RANDOM WALKS AND AMENABILITY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ATABAY MAHMUDOV

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MATHEMATICS

JULY 2024

Approval of the thesis:

RANDOM WALKS AND AMENABILITY

submitted by **ATABAY MAHMUDOV** in partial fulfillment of the requirements for the degree of **Master of Science in Mathematics Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Hasan Taşeli
Head of Department, **Mathematics** _____

Assoc. Prof. Dr. Mustafa Gökhan Benli
Supervisor, **Mathematics Department, METU** _____

Examining Committee Members:

Prof. Dr. Yıldıray Ozan
Mathematics Department, METU _____

Assoc. Prof. Dr. Mustafa Gökhan Benli
Mathematics Department, METU _____

Assist. Prof. Dr. Elif Medetoğulları
Dept. of Mathematics and Science Education, TEDU _____

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Atabay Mahmudov

Signature :

ABSTRACT

RANDOM WALKS AND AMENABILITY

Mahmudov, Atabay
M.S., Department of Mathematics
Supervisor: Assoc. Prof. Dr. Mustafa Gökhan Benli

July 2024, 94 pages

Random walks on groups and graphs has a rich history and provides a connection between group theory and probability theory. The aim of this thesis is to investigate random walks on groups and graphs and their relation to amenability. Among the main topics which will be covered in the thesis will be H. Kesten‘s theorem characterizing amenable groups in terms of random walks. Also, a generalization of this theorem to invariant random subgroups proved recently by Abert, Galsner and Virag will be studied.

Keywords: Random Walks, Finitely Generated Groups, Infinite Graphs and Networks. Amenability, Invariant Random Subgroups.

ÖZ

RASTGELE ADIMLAR VE AMENABİLİTE

Mahmudov, Atabay
Yüksek Lisans, Matematik Bölümü
Tez Yöneticisi: Doç. Dr. Mustafa Gökhan Benli

Temmuz 2024 , 94 sayfa

Çizgeler ve gruplar üzerinde rastgele adımlar zengin bir tarihçeye sahiptir ve gruplar teorisi ile olasılık teorisi arasında bir bağlantı oluşturur. Bu tezin amacı, gruplar ve çizgeler üzerindeki rastgele adımları inceleyip, bunların amenabilite ile ilişkisini araştırmaktır. Başlıca incelenecek noktalar arasında, H. Kesten'in amenable grupları rastgele adımlar ile karakterize eden teoremi vardır. Ayrıca, yakın zamanda Abert, Glasner ve Virag tarafından ispatlanan ve Kesten'in teoremini rastgele değişmez alt gruplara genelleyen sonuç da incelenecektir.

Anahtar Kelimeler: Rastgele Adımlar, Sonlu Üreteçli Gruplar, Sonsuz Çizgeler ve Ağlar, Amenabilite, Değişmez Rastgele Altgruplar

To my parents...

ACKNOWLEDGEMENTS

First of all, my deepest gratitude goes to my supervisor Mustafa Gökhan Benli. His unwavering support and attention were crucial to my studies at METU and the prospects of my academic career. Reading papers together and discussing mathematics was the most valuable throughout my Master's degree as I learned a lot from his mathematical understanding, and approach to mathematics. I am indebted to him for being more mathematically mature.

I would like to thank to Department of Mathematics for the sincere environment.

I also want to thank Prof. Dr. Yıldıray Ozan and Assoc. Prof. Dr. Elif Medetoğulları for sparing their valuable time to be the committee members.

My deep gratitude also goes to the friends from the department I met along the journey. Special thanks to my friends Franciszek Ciuk, Egemen Gover, Ben Forsyth, Yiğithan Tamer, Tan Ozalp, and El Gato.

Finally, thanks to Öykü Naz Uzunoğlu for all the love and support.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Banach and Tarski [3] proved in 1924 that the unit ball in $\mathbb{R}^3$ can be decomposed into finitely many disjoint pieces and reassembled into two unit balls using only rigid motions. Such a decomposition is called a *paradoxical decomposition*. In 1929, John Von Neumann [10] discovered that this result is due to the underlying group. He introduced the class of amenable groups during his study of the Banach-Tarski phenomenon, which turned out to be very important for many areas of mathematics. His insight was that paradoxical decompositions arise exactly when the underlying group is not amenable. The classic examples of non-amenable groups are non-abelian free groups, which lie at the heart of the Banach-Tarski paradox.

In the last century, the quest for an algebraic characterization of amenable groups was a very influential problem in group theory. In fact, Von Neumann conjectured that every non-amenable group contains an isomorphic copy of the free group F_2. However, this conjecture was disproven by O'lshanskii when he discovered that there exist non-amenable groups with all proper subgroups of order exactly p for $p > 10^{75}$. It seems that an algebraic characterization of amenable groups remains an elusive open problem.

On the other hand, there are groups whose amenability is still unknown, such as Thompson's group F. Mathematicians have been developing tools to determine whether or not a group is amenable. In fact, amenability is an interesting notion that has various (and seemingly unrelated) characterizations. One of them is a probabilistic approach developed by Harry Kesten [8] in 1959. He studied random walks on the Cayley and Schreier graphs of groups and obtained an amenability criterion. This seminal result was generalized to Invariant Random Subgroups in 2012 by Abert,

Glasner and Virag.

This thesis is devoted to the investigation of amenable groups and Kesten's characterization (and its generalization) of amenable groups via random walks.

In Chapter 2, we will start with graph-theoretic preliminaries and then introduce Markov Chains (Random Walks) in a general setting, presenting important notions such as recurrence, transience, and certain results about them. Then, we provide an important result about the simple random walk on trees which enables to understand the nature of the simple random walk on the Cayley graph of a non-amenable group. We then consider reversible Markov chains and obtain some recurrence criteria which will be applied to the infamous problem called Polya's walk. Finally, we will introduce the notion of growth rate for groups and describe its relationship with random walks.

In the next chapter, we present Amenability, a very important group-theoretic notion. We explain in detail how it relates to the Banach-Tarski paradox and paradoxical decompositions. For practical purposes, we include another formulation of amenability and provide some basic properties of amenable groups. Moreover, we show that every group of sub-exponential growth is amenable, while free groups are non-amenable. Finally, we end the chapter with an important characterization of amenability which will be used in the proof of Kesten's theorem.

In Chapter 4, we will explain Kesten's theorem and its proof in detail. We will start by providing some operator-theoretic notions. The aim will be to use the fact that the spectral radius is equal to the norm of the Markov operator. Then, we will explain how one can describe translations of sets using operator theory. After the proof of Kesten's theorem, we will present a very interesting corollary linking the spectral radius of the Cayley graph $Cay(G, S)$ and the Schreier graph $Sch(G/H, S)$ where H is an amenable subgroup of G.

In the final chapter, we will generalize Kesten's Theorem to Invariant Random Subgroups. We will start by briefly discussing invariant random subgroups and then providing all the necessary lemmas to present the proof of Kesten's theorem for invariant random subgroups.

CHAPTER 2

RANDOM WALKS ON GROUPS AND GRAPHS

In this chapter, after providing necessary preliminaries regarding general graphs and random walks, we will focus on random walks on groups and graphs. We will investigate transient and recurrent random walks. After delving into reversible Markov chains and networks, we will prove a criterion for transience. Finally, we will touch upon the connection between growth and random walks. The main references for this chapter are the excellent books [9], [7], [11].

2.1 Preliminaries on Graphs

Definition 2.1. A graph Γ is a pair (V, E) where V is a set whose elements are called vertices, and E is a set whose elements are called edges which are unordered pair of vertices.

Let's state some elementary definitions on Graphs.

- If $\{v, w\} \in E$ for any v and w, then the vertices v and w ar e called adjacent (or neighbors), and we write $v \sim w$.
- An edge $\{v, w\}$ is said to be incident with v and w.
- Two edges are called adjacent if they are incident with the same vertex.
- Cardinality of the set V is called the order of Γ.
- Let $N(v)$ denote the set of all neighbors of $v \in V$. Degree of a vertex is defined to be the number $|N(v)|$ and is denoted by $deg(v)$.

- If $deg(v) = 0$, then v is called an isolated vertex.
- A vertex is of an even(respectively odd) degree if $deg(v)$ is an even(respectively odd) number.
- A graph is called k-regular if $deg(v) = k$ for all $v \in V$.
- A **path** of length n from v to w is a finite sequence $v = v_0, v_1, ..., v_n = w$ of distinct vertices of Γ such that $v_i \sim v_{i+1}$ for every $i = 0, 1, ..., n$. A graph is called connected if for every $v, w \in V(\Gamma)$ there exists a path from v to w.
- The minimal length of a path between v and w is defined to be the distance between v and w and is denoted by $d(v, w)$
- A graph is called locally finite if $deg(v) < \infty$ for all $v \in \Gamma$

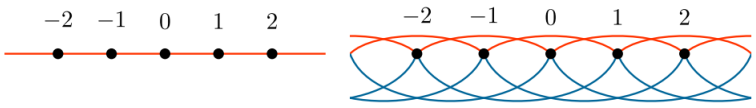
2.2 Cayley and Schreier graphs

As we have mentioned in the introduction above, our bridge from group theory to its geometry will be the Cayley graph. For every group there will be corresponding Cayley graph (possibly not unique), defined as follows:

Definition 2.2. Let G be a group and let $S \subseteq G$ be a generating set of G. The Cayley graph $Cay(G, S)$ of G is a graph whose vertex set is G, and set of edges is

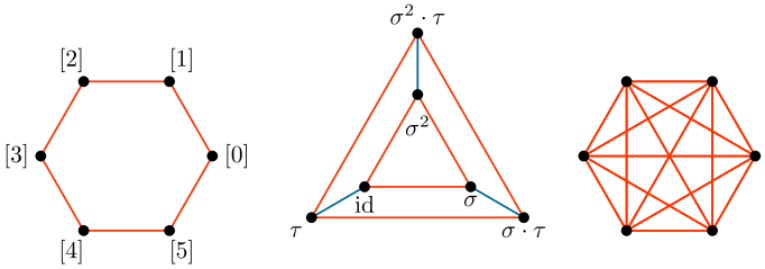
$$\{\{g, g \cdot s\} \mid g \in G, s \in (S \cup S^{-1}) \setminus \{e\}\}.$$

If $g, h \in G$ are two adjacent vertices of $Cay(G, S)$, then $h = gs$ implies that $g = hs^{-1}$. Thus the Cayley graph is a graph whose vertices are the elements of groups and, two vertices are adjacent if they differ by right multiplication by inverse of generating element.



$\mathrm{Cay}(\mathbb{Z}, \{1\})$ $\mathrm{Cay}(\mathbb{Z}, \{2, 3\})$

As seen from the example, Cayley graph of a group changes as we change the generating set. Here are more examples of Cayley graphs.



$\mathrm{Cay}(\mathbb{Z}/6, \{[1]\})$ $\mathrm{Cay}(S_3, \{\tau, \sigma\})$ $\mathrm{Cay}(S_3, S_3) \cong \mathrm{Cay}(\mathbb{Z}/6, \mathbb{Z}/6)$

Proposition 2.3 (Basic properties of Cayley graphs)**.**

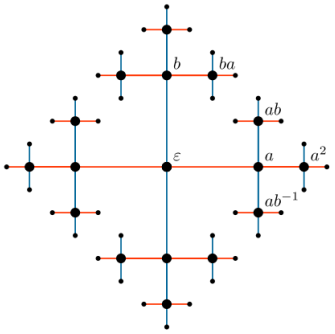
1. *Cayley graphs are regular i.e. every vertex has* $|(S \cup S^{-1}) \setminus \{e\}|$ *neighbors.*
2. *Cayley graph is locally finite if and only if* S *is finite.*
3. *Cayley graphs are connected.*

Now we will see the relation between free groups and Cayley graphs.

Theorem 2.4. *Let* F *be a free group generated by* $S \subset F$*. Then the Cayley graph of* F *with respect to the generating set* S *is a tree.*

However, the converse is not true. For instance, the Cayley graph of $\mathbb{Z}_2$ with respect to the generating set $\{1\}$ is an edge connected with two vertices. This graph is a tree but $\mathbb{Z}_2$ is not a free group.

