electric networks, which we will soon discuss in greater detail, is the Nash-Williams inequality. The Nash-Williams inequality uses the topological properties of the electric network by dividing edges in edge-cutsets.

Definition 2.4.1. An ***edge-cutset*** of a graph is a set of edges such that every path between nodes a and z contains at least one edge of the edge-cutset, i.e. after removing the edges of the edge-cutset from the graph, a and z are disconnected.

Proposition 2.4.2 (Nash-Williams inequality)**.** Let $\{\Pi_k\}$ be the disjoint edge-cutsets which separate nodes a and z, then

$$\mathscr{R}(a \longleftrightarrow z) \geq \sum_k \left(\sum_{e \in \Pi_k} c(e) \right)^{-1}. \tag{2.24}$$

We will use Proposition 2.4.2 to obtain bounds for the escape probability from a box of a simple two-dimensional random walk.

Proposition 2.4.3. Let B_n be the subset of $\mathbb{Z}^2$ contained in the box of side length $2n$ centered at 0. Let ∂B_n be the set of vertices along the perimeter of the box, then

$$\lim_{n\to\infty} \boldsymbol{P}_0\{\tau_{\partial B_n} < \tau_0^+\} = 0.$$

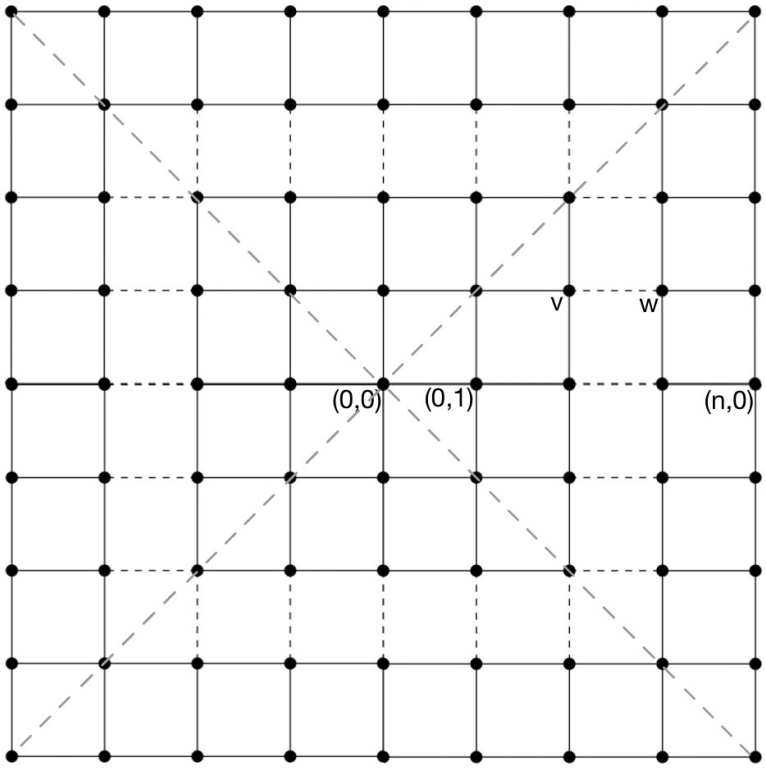

Figure 2.6: B_4, with cutset Π_3 in the black dashed lines. The diagonal grey dashed lines divide the cutsets in 4 sections, all containing $2k-1$ edges. v = (2,1), $\|v\|_\infty = 2$, w = (3,1) and $\|w\|_\infty = 3$. (v,w) is an edge from Π_3. Original from [1]

Proof. Let Π_k be the edge set $\Pi_k = \{(v,w) \in B_n : \|v\|_\infty = k-1, \|w\|_\infty = k\}$, with $\|(v_1,v_2)\|_\infty = \max\{|v_1|,|v_2|\}$. (The edge sets form "squares" around the center, see Figure 2.6.) Because every path from 0 to ∂B_n must contain an edge in Π_k for every $k \in \{1,2,\cdots,n\}$, Π_k is a cutset. Note that these edge-cutsets are disjoint.

Since every edge has unit conductance, $\left(\sum_{e\in\Pi_k} c(e)\right)^{-1} = \frac{1}{\#\text{ edges}}$. The edge set k can be divided into 4 equal parts, all containing $2k-1$ edges. Therefore $\left(\sum_{e\in\Pi_k} c(e)\right)^{-1} = \frac{1}{4(2k-1)}$ (see Figure 2.6). Using the Nash-Williams inequality, with $a = 0$ and $z = \partial B_n$, we get

$$\begin{aligned} 0 \le \mathbf{P}_0\{\tau_{\partial B_n} < \tau_0^+\} &= \frac{1}{c(0)\mathscr{R}(0 \longleftrightarrow \partial B_n)} \\ &\le \frac{1}{c(0)\sum_{k=1}^{n} \frac{1}{4(2k-1)}} \le \frac{1}{c(0)\sum_{k=1}^{n} \frac{1}{8k}} \le \frac{8}{c(0)\log(n)}, \end{aligned}$$

and therefore

$$\lim_{n\to\infty} \frac{8}{c(0)\log(n)} = 0. \tag{2.25}$$

Then, by the squeeze theorem $\lim_{n\to\infty} \mathbf{P}_0\{\tau_{\partial B_n} < \tau_0^+\} = 0$. □

We showed that the limit of the escape probability is 0, which means that on an infinitely large grid the random walk will return to the center at some finite time. Then the random walk in two dimensions is called recurrent in [4]. The random walk is also recurrent in one dimension, but not in three or more dimensions. For a random walk in three or more dimensions the escape probability is not equal to 0. We call these random walks transient.