Here's the Cayley graph of the free group on two generators $\{a, b\}$.



Another useful graph-theoretic notion is the Schreier graph defined as follows.

Definition 2.5. Let G be a group generated by $S \subset G$ and let H be a subgroup of G. The *Schreier graph* $Sch(G/H, S)$ is a graph whose vertex set is the set of right cosets $Hs = \{hs : h \in H\}$ for $s \in S$, and the set of edges is

$$\left\{\{Hg, Hgs\} : g \in G, s \in (S \cup S^{-1}) \setminus \{e\}\right\}.$$

More generally, we could define Schreier graph with respect to an action of a group G on a G-set in the same manner. Namely, if X is a G-set, then the Schreier graph $Sch(G, X, S)$ can be defined as a graph whose vertex set is X and edges are of the form $\{x, xs\}$ for $x \in X$ and $s \in S$. We notice that the Cayley graph is a Schreier graph with $H = \{e\}$ and the right multiplication of the group.

From now on, unless otherwise stated, we shall always assume that the Graph is locally finite and connected.

2.3 Markov Chain

A square (possibly infinite) matrix $A = (a_{ij})$ is called *stochastic* if all the entries are non-negative with

$$\sum_j a_{ij} = 1 \text{ for all } i.$$

Definition 2.6. Let X be a countable set and $P = (p(x, y))_{x,y \in X}$ be a stochastic matrix. Let $\lambda : X \to \mathbb{R}$ such that $\lambda(x) \geq 0$ for all $x \in X$ and $\sum_{x \in X} \lambda(x) = 1$. The matrix P is called one-step transition probability matrix and the function λ is called initial distribution. We call (X, P, λ) a Markov Chain.

By definition of the stochastic matrix, we have

$$p(x, y) \geq 0 \text{ for all } x, y \in X$$

and

$$\sum_{y \in X} p(x, y) = 1 \text{ for all } x \in X$$

Informally speaking, a Markov chain is a stochastic model on X with discrete time $t \in \mathbb{N}$. The process starts at x_0 with probability of $\lambda(x_0)$ initially i.e. at a time $t = 0$. If the random process is at the state $x(n)$ at n-th step, it moves (or stays at the same state) with the probability of $p(x, y)$, $y \in X$.
In order to formalize it, let us denote the space of trajectories by

$$\Omega := X^{\mathbb{N}} = \{\omega = (x_0, x_1, ...) : x_n \in X \text{ for all } n \in \mathbb{N}\}.$$

For fixed $y_0, y_1, ..., y_k \in X$, denote the cylinder with base as the tuple $(y_0, y_1, ..., y_k) \in X^{k+1}$ by

$$C(y_0, ..., y_k) := \{\omega = (x_0, x_1, ...) \in \Omega : x_m = y_m, 0 \leq m \leq k\}$$

This is the set of all possible ways how the Markov chain may continue after the initial steps through $y_0, ..., y_k$. With this set, we associate the probability

$$\mathbb{P}_\lambda(C(y_0, y_1, ..., y_k)) := \lambda(y_0)p(y_0, y_1)p(y_1, y_2) \dots p(y_{k-1}, y_k).$$

For $n \in \mathbb{N}_0$, let $\mathcal{A}_n$ be the σ-algebra generated by the collection of all cylinders $C(a)$ for $a \in X^{k+1}$, $k \leq n$. Thus, there is a bijection between collection of all subsets of X^{k+1} and $\mathcal{A}_n$. Furthermore, it is obvious that $\mathcal{A}_n \subset \mathcal{A}_{n+1}$. Set $\mathcal{F} = \bigcup_n \mathcal{A}_n$. Then $\mathcal{F}$ is an algebra: it clearly contains Ω and is closed with respect to taking complement and finite unions (we know that union of increasing sequence of σ-algebras is not a σ-algebra). Denote by $\mathcal{A}$ the σ-algebra generated by $\mathcal{F}$. Then well-known Kolmogorov extension theorem states that $\mathbb{P}_\lambda$ has a unique extension to a probability measure on $\mathcal{A}$. (see [12], Theorem 1.12 for Kolmogorov Extension Theorem)

Define $Z_n : \Omega \to X$ by $Z_n(\omega) := x_n$, for all $\omega = (x_0, x_1, ...) \in \Omega$ and $n \in \mathbb{N}$. So we have a sequence of X-valued random variables Z_n, $n \geq 0$, representing the random position at time $t = n$. Note that Z_n is a measurable function from (Ω, Σ) to $(X, \mathcal{P}(X))$, where $\mathcal{P}(X)$ is the σ-algebra of all subsets of X.

An event E is a measurable subset i.e. $E \in \Sigma$. For example the event of being at the position x in a given time $t = n$ is the subset

$$Z_n^{-1}(x) = \{(x_0, x_1, ...) \in \Omega : x_n = x\} = \bigcup_{y_0, y_1, ...y_n \in X} C(y_0, y_1, ...y_{n-1}, x) \in \Sigma.$$

Throughout this paper we will denote events with brackets. For instance, we will denote event in the example above by $[Z_n = x]$. For simplicity, we will use $\mathbb{P}_\lambda[Z_n = x]$ instead of $\mathbb{P}_\lambda([Z_n = x])$. If A and B are two events then the joint probability $\mathbb{P}_\lambda([Z_n = x] \cap [Z_{n+1} = y])$ will be written as $\mathbb{P}_\lambda[Z_n = x, Z_{n+1} = y]$ for any $n \in \mathbb{N}$. We denote the conditional probability of A given B by

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

We have

$$\mathbb{P}_\lambda[Z_0 = x] = \lambda(x).$$

Consider $x_0, x_1, ..., x = x_n, y = x_{n+1} \in X$ such that $\mathbb{P}_\lambda[Z_0 = x_0, ..., Z_n = x_n] > 0$. Then

$$\begin{aligned}&\mathbb{P}_\lambda[Z_{n+1} = y | Z_n = x, Z_i = x_i \text{ for all } i < n] = \\ &= \frac{\sum_{y_0,...,y_{n-1} \in X} \lambda(y_0)p(y_0, y_1) \dots p(y_{n-1}, x)p(x, y)}{\sum_{y_0,...,y_{n-1} \in X} \lambda(y_0)p(y_0, y_1) \dots p(y_{n-1}, x)} = p(x, y) \qquad (2.1)\end{aligned}$$

Now consider the events $A = [Z_{n+1} = y]$ and $B = [Z_n = x]$ in $\mathcal{A}$. We have $B = \bigcup\{C(a) \mid a \in X^{n+1}, a_n = x\}$. Assuming that $\mathbb{P}_\lambda(C(a)) > 0$ and using (2.1), we get

$$\begin{aligned}\mathbb{P}_\lambda[Z_{n+1} = y \mid Z_n = x] &= \frac{\mathbb{P}_\lambda(A \cap B}{\mathbb{P}_\lambda(B)} \\ &= \sum_{a \in X^{n+1}: a_n = x} \frac{\mathbb{P}_\lambda A \cap C(a)}{\mathbb{P}_\lambda(B)} \\ &= \sum_{a \in X^{n+1}: a_n = x} \mathbb{P}_\lambda(A \mid C(a)) \frac{\mathbb{P}_\lambda(C(a))}{\mathbb{P}_\lambda(B)} \\ &= \sum_{a \in X^{n+1}: a_n = x} p(x, y) \frac{\mathbb{P}_\lambda(C(a))}{\mathbb{P}_\lambda(B)} \\ &= p(x, y)\end{aligned}$$

for all $n \in \mathbb{N}$. Observe that the position Z_{n+1} depends on Z_n but not upon the $Z_{n-1}, ..., Z_1, Z_0$. We can formulate it as follows:

$$\mathbb{P}_\lambda[Z_{n+1} = y \mid Z_n = x, Z_{n-1} = x_{n-1}, ..., Z_0 = x_0] = \mathbb{P}_\lambda[Z_{n+1} = y \mid Z_n = x] \quad (2.2)$$

with $\mathbb{P}_\lambda[Z_n = x, Z_{n-1} = x_{n-1}, ...Z_0 = x_0] > 0$. Condition (2.2) is called the Markov property. Loosely speaking, the Markov property asserts that the future evolution of

the stochastic process depends only on the present state and does not depend on the past history.

From now on, we will fix a point $x \in X$ and the initial distribution will be the Dirac delta function δ_x, i.e. $\delta_x(y) = 1$ if $x = y$ and 0 otherwise. For simplicity, we shall write $\mathbb{P}_x$ instead of $\mathbb{P}_{\delta_x}$

Let us calculate the transition probability $\mathbb{P}_x[Z_2 = y|Z_0 = x]$ of moving from the state x to the state y in two steps. We shall factor through the first step i.e. the values of Z_1. Thus, we have

$$\begin{aligned}\mathbb{P}_x[Z_2 = y \mid Z_0 = x] &= \sum_{z\in X} \mathbb{P}_x[Z_2 = y \mid Z_1 = z]\mathbb{P}_x[Z_1 = z \mid Z_0 = x] \\ &= \sum_{z\in X} p(z,y)p(x,z) = \sum_{z\in X} p(x,z)p(z,y) \\ &= (P^2)_{xy}\end{aligned}$$

where $(P^2)_{xy}$ means the xy-th entry of the second power of the transition matrix. By induction, one can show that n-th step transition probability is the xy-th entry of n-th power of the transition matrix, i.e. $\mathbb{P}_x[Z_n = y \mid Z_0 = x] = (P^n)_{xy}$.

For $n = 0$, we set $P^0 = I = (\delta_{x,y})_{x,y\in X}$, where I is the identity matrix. And for $n \geq 1$, we denote n-th step transition probability simply by $p^{(n)}(x, y)$. We shall also write

$$P^n = \left(p^{(n)}(x,y)\right)_{x,y\in X}$$

Now we shall introduce an important property of Markov chains.

Definition 2.7. A Markov chain is called irreducible if for every $x, y \in X$ there exists an $n \in \mathbb{N}$ such that $p^{(n)}(x, y) > 0$.

We define the Green function of a Markov chain as the power series with $x, y \in X$ and with indeterminate z:

$$G(x, y|z) := \sum_{n=0}^{\infty} p^{(n)}(x, y)z^n \tag{2.3}$$

Now irreducibility yields an important result about the Green function.

Proposition 2.8. *Let (X, P, λ). For real $z > 0$, the series $G(x, y \mid z)$ converges or diverges simultaneously and for all $x, y \in X$.*

Proof. Let $x_1, y_1 \in X$ such that $G(x_1, y_1|z) < \infty$. Let $x_2, y_2 \in X$ be any states. we will show that $G(x_2, y_2|z) < \infty$. By irreducibility, there exist $k, l \in \mathbb{N} \setminus \{0\}$ such that $p^{(k)}(x_1, x_2) > 0$ and $p^{(l)}(y_2, y_1) > 0$. Now if we consider probability of random evolution through the trajectory (or path) x_1, x_2, y_2, y_1 with moving from one state to another in k, l, m amount of steps respectively, then this probability will not exceed the total probability of moving from x_1 to y_1 in $k + l + m$ steps. Thus, we have

$$p^{(k+n+l)}(x_1, y_1) \geq p^{(k)}(x_1, x_2)p^{(n)}(x_2, y_2)p^{(l)}(y_2, y_1)$$

for all $n \in \mathbb{N}$. Then we have

$$\begin{aligned} G(x_1, y_1|z) &= \sum_{m=0}^{\infty} p^{(m)}(x_1, y_1)z^m \\ &\geq \sum_{m=k+l} p^{(m)}(x_1, y_1)z^m = \sum_{n=0} p^{(k+l+n)}(x_1, y_1)z^{k+l+n} \\ &\geq p^{(k)}(x_1, x_2)p^{(l)}(y_2, y_1)z^{k+l} \sum_{n=0} p^{(n)}(x_2, y_2)z^n \\ &= p^{(k)}(x_1, x_2)p^{(l)}(y_2, y_1)z^{k+l}G(x_2, y_2|z) \end{aligned}$$

Thus, $G(x_2, y_2|z) < \infty$ □

As a consequence of Proposition 2.8, all $G(x, y|z)$, $x, y \in X, z > 0$ have the same radius of convergence $r(P) = 1/\rho(P)$, where $\rho(P)$ is given by

$$\rho(P) = \limsup_{n\to\infty} p^{(n)}(x, y)^{1/n} \in (0, 1].$$

This quantity is called spectral radius of P.

From now on we shall always assume that Markov chains are irreducible.

For all $x, y \in X$, set

$$G(x, y) := G(x, y|1) = \sum_{n=0} p^{(n)}(x, y)$$

Let us assume that a Markov chain starts at the sate x. Consider

$$V_x(y) := \sum_{n=1}^{\infty} \mathbb{1}_{\{Z_n=y\}}$$

It is a random variable and counts the number of visits to the state y starting from x. Then the expected number of visits from x to y is

$$\mathbb{E}[V_x(y)] = \mathbb{E}\sum_{n=1}^{\infty} \mathbb{1}_{\{Z_n=y\}} = \sum_{n=1}^{\infty} \mathbb{E}_x \mathbb{1}_{\{Z_n=y\}} = \sum_{n=1}^{\infty} p^{(n)}(x,y)$$

Now we see that $G(x,y)$ is actually the expected number of visits to y starting at x.

Now we define the hitting time and hitting probability. Let $x, y \in X$. The random variable $t^x : \Omega \to \mathbb{N}$ defined by

$$t^x(\omega) =: \inf\{n \geq 1 : Z_n = x\}$$

is called the hitting time. Then $t^x(\omega) = n$ indicates that the walk visits x for the first time at a time $t = n$.

We will denote the probability of visiting y for the first time in n steps after visiting x by

$$f^{(n)}(x,y) := \mathbb{P}_x[t^y = n] \tag{2.4}$$

we will call it the hitting probability.

The associated generating function will be denoted by

$$F(x,y|z) := \sum_{n=1}^{\infty} f^{(n)}(x,y) z^n$$

Lemma 2.9. *Let (X, P, λ) be an irreducible Markov Chain and $x, y \in X$ be distinct elements. Then the following holds:*

1) $G(x,y|z) = G(y,y|z)F(x,y|z)$;

2) $G(x,x|z) = G(x,x|z)F(x,x|z) + 1$;

3) If $t \in [0,\infty)$ is such that $G(x,x|t) < \infty$ and $F(x,x|t) < 1$ then

$$G(x,x|t) = \frac{1}{1 - F(x,x|t)} \tag{2.5}$$

Proof. First two results are obtained by conditioning with respect to the first visit to y. Suppose that $Z_0 = x$ and $Z_n = y$ for $n \geq 1$. Let m be the hitting time, i.e. $t^y = m$. Thus,

$$[t^y = m] = [Z_m = y, Z_l \neq y \text{ for } l = 1, 2, ..., m-1]$$

Then we have

$$p^{(n)}(x, y) = \mathbb{P}_x[Z_n = y] = \sum_{m=1}^{n} \mathbb{P}_x[Z_n = y, t^y = m]$$

But the probability of event $[Z_n = y, t^y = m]$ is equal to

$$\mathbb{P}_x[Z_n = y | t^y = m]\mathbb{P}_x[t^y = m].$$

Thus, we have

$$\begin{aligned} p^{(n)}(x, y) &= \sum_{m=1}^{n} \mathbb{P}_x[Z_n = y | Z_m = y, Z_l \neq y \; for \; l = 1, 2, ...m-1]\mathbb{P}_x[t^y = m] \\ &= \sum_{m=1}^{n} \mathbb{P}_x[Z_n = y | Z_k = y]\mathbb{P}_x[t^y = m] \\ &= \sum_{m=1}^{n} p^{(n-m)}(y, y) f^{(m)}(x, y) \end{aligned}$$

Then

$$\begin{aligned} G(x, y|z) &= \sum_{n=0}^{\infty} p^{(n)}(x, y) z^n \\ &= p^{(0)}(x, y) + \sum_{n=1}^{\infty} p^{(n)}(x, y) z^n \\ &= p^{(0)}(x, y) + \sum_{n=1}^{\infty} \sum_{m=0}^{n} p^{(n-m)}(y, y) f^{(m)}(x, y) z^n \\ &= p^{(0)}(x, y) + G(y, y|z) F(x, y|z) \end{aligned}$$

Since $p^{(0)}(x, y)$ is an entry of the matrix P^0, we have that $p^{(0)}(x, x) = 1$ and $p^{(0)}(x, y) = 0$ for distinct $x, y \in X$. Thus we have obtained 1) and 2). Lastly, we notice that

$$\sum_{n=1}^{\infty} \sum_{m=0}^{n} p^{(n-m)}(y, y) f^{(m)}(x, y) z^n$$

is the Cauchy product of two sums. We will use a fairly well-known theorem of Mertens to prove the last part of the lemma.

Theorem 2.10 (Mertens)**.** *Let* (a_n) *and* (b_n) *be two real sequences. Suppose that* $\sum_{n=0} a_n$ *converges absolutely* $S \in \mathbb{R}$ *and the sum* $\sum_{n=0} b_n$ *converges to* $T \in \mathbb{R}$*. Then their Cauchy product*

$$\sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} a_{n-m} b_m \right)$$

converges to TS

Let $t \in [0, \infty)$ such that $G(x, x|t) < \infty$ and $F(x, x|t) < 1$. Then by applying this theorem to 2), we get $G(x, x|t) = 1 + G(x, x|t)F(x, x|t)$. Since all the terms are finite and the product $G(x, x|t)F(x, x|t)$ is convergent, we can regroup the terms in last equality and obtain (2.5) . □

2.4 Recurrent and Transient Markov Chains

Definition 2.11. Let (X, P, λ) be a Markov chain. A state $x \in X$ is called recurrent if

$$\mathbb{P}_x[\exists n > 0 : Z_n = x] = 1.$$

And it is called transient if

$$\mathbb{P}_x[\exists n > 0 : Z_n = x] < 1$$

Roughly speaking, a state is recurrent if the probability of visiting the state x once it has been already visited is one, and transient otherwise.

For $x, y \in X$, define

$$K(x, y) := \mathbb{P}_x[\exists n > 0 : Z_n = y].$$

$K(x, y)$ is the probability of visiting y at least once after visiting x.

And lastly, define

$$H(x, y) := \mathbb{P}_x[Z_n = y \text{ for infinitely many } n \in \mathbb{N}],$$

the probability of visiting y infinitely many times once x has already been visited.

We notice that $K(x, x) = 1$ reduces to the definition of recurrent state. Therefore, a state x is recurrent if and only if $K(x, x) = 1$

Now notice that $\mathbb{P}_x[\exists n > 0 : Z_n = y] = \sum_{n=1}^{\infty} \mathbb{P}_x[t^y = n]$. But we denoted the hitting probability $\mathbb{P}_x[t^y = n]$ by $f^{(n)}(x, y)$. Thus, we have

$$K(x, y) = \sum_{n=1}^{\infty} f^{(n)}(x, y) = \sum_{n=0}^{\infty} f^{(n)}(x, y) = F(x, y|1)$$

Here we used the simple intuitive observation that $f^{(0)}(x, y) = 0$.

Now let us make one important observation about recurrent states. If $x \in X$ be a recurrent state. The random walk returns to this state with probability one after visiting the state $x \in X$. Clearly, that the random walk will return to the state x infinitely many times after visiting it once. Therefore the state $x \in X$ is recurrent if and only if $K(x, x) = H(x, x) = 1$.

Theorem 2.12. *Suppose (X, P, λ) is an irreducible Markov chain. For a state $x \in X$ the following are equivalent:*

(1) x is recurrent (i.e. $K(x, x) = 1$);

(2) $G(x, x) = \infty$;

(3) $H(x, x) = 1$;

(4) $H(x, x) > 0$;

Thus, x is transient if and only if $H(x, x) = 0$.

Proof. Equivalence of (1) and (2) follows from Lemma 2.9, 3). Now for $m \in \mathbb{N}$, we will define the probability $H^{(m)}(x, y)$ that random walk will visit y at least m times, after visiting x as follows

$$H^{(m)}(x, y) := \mathbb{P}_x[\exists 0 < n_1 < n_2 < ... < n_m : Z_{n_i} = y \text{ for all } i = 1, 2, ...m].$$

Observe that $H^{(1)}(x, y) = F(x, y)$ for all $x, y \in X$. Moreover, $H^{(m)}(x, y) \geq H^{(m+1)}(x, y)$ for all $m \in \mathbb{N}$ and

$$H(x, y) = \lim_{m \to \infty} H^{(m)}(x, y) \tag{2.6}$$

We can prove by induction that

$$H^{(m)}(x, x) = K(x, x)^m \text{ for all } m \geq 1. \tag{2.7}$$

Recall that

$$\mathbb{P}_x[t^y = k] = f^{(k)}(x,y) \text{ and } [t^y = k] = [Z_k = y, Z_j \neq y \text{ for } i = 1, 2, ..., k-1].$$

Then $H^{(m+1)}(x,y) =$

$$= \sum_{k=1}^{\infty} \mathbb{P}_x[t^y = k \text{ and } \exists k = n_1 < n_2 < ... < n_{m+1} : Z_{n_i} = y, i = 1, 2, ...m+1]$$
$$= \sum_{k=1}^{\infty} \mathbb{P}_x[t^y = k]\mathbb{P}_x[\exists k = n_1 < n_2 < ... < n_{m+1} : Z_{n_i} = y]$$
$$= \sum_{k:f^{(k)}(x,y)>0} f^{(k)}(x,y)\mathbb{P}_x[\exists k = n_1 < n_2 < ... < n_{m+1} :$$
$$Z_{n_i} = y, i = 1, 2, ..., m | Z_k = y \text{ and } Z_i \neq y \text{ for } j = 1, 2, ..., k-1]$$
$$= \sum_{k:f^{(k)}(x,y)>0} f^{(k)}(x,y)\mathbb{P}_y[\exists 0 < n_2 < ... < n_{m+1} : Z_{n_i} = y \text{ for all } i = 2, ..., m+1]$$
$$= \sum_{k=1}^{\infty} f^{(k)}(x,y)H^{(m)}(y,y)$$
$$= K(x,y)H^{(m)}(y,y) = K(x,y)^{m+1}.$$

Thus, we have $H^{(m)}(x,x) = K(x,x)^m$ for all $m \in \mathbb{N}$. Now that if $K(x,x) = 1$, then by (2.6) and (2.7), we obtain that $H(x,x) = 1$, proving $(3) \Rightarrow (1)$. The implication $(3) \Rightarrow (4)$ is trivial. Now suppose that $H(x,x) > 0$ and $K(x,x) < 1$. Then again by (2.6) and (2.7), we infer that $H(x,x) = 0$, a contradiction. □

Corollary 2.13. *Suppose (X, P, λ) is an irreducible Markov chain. Then either all states are recurrent or they are all transient.*

Proof. By the previous theorem, we know that a state $x \in X$ is recurrent if and only if $G(x,x) = \infty$. On the other hand, by the Proposition 2.8, the series $G(x, y|z)$ converges and diverges simultaneously for all $x, y \in X$. Therefore, (X, P, λ) is either recurrent or transient for all $x, y \in X$ □

If all states in a Markov chain are recurrent, then we call it a recurrent Markov chain. Similarly, if all states are transient, the Markov chain is called transient.

There is another noteworthy characterization of recurrence. Let P be a one-step tran-

sition matrix as usual. Define an action of P on functions $f : X \to \mathbb{R}$ by

$$Pf(x) = \sum_y p(x, y)f(y).$$

assuming that the sum $P|f|$ is finite.

Definition 2.14. A function $f : X \to \mathbb{R}$ is called superharmonic if $Pf \leq f$ pointwise, and harmonic if $Pf = f$.