Testing the Nash-Williams inequality

How accurate is the Nash-Williams inequality when estimating the escape probability described in Proposition 2.4.3? To answer this we simulate the random walk for different n and compare the results with the upper bound derived from the Nash-Williams inequality.

Taking into account unit conductance we have the following upper bound

$$\mathbf{P}_0\{\tau_{\partial B_n} < \tau_0^+\} \le \frac{1}{4\sum_{k=1} \frac{1}{\#\text{edges}}}. \tag{2.26}$$

The Nash-Williams inequality does not impose many restrictions on the choice of edge-cutsets. For example, it is possible to use only one edge-cutset, which gives the upper bound

$$\mathbf{P}_0\{\tau_{\partial B_n} < \tau_0^+\} \le \frac{\#\text{edges}}{4}.$$

Since the smallest possible cutset contains the 4 edges around the center, the upper bound is ≥ 1. This upper bound is not useful, because the escape probability must always be between 0 and 1. In this case, more edge-cutsets lead to tighter upper bounds (see (2.26)).

It is also possible to create larger edge-cutsets. The minimal size is $4(2k-1)$ edges for set $k = 1,\ldots,n$, as in Figure 2.6. The maximal size is $4(4k-1)$ edges for set k, i.e. when it contains all edges in B_n.

We simulated a simple random walk starting in the center of a two-dimensional grid of size $2n \times 2n$ (see Figure 2.6). The simulation stops if the chain returns to the center ("returned") or if the chain reaches the perimeter of the grid ∂B_n ("escaped").

For each n we simulate this random walk 10.000 times and keep track of how often the chain returns and how often the chain escapes. We also track the stopping times of the chains to calculate the average amount of steps of the chain, i.e. the average amount of steps needed to either return or reach the perimeter of the grid.

Using the law of large numbers from Section 13 of [11], we have for m simulations

$$\lim_{m\to\infty} \frac{\#\text{ escaped}}{m} = \mathbf{P}\{\tau_{\partial B_n} < \tau_0^+\}. \tag{2.27}$$

Therefore we approximate the escape probability by simulating the same chain 10.000 times, i.e.

$$\frac{\#\text{ escaped}}{10000} \approx \mathbf{P}\{\tau_{\partial B_n} < \tau_0^+\}. \tag{2.28}$$

We compare the approximated escape probability with two upper bounds, provided by the Nash-Williams inequality. The first upper bound results from taking n minimal size edge-cutsets and the second upper bound results from taking n maximal size edge-cutsets (using all edges of the network). For the first upper bound we calculate the rate

$$\text{rate } = \frac{\text{upper bound}}{\text{escape probability}}. \tag{2.29}$$

The higher the rate, the closer the upper bound is to the escape probability, with a perfect upper bound resulting in a rate of 1.

The results are in Table 2.1. Details of the simulation are in Appendix A.

Table 2.1: For different n, 10.000 simulations of random walks compared with two upper bounds, which are derived from the Nash-Williams inequality.

n	average # steps	returned	escaped	approx. esc. prob.	upper bound $4(2k-1)$	rate	upper bound $4(4k-1)$
2	3	3304	6696	0.6696	0.7500	0.8928	2.1000
5	14	5229	4771	0.4771	0.5595	0.7991	1.4569
10	46	6108	3892	0.3892	0.4688	0.8302	1.1718
20	160	6604	3396	0.3396	0.4033	0.8421	0.9770
50	835	7143	2857	0.2857	0.3404	0.8393	0.7995
100	2916	7504	2496	0.2496	0.3045	0.8197	0.7026

The smaller set size results in a tighter upper bound. The larger upper bound performs significantly worse, as it is at least double the smaller upper bound and does not directly drop below 1. This confirms that the choice of edge-cutsets greatly impacts the resulting upper bound.

The rate between the escape probability and the tighter upper bound is somewhere between 0.90 and 0.79 and between 0.85 and 0.80 for larger n. It is unclear whether the rate is dependent on n. For the n we studied, the rate is bounded above by 0.9. Therefore multiplying the smallest upper bound by 0.9 would result in a more accurate approximation of the escape probability, while still remaining an upper bound. This can only be done for general n if the rate is dependent on n or if the rate has a maximum smaller than 1.

Conclusion and discussion

The research question was: When do we use the electric network approach and in what way does this approach contribute to the understanding of Markov chains?

If the Markov chain of interest is a random walk, it is possible to use the electric networks approach. We have also shown that every reversible Markov chain is a random walk. This approach is built on translating the random walk to an electric network by relating the conductances to the weights. The electric network approach provides problem simplification tools, such as the Series Law, and powerful inequalities, such as the Nash-Williams inequality. These tools are based in physics and are often more intuitive than probability theory.