Theorem 2.15 (Minimum Principle)**.** *If f is superharmonic and attains a minimum value on X, then f is a constant function.*

Proof. Let $x \in X$ such that $f(x) = \min_X f$. Also, let $z \in X$ such that $f(z) > f(x)$. Then there exists an integer $k \in \mathbb{N}$ such that $p^{(k)}(x, z) > 0$. We can prove by induction that $P^n f \leq f$ for every $n \geq 1$. Then for every k, we have $\sum_{y \in X} p^{(k)}(x, y)f(y) \leq f(x)$. On the other hand, since P^k is also stochastic matrix as we have mentioned above, we have that

$$\sum_y p^{(k)}(x, y)f(y) \leq \sum_y p^{(k)}(x, y)f(x)$$

Thus, we have the inequality

$$\sum_y p^{(k)}(x, y)(f(y) - f(x)) \leq 0.$$

Since $f(z) > f(x)$ and $p^{(k)}(x, z) > 0$, we obtain that there must be some $y \in X$ such that $f(y) - f(x) < 0$. Which is a contradiction to the assumption that $f(x)$ is the minimum of the function $f(x)$. Therefore, f is constant. □

If we apply Minimum principle to the function $-f$, then we get maximum principle for harmonic function.

Theorem 2.16. *(X, P, λ) is a recurrent irreducible Markov chain if and only if all non-negative superharmonic functions are constant.*

Proof. Let (X, P, λ) be transient. Define $f : X \to \mathbb{R}$ by $x \mapsto G(x, x)$. we will show

that f is superharmonic.

$$\begin{aligned}
Pf(x) &= \sum_y p(x,y)f(y) = \sum_y p(x,y)G(y,x) \\
&= \sum_y p(x,y)\sum_{n=0}^{\infty} p^{(n)}(y,x) = \sum_{n=0}^{\infty}\sum_y p(x,y)p^{(n)}(y,x) \\
&= \sum_{n=0}^{\infty} p^{(n+1)}(x,x) = \sum_{n=1}^{\infty} p^{(n)}(x,x) \\
&< 1 + \sum_{n=1}^{\infty} p^{(n)}(x,x) \\
&= \sum_{n=0}^{\infty} p^{(n)}(x,x) = f(x).
\end{aligned}$$

Moreover, from the above calculations, we see that f is not harmonic. If f were constant, then it would clearly be harmonic.

Now assume that (X, P, λ) is recurrent. Let $f : X \to \mathbb{R}$ be any nonnegative superharmonic function. Our first goal is showing that every superharmonic function is harmonic. Set $g : X \to \mathbb{R}$ as $g = f - Pf$. Thus, we claim that $g \equiv 0$. Suppose there exists $y \in X$ such that $g(y) > 0$. Choose $x \in X$. For each n, we have

$$\begin{aligned}
\sum_{k=0}^{n} p^{(k)}(x,y)g(y) &\leq \sum_{k=0}^{n}\sum_y p^{(k)}(x,y)g(y) \\
&= \sum_{k=0}^{n} P^k g(x) = f(x) - P^{n+1}f(x) \leq f(x).
\end{aligned}$$

Now passing to limits as $n \to \infty$ yields that $G(x,y) \leq \dfrac{f(x)}{g(y)}$ which is in contradiction with recurrence. Thus the claim is proven. For the superharmonic function $f \geq 0$ choose $x \in X$ and set $M = f(x)$. Then we claim that pointwise minimum function $h = \min\{f, M\}$ is also superharmonic. For the first case, let $y \in X$ such that

$M = f(x) < f(y)$. Then,

$$\begin{aligned} Ph(y) &= \sum_{z \in X} p(y,z)h(z) \\ &= \sum_{z:f(z)>f(x)} p(y,z)f(x) + \sum_{z:f(z)\leq f(x)} p(y,z)f(z) \\ &\leq \sum_{z:f(z)>f(x)} p(y,z)f(x) + \sum_{z:f(z)\leq f(x)} p(y,z)f(x) \\ &= \sum_{z} p(y,z)f(x) = f(x) = h(y) \end{aligned}$$

Now suppose that $f(y) < f(x)$ so that $h(y) = f(y)$.

$$\begin{aligned} Ph(y) &= \sum_{z} p(y,z)h(z) \\ &= \sum_{z:f(z)<f(x)} p(y,z)f(z) + \sum_{f(x)\leq f(z)} p(y,z)f(x) \\ &= \sum_{z} p(y,z)f(z) = Pf(y) \leq f(y) = h(y) \end{aligned}$$

Thus, the function h is also superharmonic. But as we have showed in the first part of the proof, every superharmonic function over X is harmonic. Hence, so does the function h. It then follows by the Maximum principle that the function h is constant. Since h is constant, f must be constant too. □

P acts on non-negative measures on X by

$$\mu P(y) = \sum_{x \in X} \mu(x)p(x,y)$$

We call μ excessive if $\mu P \leq \mu$ pointwise, and invariant if $\mu P = \mu$.

Now we state another useful criterion for recurrence:

Theorem 2.17. *A Markov chain (X, P, λ) is recurrent if and only if there exists an invariant measure μ such that every excessive measure is a multiple of μ.*

Proof. ($\implies$) Let μ and ν be an invariant and excessive measures, respectively. Define an irreducible Markov chain (X, Q, λ) by

$$q(y,x) = p(x,y)\frac{\mu(x)}{\mu(y)}$$

This indeed defines a Markov chain:

$$\sum_x q(y,x) = \sum_x p(x,y)\frac{\mu(x)}{\mu(y)} = \frac{1}{\mu(y)}\sum_x p(x,y)\mu(x) = \frac{\mu(y)}{\mu(y)}$$

where the last equality follows from the invariance of μ. Moreover, notice that (X,P) is irreducible if and only if (X,Q) is irreducible. We will show that $\frac{\nu}{\mu}$ is a superharmonic function so that it is constant by the virtue of Theorem 2.16. Using excessiveness of ν, we have

$$\begin{aligned}\sum_y p(x,y)\frac{\nu(y)}{\mu(y)} &= \sum_y q(y,x)\frac{\mu(y)}{\mu(x)}\frac{\nu(y)}{\mu(y)} \\ &= \frac{1}{\mu(x)}\sum_y q(y,x)\nu(y) \\ &\leq \frac{\nu(x)}{\mu(x)}\end{aligned}$$

Therefore, $\frac{\nu}{\mu}$ is a superharmonic function and hence every excessive measure is a multiple of the invariant measure. We have proved that a recurrent irreducible Markov chain, if exists, has a unique invariant measure up to multiplication by a constant. In other words, the space of invariant measures is one-dimensional for a recurrent irreducible Markov chain. Now we will explicitly present the invariant measure.
For a state $y \in X$. Denote the time of the first return to y by

$$T_y = \min\{n \in \mathbb{N} : X_n = y\} \in \mathbb{N} \cup \{\infty\}$$

We allow it to attain infinity so that the minimum is infinity by the convention. For a state $x \in X$, denote the expected number of visits to x before first visit to y by

$$\mu_y(x) = \mu(x) = \mathbb{E}_x \sum_{n=0}^{T_y - 1} \mathbb{1}_{\{X_n = x\}} \in [0,\infty]$$

We claim that μ_y is an invariant measure.
Firstly, observe that $\mu(y) = 1$ as $\mathbb{1}_{\{X_0=y\}} = 1$ and $\mathbb{1}_{\{X_i=y\}} = 0$, $i = 1, ..., T_y - 1$. Thus, we have $\mu(y) = \mathbb{E}_y 1 = 1$.
Now let $z \in X$. We will use the infinite sum trick to show that $P\mu(z) = \mu(z)$. We

know that $\mathbb{1}_{\{X_0=y\}} = \mathbb{1}_{\{X_{T_y}=y\}} = 1$. We have

$$\begin{aligned}\mu(z) &= \mathbb{E}_y \sum_{n=0}^{T_y-1} \mathbb{1}_{\{X_z\}} = \mathbb{E}_y \sum_{n=1}^{T_y} \mathbb{1}_{\{X_n=z\}} \\ &= \mathbb{E}_y \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=z, n\le T_y\}} = \sum_{n=1}^{\infty} \mathbb{P}_y[X_n = z, n \le T_y] \\ &= \sum_{\omega\in X} \sum_{n=1}^{\infty} \mathbb{P}_y[X_n = z, X_{n-1} = \omega, T_y \ge n] \\ &= \sum_{\omega} p(\omega, z)\mathbb{P}_y[X_{n-1} = \omega, T_y \ge n]\end{aligned}$$

Now we will change summation index to $m = n - 1$. Hence,

$$\begin{aligned}\mu(z) &= \sum_{\omega} p(\omega, z) \sum_{m=0}^{\infty} \mathbb{E}_y \mathbb{1}_{\{X_m=\omega, T_y\ge m+1\}} = \sum_{\omega} p(\omega, z)\mathbb{E}_y \sum_{m=0}^{T_y-1} \mathbb{1}_{\{X_m=\omega\}} \\ &= \sum_{\omega} p(\omega, z)\mu(\omega) = \mu(z).\end{aligned}$$

Moreover, we will show that $0 < \mu_y < \infty$. Let $z \in X$ be any state. By irreducibility, there exist $m, n \in \mathbb{N}$ such that $p^{(m)}(z, y) > 0$ and $p^{(n)}(y, z) > 0$. We obtain

$$\mu(z) = P\mu(z) = P^n\mu(z) = \sum_x p^{(n)}(x, z)\mu(x) \ge p^{(n)}(y, z)\mu(y) = p^{(n)}(y, z) > 0.$$

Therefore, $\mu_y > 0$.

On the other hand,

$$1 = \mu(y) = \sum_x p^{(m)}(x, y)\mu(x) \ge p^{(m)}(z, y)\mu(z).$$

Thus, $\mu(z) \le \dfrac{1}{p^{(m)}(z, y)} < \infty$. Hence, μ_y is a unique invariant measure. This completes first part of the proof.

($\Leftarrow$) Recall that we defined (X, Q) by $q(y, x) = p(x, y)\dfrac{\mu(y)}{\mu(x)}$. In addition, we can show that

$$q^{(n)}(y, x) = p^{(n)}(x, y)\frac{\mu(y)}{\mu(x)}.$$

Indeed,

$$\begin{aligned}q^{(n+1)}(y, x) &= \sum_z q^{(n)}(y, z)q(z, x) = \sum_z p^{(n)}(z, y)\frac{\mu(z)}{\mu(y)}p(x, z)\frac{\mu(x)}{\mu(z)} \\ &= \sum_z p^{(n)}(x, z)p(z, y)\frac{\mu(x)}{\mu(y)} = p^{(n+1)}(x, y)\frac{\mu(x)}{\mu(y)}\end{aligned}$$

One observes that (X, P) is recurrent if and only if (X, Q) is recurrent by the virtue of the Corollary 2.13.

Now let μ, f be an P-invariant measure and a non-negative Q-superharmonic function, respectively. Define $\nu = \mu f$, which is P-excessive:

$$\begin{aligned} P\nu(y) &= \sum_x p(x,y)f(x)\mu(x) = \sum_x q(y,x)f(x)\mu(y) \\ &= \mu(y)\sum_x q(y,x)f(x) \leq \mu(y)f(y). \end{aligned}$$

Then by the hypothesis, $\nu = c\mu$ for some constant c. Therefore, f is constant. By the Theorem 2.16 again, (X, Q) is recurrent, so is (X, P).

□

2.5 Simple Random Walks on Trees

Let X be a locally finite graph. Now, we shall consider a Markov chain with the state space consisting of vertices of X and transition probabilities $p(x, y) = \frac{1}{deg(x)}$ if $x \sim y$, and 0 otherwise. This is called the simple random walk. We also say that x is recurrent (transient) if the corresponding Markov chain is of the same property. In this section, we will calculate the spectral radius of a random walk over a homogeneous tree $\mathbb{T}_N$ i.e. a tree in which all vertices have the same degree N. Recall that a *tree* is a connected graph without cycles i.e. paths with the same initial and terminal vertex. Since a tree is a connected graph with no cycles, it can be observed that for every $x, y \in X$, there is a unique path $\pi(x, y)$ of length $d(x, y)$ connecting them.

Lemma 2.18. *Let t be any vertex in between x and y i.e., $t \in \pi(x, y)$. Then we have $F(x, y|z) = F(x, t|z)F(t, y|z)$*

Proof. As we have mentioned above, by the tree structure there is a unique path from x to y. Thus, the random walk must pass through t. Conditioning with respect to the first visit in t, we have

$$f^{(n)}(x, y) = \sum_{k=0}^{n} f^{(k)}(x, t)f^{(n-k)}(t, y).$$

Therefore,

$$\begin{aligned} F(x,y|z) &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} f^{(k)}(x,t)f^{(n-k)}(t,y)z^n \\ &= \sum_{n=0}^{\infty} f^{(n)}(x,t)z^n \sum_{n=0}^{\infty} f^{(n)}(t,y)z^n = F(x,t|z)F(t,y|z). \end{aligned}$$

□

Lemma 2.19. *Let (X,P,λ) denote the simple random walk (Markov chain) on the homogeneous tree $\mathbb{T}_N$. For any $x,y \in X$, one has*

$$G(x,y|z) = \frac{2(N-1)}{N-2+\sqrt{N^2-4(N-1)z^2}} \left(\frac{N-\sqrt{N^2-4(N-1)z^2}}{2(N-1)z} \right)^{d(x,y)}$$

As a consequence, $\rho(P) = \frac{2\sqrt{N-1}}{N}$.

Proof. Since we are considering a homogeneous tree, it is clear that $F(x,y|z)$ is the same for every pair $x,y \in X$ such that $x \sim y$. Let us denote this quantity by $F(z)$ as this quantity now only depends on z in a sense. Thus, if $x = x_0, x_1, ..., x_n = y$ is the unique path connecting x and y, then by the previous lemma we obtain

$$F(x,y|z) = F(x,x_1|z)F(x_1,x_2|z)\dots F(x_{n-1},y) = F(z)^{d(x,y)}$$

Now we will need the following lemma.

Lemma 2.20. *For $x \neq y$, we have*

$$F(x,y|z) = \sum_t p(x,t)zF(t,y|z).$$

Proof. This identity can be proven by factoring through the first step:

$$\begin{aligned} F(x,y|z) &= \sum_{n=0}^{\infty} f^{(n)}(x,y)z^n = \sum_{n=1}^{\infty} f^{(n)}(x,y)z^n \\ &= \sum_{n=0}^{\infty} f^{(n+1)}(x,y)z^{(n+1)} = \sum_{n=0}^{\infty}\sum_t p(x,t)zf^{(n)}(t,y)z^n \\ &= \sum_t p(x,t)z\sum_{n=0}^{\infty} f^{(n)}(t,y)z^n = \sum_t p(x,t)zF(t,y|z) \end{aligned}$$

□

Now applying previous the lemma we just proved yields

$$F(z) = F(x, y|z) = \sum_{t\sim x} \frac{1}{N} zF(z)^{d(t,y)}.$$

Notice that t can be y itself too since x and y are adjacent vertices from the hypothesis. In this case, $d(t, y) = 0$. Otherwise, $t \sim x \sim y$ and hence $d(t, y) = 2$. Therefore,

$$\sum_{t\sim x} \frac{1}{N} zF(z)^{d(t,y)} = \frac{1}{N} z + \frac{N-1}{N} zF(z)^2$$

We observe that we have obtained a quadratic equation in $F(z)$. By taking into account the fact that $F(0) = 0$, we get

$$F(z) = \frac{1}{2(N-1)z} \left(N - \sqrt{N^2 - 4(N-1)z^2} \right).$$

Now using the Lemma 2.20, we get that

$$F(x, x|z) = \sum_t p(x, t) zF(t, x) = \sum_{t\sim x} \frac{1}{N} zF(t, x) = \frac{1}{N} zNF(z) = zF(z)$$

Hence, we can calculate $G(x, x|z)$ from part $3)$ of the the lemma. And using part $1)$, we can calculate $G(x, y|z)$. By Pringsheim's theorem (see Hille[5], p. 133), the radius of convergence must be the smallest positive singularity. Hence, $N^2-4(N-1)z^2 = 0$. From this we get that $r(P) = 1/\rho(P) = \frac{N}{2\sqrt{N-1}}$. Consequently,

$$\rho(P) = \frac{2\sqrt{N-1}}{N}.$$

□

Notice that if $\rho(P) < 1$, the Markov chain is transient (the converse is not true). As a consequence, we deduce that all simple random walks over N-regular homogeneous trees for $N \geq 3$ are transient.

2.6 Random Walks on Finitely Generated Groups

Definition 2.21. Let G be a group and S be a finite symmetric generating subset of G. The *word length* of an element $g \in G$ is defined to be

$$\ell_S(g) := \min\{n \geq 0 : g = g_1 g_2 \cdots g_n, g_i \in S, 1 \leq i \leq n\}.$$

In other words, it is the minimal integer $n \geq 0$ such that g can be written as a product of n elements from S.

Consider the function $d_S : G \times G \to \mathbb{N}$ defined by

$$d_S(g, h) := \ell(g^{-1}h)$$

for all $g, h \in G$.

Proposition 2.22. *The map d_S is a left-invariant metric on G.*

Proof. It is clear that $d_S(g, h) = 0$ if and only if $g = h$
Let $d_S(g, h) = \ell(g^{-1}h) = n$ for $g, h \in G$. Then $g^{-1}h = g_1 \cdots g_n$ for some $g_1, ..., g_n \in S$. On the other hand, $h^{-1}g = (g^{-1}h)^{-1} = g_n^{-1} \cdots g_1^{-1}$. Thus, $d_S(h, g) \leq d_S(g, h)$. But if we reverse the roles of g and h, we obtain $d_S(g, h) \leq d_S(h, g)$. Therefore $d_S(g, h) = d_S(h, g)$.
Let $g, h, k \in G$ such that $d_S(g, h) = n$, $d_S(h, k) = m$. Suppose that $g^{-1}h = g_1 \cdots g_n$ and $h^{-1}k = h_1 \cdots h_m$. But $g^{-1}k = g_1 \cdots g_n h_1 \cdots h_m$. Henceforth, $d_S(g, k) \leq m + n = d_S(g, h) + d_S(h, k)$.
Moreover, it can easily be seen that this metric is also invariant:

$$d_S(kg, kh) = \ell((kg)^{-1}kh) = \ell(g^{-1}h) = d_S(g, h)$$

□

Definition 2.23. Let (G, P, λ) be an irreducible Markov chain on G. We say that (G, P, λ) is a *random walk* on G if the following properties are satisfied:

(1) $p(g, h) = p(h, g)$ for all $g, h \in G$ (symmetry)

(2) $p(kg, kh) = p(g, h)$ for all $k, g, h \in G$ (space homogeneity)

(3) $\sum_{h \in G} d_S(g, h)^2 p(g, h) < \infty$ for all $g \in G$ (finiteness of the second moment).

Notice that for any $k, g \in G$, by the space homogeneity and left invariance of the

word metric, we have

$$\begin{aligned}\sum_{h\in G} d_S(g,h)^2 p(g,h) &= \sum_{h\in G} d_S^2((kg^{-1}g, kg^{-1}h)p(kg^{-1}g, kg^{-1}h) \\ &= \sum_{h\in G} d_S(k, kg^{-1}h)p(k, kg^{-1}h) \\ &= \sum_{h\in G} d_S(k,h)p(k,h)\end{aligned}$$

since the second coordinate will run through G again. Suppose that (G, P, λ) is a Markov chain satisfying $(1), (2)$, and the condition

$$p(g,h) = 0 \text{ if } g^{-1}h \notin S.$$

It is clear that the Markov chain will also satisfy the finiteness of the second moment since there will be only finitely many non-zero terms. Hence, it is a random walk. This is equivalent to the condition that if $p(g,h) > 0$, then $d_S(g,h) \leq 1$. In this case, we say that the random walk is of *nearest neighbor* type with respect to S.

Let (G, P, λ) be a random walk. One defines an atomic probability measure on $(G, P) = (G, P, \lambda)$ by setting

$$\mu(g) := p(1_G, g) \tag{2.8}$$

for all $g \in G$. Define *n-th convolution* of μ by setting $\mu^{(1)} := \mu$ and $\mu^{(n)}(g) = \sum_{h\in G} \mu^{(n-1)}(gh^{-1})\mu(h)$ for all $g \in G$. It is easy to see that convolutions are also atomic probability measure. Indeed, using the fact that P^n, $n \in \mathbb{N}$ are stochastic matrices, we show that it is a probability measure:

$$\begin{aligned}\sum_{g\in G} \mu^{(n)}(g) &= \sum_g \sum_h \mu^{(n-1)}(gh^{-1})\mu(h) \\ &= \sum_g \sum_h p^{(n-1)}(1_G, gh^{-1})p(1_G, h) \\ &= \sum_h p(1_G, h) \sum_g p^{(n-1)}(1_G, gh^{-1}) = 1.\end{aligned}$$

By the way the convolutions are defined, it is clear that they are atomic.
Moreover, we claim that $\mu^{(n)}(g) = p^{(n)}(1_G, g)$. Suppose that the equality holds for n.

Then

$$\begin{aligned}\mu^{(n+1)}(g) &= \sum_h \mu^{(n)}(gh^{-1})\mu(h) = \sum_h p^{(n)}(1_G, gh^{-1})p(1_G, h) \\ &= \sum_h p^{(n)}(1_G, gh^{-1})p(h^{-1}, 1_G) = \sum_{h^{-1}} p^{(n)}(1_G, gh^{-1})p(gh^{-1}, g) \\ &= p^{(n+1)}(1_G, g)\end{aligned}$$

which concludes the proof by induction. We have the following properties analogous to those in the Definition 2.23.

Proposition 2.24. *Let μ be an atomic probability measure defined as in (2.8). Then*

(1') $\mu(g) = \mu(g^{-1})$ *for all* $g \in G$ *(symmetry);*

(2') $\bigcup_{n=0}^{\infty} supp(\mu^{(n)}) = G$ *(irreducibility);*

(3') $\sum_{g\in G} d_S(1_G, g)^2\mu(g) = \sum_{g\in G} \ell(g)^2\mu(g) < \infty$ *for all* $g \in G$ *(finiteness of the second moment).*

Proof. For proving these properties, we will use the properties (1) and (2) from the Definition 2.23.

For $(1')$, $\mu(g) = p(1_G, g) = p(g^{-1}, g^{-1}g) = p(g^{-1}, 1_G) = p(1_G, g^{-1}) = \mu(g^{-1})$.
As for $(2')$, we only have to show that any element is in $\bigcup_{n=0}^{\infty} supp(\mu^{(n)}$. Irreducibility and the fact that $\mu^{(n)}(g) = p^{(n)}(1_G, g)$ together imply that there exists $n \in \mathbb{N}$ such that $\mu^{(n)}(g) = p^{(n)}(1_G, g) > 0$. Thus, proving $(2')$.
Finally, $(3')$ immediately follows. □

For the simple random walk, it is clear that $\mu(g) = p(1_G, g) > 0$ if and only if $g \in S$. Therefore, $supp(\mu) = S$ and the associated measure is $\mu = \dfrac{1}{|S|}\mathbb{1}_S$

Conversely, given a probability measure satisfying (1'),(2'), and (3') above, one can define a random walk (G, P) on G:

Proposition 2.25. *Let μ be a probability measure satisfying (1'),(2'), and (3') above. One can define a random walk (G, P) by setting*

$$p(g, h) := \mu(g^{-1}h)$$

for all $g, h \in G$. Moreover, (G, P) is irreducible.

Proof. (1) $p(g,h) = \mu(g^{-1}h) = \mu(h^{-1}g) = p(h,g)$ for all $g, h \in G$

(2) $p(kg, kh) = \mu((kg)^{-1}kh) = \mu(g^{-1}h) = p(g,h)$ for all $g, h, k \in G$

(3) Using left invariance of the word metric, we have

$$\begin{aligned}\sum_h d_S(g,h)^2 p(g,h) &= \sum_h d_S(g,h)^2 \mu(g^{-1}h) \\ &= \sum_h d(1_G, g^{-1}h)^2 \mu(g^{-1}h) < \infty\end{aligned}$$

for all $g, h \in G$.