We have shown how the Nash-Williams inequality provides an upper bound for the escape probability on a two-dimensional grid. In this case the chain starts in the center and has escaped if it reaches the outer edges of a $2n$ by $2n$ grid. The upper bound, given by the Nash-Williams inequality, depends on the choice of edge-cutsets, where both more and smaller sets resulted in a better upper bound. The relation between the upper bound and the escape probability seems to be independent of the size of the grid.

Relating the Nash-Williams inequality to the escape probability, where the latter often requires extensive calculations, is a prime example of the advantages of the electric network approach.

It is, however, suboptimal to show the advantages of the electric network approach, when it is not compared to other approaches. For further research, we recommend comparing different approaches by applying them to the same problems. This would give the reader more insight in the advantages and disadvantages of the electric network approach, as well as give insight in the classical theoretical probability approach.

It would have been nice to add more simulations to the report. This would compensate for the amount of definitions and put the focus more on the different applications of Markov chains. Including more simulations would also help to show the reader the bigger picture. For example, elaborating why Markov chains are used and in what fields.

Although everything is written from a mathematics standpoint, we wished to incorporate more graph theory and linear algebra. We did, however, include some graph theory to support the material from [1]. With this we introduced how the study of graphs is closely related to the study of random walks.

Bibliografie

[1] Levin, D. A., Peres, Y., & Wilmer, E. L. (2008). *Markov Chains and Mixing Times*. American Mathematical Society.

[2] Markov chain. (n.d.). *Encyclopedia of Mathematics*. `http://encyclopediaofmath.org/index.php?title=Markov_chain&oldid=49254`

[3] Markov, A. (1907). Issledovanie zamechatel'nogo sluchaya zavisimyh ispytanij, *Petersburg Bull. (6) 1, No. 3*, 61-80

[4] Doyle, P. G., & Snell, L. J. (2006). *Random walks and electric networks*. The Mathematical Association of America.

[5] Aldous, D., & Fill, J. A. (2002). *Reversible Markov Chains and Random Walks on Graphs*.

[6] Häggström, O. (2002). *Finite Markov Chains and Algorithmic Applications*. London Mathematical Society.

[7] Boyle, M. (n.d.). *Notes on the Perron-Frobenius theory of nonnegative matrices*. Mathematics department, University of Maryland.

[8] Diestel, R. (2017). *Graph Theory*. Springer.

[9] NS travelplanner site. Last seen on 16$^{\text{th}}$ of March. `https://www.ns.nl/reisplanner/#/`

[10] Avena, L., Heydenreich, M., den Hollander, F., Verbitskiy, E., & van Zuijlen, W. (2016). *Random Walks (lecture notes)*. Leiden University.

[11] Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics*. Springer-Verlag London.

Appendix A

```
#Escape from the center on a 2D grid of 2n by 2n nodes.
import random

def process(upperbound, n):
    x = 0
    y = 0 #start in the middle
    time = 0
    while (abs(x) + abs(y) != 0 or time == 0) and abs(x)<n and abs(y)<n and time<upperbound:
        direction = random.randint(1,4)
        if direction == 1: #up
            y += 1
        elif direction == 2: #right
            x += 1
        elif direction == 3: #down
            y -= 1
        else: #left
            x -= 1
        time += 1
    if abs(x) + abs(y) == 0:
        stoppingreason = 0 #"returned"
    elif abs(x) == n or abs(y) == n:
        stoppingreason = 1 #"escaped"
    else:
        stoppingreason = 2 #"time out"
    return x, y, time, stoppingreason

def processes(i, upperbound, n):
    listofstoppingreasons = [0, 0, 0]
    listofstoppingtimes   = []
    for processes in range(i):
        stop_x, stop_y, stop_time, stoppingreason = process(upperbound, n)
        listofstoppingreasons[stoppingreason] += 1
        listofstoppingtimes.append(stop_time)
    return listofstoppingreasons, listofstoppingtimes

def bound(n):
    edgelist1 = 0.0
    edgelist2 = 0.0
    for k in range(1, n+1):
        edgelist1 += 1/(4*(2*k - 1)) #minimal sized edge-cutsets
        edgelist2 += 1/(4*(4*k-1))  #maximal sized edge-cutsets
    return 1/(4*edgelist1), 1/(4*edgelist2) #c(0) = 4 for unit conductance

#starting values
upperbound = 100000 #upperbound for timesteps
```

```
n = 100
tries = 10000

#simulate the processes
listofstoppingreasons, listofstoppingtimes = processes(tries, upperbound, n)

#print the results
print("Grid of length ", n)
print("Averagestoppingtime =", sum(listofstoppingtimes)/len(listofstoppingtimes))
print()
print("Times returned :", listofstoppingreasons[0])
print("Times escaped  :", listofstoppingreasons[1])
print("Times timed out:", listofstoppingreasons[2])
print()
bound1, bound2 = bound(n)
print("Minimal sized sets upper bound for escape probability =", bound1)
print("Maximal sized sets upper bound for escape probability =", bound2)
```