Moreover, we again have $p^{(n)}(g,h) = \mu^{(n)}(g^{-1}h)$:

$$\begin{aligned}\mu^{(n+1)}(g^{-1}h) &= \sum_{k\in G} \mu^{(n)}(g^{-1}hk^{-1})\mu(k) \\ &= \sum_k \mu((h^{-1}g)^{-1}k^{-1})\mu(k^{-1}) \\ &= \sum_k p^{(n)}(h^{-1}g, k^{-1})p(1_G, k^{-1}) \\ &= \sum_k p^{(n)}(g, hk^{-1})p(k^{-1}, 1_G) \\ &= \sum_k p^{(n)}(g, hk^{-1})p(hk^{-1}, h) \\ &= p^{(n+1)}(g,h)\end{aligned}$$

Now let $g, h \in G$ be any two elements. Then by (2'), there exists an $n \in \mathbb{N}$ such that $g^{-1}h \in supp(\mu^{(n)})$, i.e., $p^{(n)}(g,h) > 0$. Thus, the random walk is irreducible. □

2.7 Reversible Markov Chains and Networks

Let X be an infinite, countable set without assuming that it has a locally finite graph structure.

Definition 2.26. A Markov chain (X, P, λ) is called reversible if there exists a map $m : X \to (0, \infty)$ such that

$$m(x)p(x,y) = m(y)p(y,x) \text{ for all } x, y \in X.$$

In this case, the associated quantity

$$a(x,y) := m(x)p(x,y) = m(y)p(y,x) = a(y,x)$$

is called the conductance between x and y for $x, y \in X$, and $m(x)$ is called the total conductance at x. Notice that $m(\cdot)$ is an invariant measure with respect to P. Indeed,

$$mP(y) = \sum_{x\in X} m(x)p(x,y) = \sum_{x\in X} m(y)p(y,x) = m(y).$$

Conversely, if there exists a symmetric function $a : X \times X \to [0,\infty)$ such that $m(x) = \sum_y a(x,y)$, setting $p(x,y) = \dfrac{a(x,y)}{m(x)}$ yields a reversible Markov Chain. We also observe that conductances uniquely determine the Markov chain. Indeed,

$$p(x,y) = \frac{a(y,x)}{m(x)} = \frac{a(y,x)}{\sum_z a(x,z)}.$$

Example 2.27. Let (G, P, λ) be a simple random walk, where G is generated by finite set S. Then by symmetry, we have $p(g,h) = p(h,g)$ for all $g, h \in G$. Thus, in the case of simple random walk on finitely generated groups, $m(\cdot) = 1$.

Equipping X with an edge set such that (X, P, λ) is of the nearest neighbor type makes theoretical building more convenient. Therefore, $\{x, y\} \in E$ if and only if $a(x,y) > 0$. In this case (X, E) is connected when viewed as a graph structure, but it is not necessarily locally finite. For every edge, we will fix an orientation once and for all. For an edge $e \in E(P)$, we will specify one of its two endpoints as initial and the other one as terminal vertex. The usual choice for the notation is $e = \{e^+, e^-\}$. Although, the rest of the section will not depend on the particular choice of initial and terminal vertices, let us fix e^+ for the initial vertex and e^- for the terminal vertex. One should think of functions on E as a flow moving from e^+ to e^-.

Definition 2.28. Let $e = \{e^+, e^-\}$ be any edge. The function $r : E \to (0,\infty)$ defined by

$$r(e) := \frac{1}{a(e^+, e^-)}$$

for all $e \in E$ is called the *resistance*. We call the triple $\mathcal{N} = (X, E, r)$ a *network*.

The network $\mathcal{N}$ cen be thought of as an infinite electrical network where the edges are wires with associated resistances $r(e)$. The network (X, E, r) is called recurrent

if (X, P) is recurrent.

Notice that the total conductance is totally defined by resistance of edges whose initial vertex is x, i.e.,

$$m(x) = \sum_{\substack{e \in E \\ e^+ = x}} \frac{1}{r(e)}$$

We may see $\mathcal{N}$ as an infinite network where each edge is a wire with associated resistance. Viewing these notions in this manner will help us to interpret the upcoming results.

Recall that $\mathbb{R}^X$ denotes the set of all functions from X to $\mathbb{R}$. Consider the Hilbert spaces

$$\ell^2(X, m) = \{f \in \mathbb{R}^X : \sum_{x \in X} m(x) f^2(x) < \infty\}$$

and

$$\ell^2(E, r) = \{u \in \mathbb{R}^E : \sum_{e \in E} r(e) u^2(e) < \infty\}$$

with the inner products given by

$$\langle f, g \rangle_X = \sum_{x \in X} m(x) f(x) g(x)$$

and

$$\langle u, v \rangle_E = \sum_{e \in E} r(e) u(e) v(e)$$

respectively. Then we have the corresponding ℓ^2 norms defined by

$$||f||_X = \sqrt{\langle f, f \rangle_X} \text{ for all } f \in \ell^2(X, m)$$

and

$$||u||_E = \sqrt{\langle u, u \rangle_E} \text{ for all } u \in \ell^2(E, r).$$

Now consider the linear operator $\nabla : \mathbb{R}^X \to \mathbb{R}^E$ given by

$$(\nabla f)(e) := \frac{f(e^+) - f(e^-)}{r(e)}$$

for all $e \in E$ and $f \in \mathbb{R}^X$. It id called the *nabla operator* of $\mathcal{N} = (X, E, r)$.

One observes that $\nabla(\ell^2(X,m)) \subset \ell^2(E,r)$. Indeed,

$$\begin{aligned}\langle \nabla(f), \nabla(f) \rangle_E &= \sum_{e \in E} \frac{(f(e^+) - f(e^-))^2}{r(e)} \\ &= \frac{1}{2} \sum_{x,y \in X} m(x)p(x,y)(f(x) - f(y))^2\end{aligned}$$

where second equality comes from the fact that

$$\frac{1}{r(e)} = a(e^+, e^-) = m(e^+)p(e^+, e^-) = m(e^-)p(e^-, e^+).$$

Thus, changing summation to all pairs of $\{x, y\}$ over would result in doubling the original sum. Now note that $(f(x) + f(y))^2 \geq 0$ implies $-2f(x)f(y) \leq f(x)^2 + f(y)^2$. Therefore,

$$\begin{aligned}\langle \nabla(f), \nabla(f) \rangle_E &\leq \frac{1}{2} \sum_{x,y \in X} m(x)p(x,y)2(f(x)^2 + f(y)^2) \\ &= \sum_{x,y \in X} m(x)p(x,y)f(x)^2 + \sum_{x,y \in X} m(y)p(y,x)f(y)^2 \\ &\leq \sum_{x,y \in X} m(x)f(x)^2 + \sum_{x,y \in X} m(y)f(y)^2 \\ &= 2\|f\|_X^2 < \infty.\end{aligned}$$

Moreover, one observes that the norm of the nabla operator restricted to the Hilbert space $\ell^2(X,m)$ is bounded by $\sqrt{2}$, i.e., $\|\nabla|_{\ell^2(X,m)}\| \leq \sqrt{2}$.

Let $e \in E$ and $u \in \mathbb{R}^E$. Consider

$$(\nabla^* u)(x) := \frac{1}{m(x)} \left(\sum_{\substack{e \in E \\ e^+ = x}} u(e) - \sum_{\substack{e \in E \\ e^- = x}} u(e) \right).$$

We may see u as a flow in the network. Then

$$\sum_{\substack{e \in E \\ e+ = x}} u(e) - \sum_{\substack{e \in E \\ e^- = x}} u(e)$$

is the difference between the amount of flow moving and flowing into x. Thus, roughly speaking, $m(x)\nabla^* u(x)$ is the "loss" at x.

Remark. For a locally finite graph structure, there will be a finite number of edges such that $e^+ = x$ and $e^- = x$. Therefore, it will be a finite sum for every $u \in \mathbb{R}^X$ and $x \in X$.

We must show that above-written sum is convergent so that ∇^* defines a linear operator from $\ell^2(X,m)$ to $\mathbb{R}^X$.

For $x, y \in X$, define $\sigma_x(y) = \frac{1}{m(x)}$ if $y = x$ and 0 otherwise. Clearly, $\sigma_x \in \ell^2(X,m)$ and $\langle f, \sigma_x \rangle_X = f(x)$ for all $f \in \ell^2(X,m)$. Then $\nabla\sigma_x \in \ell^2(E,r)$. For $u \in \ell^2(E,r)$, we have

$$\begin{aligned}\langle u, \nabla\sigma_x \rangle_E &= \sum_{e\in E} u(e)\left(\sigma_x(e^+) - \sigma_x(e^-)\right) \\ &= \frac{1}{m(x)} \sum_{\substack{e\in E \\ e^+=x}} u(e) - \sum_{\substack{e\in E \\ e^-=x}} u(e).\end{aligned}$$

Therefore, $(\nabla^* u)(x) = \langle u, \nabla\sigma_x \rangle_E \in \mathbb{R}$ for every $x \in X$.

Lemma 2.29. *We have $\nabla^*(\ell^2(E,r)) \subset \ell^2(X,m)$. Furthermore, $\nabla^*\big|_{\ell^2(X,m)}$ is the adjoint of the nabla operator $\nabla\big|_{\ell^2(E,r)}$, i.e.*

$$\langle \nabla f, u \rangle_E = \langle f, \nabla^* u \rangle_X$$

for all $f \in \ell^2(X,m)$ and $u \in \ell^2(E,r)$.

Proof. We will show that $\nabla^*\big|_{\ell^2(X,m)}$ coincides with $\nabla\big|_{\ell^2(E,r)}$ in a dense subspace of E. Recall that the nabla operator was bounded, therefore it is continuous. Then continuity will imply that they are identical since they coincide on a dense subset.

Let $f \in \ell^2(X,m)$ and $u \in \mathbb{R}^E$ of finite support. We obtain

$$\begin{aligned}\langle \nabla f, u \rangle_E &= \sum_{e\in E} (f(e^+) - f(e^-))u(e) \\ &= \sum_{e\in E} f(e^+)u(e) - \sum_{e\in E} f(e^-)u(e) \\ &= \sum_{x\in X} f(x) \sum_{\substack{e\in E \\ e^+=x}} u(e) - \sum_{x\in X} f(x) \sum_{\substack{e\in E \\ e^-=x}} u(e) \\ &= \sum_{x\in X} f(x) \left(\sum_{\substack{e\in E \\ e^+=x}} u(e) - \sum_{\substack{e\in E \\ e^-=x}} u(e) \right) \\ &= \sum_{x\in X} m(x) f(x) \frac{1}{m(x)} \left(\sum_{\substack{e\in E \\ e^+=x}} u(e) - \sum_{\substack{e\in E \\ e^-=x}} u(e) \right) \\ &= \langle f, \nabla^* u \rangle .\end{aligned}$$

□

Consider the space $\mathcal{D}(\mathcal{N})$ of all functions $f : X \to \mathbb{R}$ such that $\nabla f \in \ell^2(E, r)$. Note that $\ell^2(X, m) \subset \mathcal{D}(\mathcal{N})$ and f does not have to be in $\ell^2(X, m)$ necessarily.

Definition 2.30. Let $f \in \mathcal{D}(\mathcal{N})$. Then *Dirichlet norm* of the function f is defined to be

$$D(f) = \langle \nabla f, \nabla f \rangle_E .$$

Note that

$$\begin{aligned} D(f) = \langle \nabla f, \nabla f \rangle_E &= \sum_{e \in E} \frac{(f(e^+) - f(e^-))^2}{r(e)} \\ &= \frac{1}{2} \sum_{x,y \in X} m(x)p(x, y)(f(x) - f(y))^2. \end{aligned}$$

This is a quasi-norm (sometimes it is also called *Dirichlet seminorm*. Its kernel is the set of all constant functions on X. We equip the space $\mathcal{D}(\mathcal{N})$ with inner product by choosing a reference point $x_0 \in X$ and setting

$$(f, g)_D = (f, g)_{D,x_0} = \langle \nabla f, \nabla g \rangle_E + f(x_0)g(x_0).$$

The associated norm will be denoted by $|| \cdot ||_D$.

Lemma 2.31. *The space $\mathcal{D}(\mathcal{N})$ is a Hilbert space. Moreover, the convergence in the associated norm $|| \cdot ||_D$ implies pointwise convergence.*

Proof. In order to prove the lemma, we need an inequality.
Let $x_0 \neq x \in X$. By connectedness, there exists a path $x_0, x_1, ..., x_n = x \in X$ such that $e_i = [x_{i-1}, x_i] \in E$. For $f \in \mathcal{D}(\mathcal{N})$, we have

$$\begin{aligned} (f(x) - f(x_0))^2 &= \left(\sum_{i=1}^{n} f(x_i) - f(x_{i-1}) \right)^2 \\ &= \left(\sum_{i=1}^{n} \frac{(f(x_i) - f(x_{i-1}}{\sqrt{r(e)}} \sqrt{r(e)} \right)^2 \end{aligned}$$

Using Cauchy-Schwarz inequality, we get

$$(f(x) - f(x_0))^2 \leq \sum_{i=1}^{n} \frac{(f(x_i) - f(x_{i-1})^2}{r(e)} \sum_{i=1}^{n} r(e_i) \leq D(f) \sum_{i=1}^{n} r(e_i)$$

Let $C(x) = \sum_{i=1}^{n} r(e_i)$. Hence, we have

$$(f(x) - f(x_0))^2 \leq C(x)D(f).$$

Let $(f)_n$ be a Cauchy sequence in $\mathcal{D}(\mathcal{N})$. This yields that $(f_n(x))$ is convergent in $\mathbb{R}$ for all $x \in X$. Indeed, since

$$||f_n - f_m||_D^2 = D(f_n - f_m) + (f_n(x_0) - f_m(x_0))^2$$

we see that $(f_n(x_0))$ is Cauchy sequence in $\mathbb{R}$. Moreover, for $f, g \in \mathcal{D}(\mathcal{N})$ we have

$$\begin{aligned} |f(x) - g(x)| &= |f(x) - f(x_0) + f(x_0) + g(x_0) - g(x) - g(x_0)| \\ &\leq |(f-g)(x) - (f-g)(x_0)| + |(f-g)(x_0)| \\ &\leq \sqrt{C(x)D(f-g)} + |(f-g)|(x_0)|. \end{aligned} \tag{2.9}$$

From (2.9), we obtain that $(f_n(x))$ is Cauchy for every $x \in X$.

Hence, there exists a function $h \in \mathbb{R}^X$ such that $f_n(x)$ converges to $h(x)$ for all $x \in X$.
We claim that $h \in \mathcal{D}(\mathcal{N})$ and (f_n) converges to $h \in \mathcal{D}(\mathcal{N})$.
On the other hand, ∇f_n is Cauchy in $\ell^2(E, r)$ as

$$||\nabla f_n - \nabla f_m||_E^2 = D(f_n - f_m).$$

Thus, $\nabla f_n \to u$ for some $u \in \ell^2(E, r)$. But then we must have $\nabla h = u$ since

$$\begin{aligned} ||\nabla f_n - \nabla h||_E^2 &= \sum_{e \in E} r(e)(\nabla f_n - \nabla h)^2 \\ &= \sum_{e \in E} \frac{(f_n - h)(e^+) + (h - f_n)(e^-)}{r(e)} \to 0. \end{aligned}$$

Then $\nabla h = u$ and by the definition of $\mathcal{D}(\mathcal{N})$, we have that $h \in \mathcal{D}(\mathcal{N})$. Now we see that f_n converges to h in $\mathcal{D}(\mathcal{N})$. Indeed,

$$||f_n - h||_D = D(f_n - h) + (f_n - h)^2(x_0) \to 0$$

as $n \to \infty$. □

Proposition 2.32. *For every $f \in \mathcal{D}(\mathcal{N})$ and $x \in X$,*

$$(\nabla^*\nabla f)(x) = f(x) - Pf(x) = (I - P)f(x).$$

Proof. This proof is just a matter of technicality, nevertheless we will go through it.

$$\begin{aligned}
(\nabla^*\nabla f)(x) &= \frac{1}{m(x)} \left(\sum_{\substack{e \in E \\ e^+ = x}} \nabla f(e) - \sum_{\substack{e \in E \\ e^- = x}} \nabla f(e) \right) \\
&= \frac{1}{m(x)} \left(\sum_{\substack{e \in E \\ e^+ = x}} \frac{f(e^+) - f(e^-)}{r(e)} - \sum_{\substack{e \in E \\ e^- = x}} \frac{f(e^+) - f(e^-)}{r(e)} \right) \\
&= \sum_{\substack{e \in E \\ e^+ = x}} \frac{f(e^+) - f(e^-)}{m(e^+)r(e)} - \sum_{\substack{e \in E \\ e^- = x}} \frac{f(e^+) - f(e^-)}{m(e^-)r(e)} \\
&= \sum_{\substack{e \in E \\ e^+ = x}} (f(e^+) - f(e^-))p(e^+, e^-) - \sum_{\substack{e \in E \\ e^- = x}} (f(e^+) - f(e^-))p(e^-, e^+) \\
&= \sum_{\substack{e \in E \\ e^+ = x}} (f(x) - f(e^-))p(x, e^-) - \sum_{\substack{e \in E \\ e^- = x}} (f(e^+) - f(x)p(x, e^+) \\
&= f(x) \left(\sum_{\substack{e \in E \\ e^+ = x}} p(x, e^-) + \sum_{\substack{e \in E \\ e^- = x}} p(x, e^+) \right) - \\
&\left(\sum_{\substack{e \in E \\ e^+ = x}} f(e^-)p(x, e^-) + \sum_{\substack{e \in E \\ e^- = x}} f(e^+)p(x, e^+) \right)
\end{aligned}$$

Now observe that

$$\sum_{\substack{e \in E \\ e^+ = x}} p(x, e^-) + \sum_{\substack{e \in E \\ e^- = x}} p(x, e^+) = \sum_{\substack{y \in X \\ \{x,y\} \in E}} p(x, y).$$

And similarly,

$$\sum_{\substack{e \in E \\ e^+ = x}} f(e^-)p(x, e^-) + \sum_{\substack{e \in E \\ e^- = x}} f(e^+)p(x, e^+) = \sum_{\substack{y \in X \\ \{x,y\} \in E}} p(x, y)f(y)$$

Therefore,

$$\begin{aligned}(\nabla^*\nabla f)(x) &= f(x)\sum_{\substack{y\in X\\ \{x,y\}\in E}} p(x,y) - \sum_{\substack{y\in X\\ \{x,y\}\in E}} p(X,y)f(,y)\\ &= f(x)\sum_{y\in X} p(x,y) - \sum_{y\in X} p(x,y)f(y)\\ &= f(x) - Pf(x).\end{aligned}$$

□

Thus, $P : \mathcal{D}(\mathcal{N})$ defines a linear operator. Furthermore it is a bounded operator over the subspace $\ell^2(X,m)$. Indeed, let $f \in \ell^2(X,m)$. Then,

$$\begin{aligned}||Pf||_X^2 &= \sum_{x\in X} m(x)(Pf(x))^2\\ &= \sum_x m(x)\left(\sum_y p(x,y)f(y)\right)^2\\ &= \sum_x m(x)\sum_y p(x,y)f(y)\sum_z p(x,z)f(z)\\ &= \sum_x m(x)\sum_y p(x,y)\sum_z p(x,z)f(y)f(z)\end{aligned}$$

Using the inequality on Arithmetic-Geometric means, we have

$$\begin{aligned}||Pf||_X^2 &\le \frac{1}{2}\sum_x m(x)\sum_y p(x,y)\sum_z p(x,z)(f^2(y)+f^2(z))\\ &= \frac{1}{2}\sum_x m(x)\left(\sum_y p(x,y)f^2(y) + \sum_z p(x,z)f^2(z)\right)\\ &= \sum_x m(x)\sum_y p(x,y)f^2(y) = \sum_x\sum_y m(x)p(x,y)f^2(y)\end{aligned}$$

Using reversibility, we get

$$\begin{aligned}||Pf||_X^2 &= \sum_x\sum_y m(y)p(y,x)f^2(y)\\ &= \sum_y m(y)f^2(y)\sum_x p(y,x) = \sum_y m(y)f^2(y)\\ &= ||f||_X^2.\end{aligned}$$

Thus, $P : \ell^2(X,m) \to \ell^2(X,m)$ is a bounded operator with $||P|| \le 1$. It is called the Markov operator of the Markov chain (X,P).

We denote the linear space of finitely supported functions on X by $\ell_0(X)$, and its closure in $\mathcal{D}(\mathcal{N})$ by $\mathcal{D}_0(\mathcal{N})$. It can be shown that $\ell^2(X,m) \subset \mathcal{D}_0(\mathcal{N})$ since we had $||f||_D \leq \sqrt{2}||f||_X$.

For $A \subset X$, we denote by $P_A = (p_A(x,y))_{\{x,y\}\in X}$ the restriction of P to A:

$$p_A(x,y) = p(x,y), \text{ if } x,y \in A \text{ and } p(x,y) = 0, \text{ otherwise.}$$

In other words, P_A is the transition probability remaining in A. Notice that for all $x, y \in X$, we have $0 \leq p_A(x,y) \leq p(x,y)$. One has the equality $p_A(x,y) = p(x,y)$ if and only if $x, y \in A$.

Proposition 2.33. *In operator notation we have*

$$(I_A - P_A)G_A = I_A.$$

Proof. It is clear that $I_A G_A = G_A$. For $P_A G_A$, considering the following calculation,

$$\begin{aligned} P_A(G_A f)(x) &= \sum_{y\in X} p_A(x,y) G_A f(y) \\ &= \sum_{y\in A} p_A(x,y) \sum_{z\in A} G_A(y,z) f(z) \\ &= \sum_{y\in A} p_A(x,y) \sum_{z\in A} \sum_{n=0}^{\infty} p_A^n(y,z) f(z) \\ &= \sum_{z\in A} f(z) \sum_{n=0}^{\infty} \sum_{y\in A} p_A(x,y) p_A^{(n)}(y,z) \\ &= \sum_{z\in A} f(z) \sum_{n=0}^{\infty} p_A^{(n+1)}(x,z) \end{aligned}$$

we obtain that

$$(I_A - P_A)G_A = \sum_{n=0}^{\infty} p_A^{(n)}(x,y) - \sum_{n=0}^{\infty} p_A^{(n+1)}(x,y) = p^{(0)}(x,y) = \delta_{x,y}$$

□

We can define n-th step probabilities inductively as follows:

$$p_A^{(n)}(x,y) = \sum_{z\in A} p_A(x,z) p_A^{(n-1)}(z,y)$$

for all $x, y \in X$. Similarly, for all $x, y \in X$, we have $0 \leq p_A^{(n)}(x,y) \leq p^{(n)}(x,y)$. Note that for $n \geq 2$, we may have $p_A^{(n)}(x,y) < p^{(n)}(x,y)$ even if $x, y \in A$ since we might have more paths connecting x and y by going outside the set A in n steps.

By the same token, one can define the Green function for P_A by

$$G_A(x,y|z) := \sum_{n=0}^{\infty} p_A^{(n)}(x,y) z^n.$$

Set $G_A(x,y) = G_A(x,y|1) = \sum_{n=0}^{\infty} p_A^{(n)}(x,y)$.

Remark. If $A \subset X$ is a finite set, then $G_A(x,y) < \infty$ for all $x, y \in X$

For a finite set $A \subset X$, the function $G_A(\cdot, x)$ is finitely supported since $G_A(x,y) = 0$ if and only if x and y do not both belong to A. Let $f \in \mathbb{R}^X$, $x \in X$ and set

$$G_A f(x) := \sum_{y \in X} G_A(x,y) f(y) = \sum_{y \in A} G_A(x,y) f(y).$$

Since $G_A(x,y) = 0$ whenever both x and y do not belong to A, we have $G_A f(x) = 0$ if $x \notin A$. Therefore, for $f \in \ell^2(X,m)$ we have $G_A f \in \ell^2(X,m)$:

$$\sum_{x \in X} m(x)(G_A f(x))^2 = \sum_{x \in A} \sum_{y \in A} G_A(x,y) f(y) < \infty$$

as both sums are over finite set A.

Define $I_A : \ell^2(X,m) \to \ell^2(X,m)$ by setting

$$I_A f(x) = f(x) \text{ if } x \in A \text{ and } 0 \text{ otherwise}$$

for all $f \in \ell^2(X,m)$.

Lemma 2.34. *Let $A \subset X$ be a finite set. Assume $x \in A$ and $f \in \ell^2(X,m)$ be such that $supp(f) \subset A$. Then*

$$\langle \nabla f, \nabla G_A(\cdot, x) \rangle_E = m(x) f(x).$$

Proof. Recalling that ∇^* is adjoint of the nabla operator ∇ and that $\nabla^* \nabla f(x) =$

$(I - P)f(x)$. Then

$$\begin{aligned}
\langle \nabla f, \nabla G_A(\cdot, x)\rangle_E &= \langle f, \nabla^* \nabla G_A(\cdot, x)\rangle_X \\
&= \langle f, (I - P)G_A(\cdot, x)\rangle_X \\
&= \langle f, (I_A - P_A)G_A(\cdot, x)\rangle \\
&= \langle f, \delta_x\rangle_X \\
&= \sum_{y \in A} m(y) f(y) \delta_x(y) \\
&= m(x) f(x).
\end{aligned}$$

□

Lemma 2.35. *Let $A \subset B$ be finite subsets of X. We have*

$$D(G_B(\cdot, x) - G_A(\cdot, x)) = m(x)(G_B(x, x) - G_A(x, x))$$

for all $x \in A$.

Proof. Applying Lemma 2.34, we get

$$\begin{aligned}
D(G_B(\cdot, x) - G_A(\cdot, x)) &= \langle \nabla(G_B(\cdot, x) - G_A(\cdot, x)), \nabla(G_B(\cdot, x) - G_A(\cdot, x))\rangle \\
&= \langle \nabla G_B(\cdot, x), \nabla G_B(\cdot, x)\rangle - 2\langle \nabla G_B(\cdot, x), \nabla G_A(\cdot, x)\rangle \\
&+ \langle \nabla G_A(\cdot, x), \nabla G_A(\cdot, x)\rangle \\
&= m(x)G_B(x, x) - 2m(x)G_B(x, x) + m(x)G(x, x) \\
&= m(x)(G_A(x, x) - G_B(x, x)).
\end{aligned}$$

□

Lemma 2.36. *Let (X, P) be a transient Markov chain. Then $G(\cdot, x) \in \mathcal{D}_0(\mathcal{N})$ for all $x \in X$.*

Proof. Let $(A_n)_n$ be an exhaustion of X, i.e., a sequence such that $A_n \subset A_{n+1}$ and $\bigcup_{n\in\mathbb{N}} A_n = X$. It is clear that $p^{(k)}_{A_n}(x, y) \to p^{(k)}(x, y)$ as $n \to \infty$ for all $x, y \in X$. Consequently, we have $G_{A_n}(x, y) \leq G(x, y)$. Since the Markov chain is transient, $G(x, y) < \infty$. Thus, $G_{A_n}(x, y) \to G(x, y)$ by the monotone convergence theorem. By the Lemma 2.35, the sequence $(G_{A_n}(\cdot, x))_n$ is a Cauchy sequence in $\mathcal{D}(\mathcal{N})$. The sequence $G_{A_n}(\cdot, x)$ converges to the pointwise limit $G(\cdot, x)$. As a consequence, $G(\cdot, x) \in \mathcal{D}_0(\mathcal{N})$. □

Now we are all set in order to present some transience criterions.

2.8 Criteria for Transience

Let (X, P) be an irreducible reversible Markov chain with the total conductance m and let $\mathcal{N} = (X, E, r)$ be the associated network.

Recall that we interpret the function $u \in \ell^2(E, r)$ as an energy flow and the quantity $m(x)\nabla^*(x)$ as the loss at $x \in X$.

Definition 2.37. For a vertex $x_0 \in X$ and $i_0 \in \mathbb{R}$, a *finite energy flow* from x with *input* i_0 is a function $u \in \ell^2(E, r)$ such that

$$\nabla^*(y) = -\frac{i_0}{m(x_0)}\delta_{x_0}(y).$$

Its energy is the positive quantity $\langle u, u\rangle$. Note that the energy can be infinite too.
We may think of the network $\mathcal{N}$ as the system of tubes filled with liquid such that i_0 litres of liquid being pumped in at a source x_0.

Definition 2.38. The quantity

$$cap(A) = min\{D(f) : f \in \mathcal{D}_0(\mathcal{N}), f \equiv 1 \text{ on } A\}$$

is called the *capacity* of the set $A \subset X$.

The following set of useful necessary and sufficient conditions for transience.

Theorem 2.39. *Let (X, P) be a reversible Markov chain with total conductance m and associated network $\mathcal{N}$. The following are equivalent:*

(1) The network is transient;

(2) For some $x \in X$, there is a finite energy flow from x with non-zero input i_0;

(3) For some $x \in X$, $cap(x) > 0$;

(4) The constant function 1 on X does not belong to $\mathcal{D}_0(\mathcal{N})$.

Proof. (1) $\Rightarrow$ (2). Set $u = -\dfrac{i_0}{m(x_0)}\nabla G(\cdot, x_0)$. By the Lemma 2.36, $G(\cdot, x) \in \mathcal{D}_0(\mathcal{N}) \subset \mathcal{D}(\mathcal{N})$ for every $x \in X$. Then by the definition of the space $\mathcal{D}(\mathcal{N})$, $\nabla G(\cdot, x_0) \in \ell^2(E, r)$. As a consequence, $u \in \ell^2(E, r)$. Let's verify that u is indeed a finite energy flow:

$$\begin{aligned}\nabla^* u &= \nabla^*(-\frac{i_0}{m(x_0)}\nabla G(\cdot, x_0)) = -\frac{i_0}{m(x_0)}\nabla^*\nabla G(\cdot, x_0) \\ &= -\frac{i_0}{m(x_0)}(I - P)G(\cdot, x_0) = -\frac{i_0}{m(x_0)}\delta_{x_0}\end{aligned}$$

where last equality followed from the equality $(I - P)G(\cdot, x_0) = \delta_{x_0}$ as calculated above (see page 35).

(2) $\Rightarrow$ (3). We may assume that $i_0 = -1$ for the sake of simplicity. Now let $f \in \ell_0(X)$ such that $f(x_0) = 1$. We have

$$\begin{aligned}\langle \nabla f, u\rangle = \langle f, \nabla^* u\rangle &= \left\langle f, \frac{1}{m(x_0)}\delta_{x_0}\right\rangle \\ &= \sum_y m(y) f(y)\frac{1}{m(y)}\delta_{x_0} \\ &= f(x_0) = 1.\end{aligned}$$

Then, Cauchy-Schwarz inequality implies

$$1 = \langle \nabla f, u\rangle_E^2 \leq \langle \nabla f, \nabla f\rangle_X \langle u, u\rangle_E = D(f)\langle u, u\rangle_E .$$

Therefore,

$$cap(x_0) \geq D(f) \geq \frac{1}{\langle u, u\rangle} > 0$$

as $\langle u, u\rangle < \infty$.

(3) $\iff$ (4). We can easily prove this by contrapositive.

Suppose 1_X belongs to $\mathcal{D}_0(\mathcal{N})$. Then clearly, $cap(x_0) = 0$.

Conversely, suppose $cap(x_0) = 0$ so that $D(f) = 0$. As we have mentioned already, the kernel of the Dirichlet semi-norm D is the set of constant functions. Since $f(x_0) = 1$ by assumption, $f \equiv 1$.

(3) $\Rightarrow$ (1). Let $A \subset X$ such that $x_0 \in A$ and set $f = \dfrac{G_A(\cdot, x_0)}{G_A(x, x)}$. It is obvious that $f \in \mathcal{D}_0(\mathcal{N})$ and $f(x_0) = 1$. Applying Lemma 2.34, we obtain

$$cap(x_0) \leq D(f) = \langle \nabla f, \nabla f \rangle = \frac{1}{G_A(x_0,x_0)^2} \langle \nabla G_A(\cdot, x_0), \nabla G_A(\cdot, x_0) \rangle$$
$$= \frac{1}{G_A(x_0,x_0)^2} G_A(x_0,x_0) m(x_0) = \frac{m(x_0)}{G_A(x_0,x_0)}.$$

Thus, $G_A(x_0,x_0) \leq \dfrac{m(x_0)}{cap(x_0)}$. Consider the exhaustion $(A_n)_{\{n\in\mathbb{N}\}}$ of X. Monotone convergence theorem yields that $G(x_0,x_0) = \lim_{n\to\infty} G_{A_n}(x_0,x_0) \leq \dfrac{m(x_0)}{cap(x_0)}$. Hence, the Markov chain is transient by the Theorem 2.12 □

Corollary 2.40. *Let $(X,P_1),(X,P_2)$ be two Markov chains on X and D_1, D_2 associated Dirichlet semi-norms, respectively. Suppose there exsits $\varepsilon > 0$ such that*

$$D_2(f) \geq \varepsilon D_1(f) \text{ for all } f \in \ell_0(X).$$

Then transience of (X,P_1) implies the transience of (X,P_2).

Proof. If (X,P_1) is transient, then $cap(x) > 0$ for every $x \in X$. Then

$$D_2(f) \geq \varepsilon \cdot cap(x) > 0.$$

Hence, the Markov chain is transient as $cap(x_0) > 0$ for (X,P_2). □

Corollary 2.41. *Suppose (G,P_1) and $G,P_2)$ are two irreducible random walks on a finitely generated group G. Assume that for all $g,h \in G$, there exists a $\delta > 0$ such that*

$$p_1(g,h) \geq \delta p_2(g,h).$$

Then, transience of (G,P_1) implies the transience of (G,P_2).

Proof. Recall that a random walk on a group is symmetric i.e. $p(g,h) = p(h,g)$, so that $m(\cdot) = 1$. Then,

$$D_1(f) = \sum_{e\in E} \frac{(f(e^+ - f(e^-))^2}{r(e)}$$
$$= \sum_{e\in E} (f(e^+ - f(e^-))^2 p_1(g,h) \geq \sum_{e\in E} (f(e^+) - f(e^-))^2 \delta p_2(g,h)$$
$$= \delta D_2(f).$$

Now the result follows from the previous corollary. □

A *subnetwork* $\mathcal{N}'$ is a connected subgraph together with the restriction of the resistance to the remaining edges.

Remark. Suppose that $\mathcal{N}'$ is transient. Then by the definition of transience, then the probability that random walk can hit the vertices infinitely many times is zero, i.e., it only passes through a vertex finitely times. Then it is clear that the network $\mathcal{N}$ is also transient. Therefore, a finite energy flow in $\mathcal{N}'$ is also a finite energy flow in $\mathcal{N}$.

Following is a very important example in a sense that we shall use it in the derivation of a Nash-Williams' criterion which is very useful.

Example 2.42. Let $\mathbb{N}$ denote the set $\mathbb{N} \cup \{0\}$ for the sake of simplicity. As a graph, it is a one-way infinite path were the edges are $e_n = [n-1, n], n \geq 1$. For a nearest neighbor random walk, one defines the probabilities $p(n-1, n) > 0$ for $n \geq 1$, and $p(n, n) \geq 0$ for $n \geq 0$. Define m as

$$m(n) = \frac{p(0,1) \cdots p(n-1,n)}{p(n,n-1) \cdots p(1,0)}.$$

It can be easily checked that m defines a total conductance. It can also be easily checked that

$$r(e_n) = \frac{p(n-1,n-2) \cdots p(1,0)}{p(0,1) \cdots p(n-1,n)}.$$

We may assume that $i_0 = 1$. Note that there exists only one edge $e = \{e^+, e^-\}$ such that $e^+ = 0$, namely, $[0, 1]$. Then the constant function 1 is indeed an energy flow. We must have

$$\sum_{\substack{e \in E \\ e^+ = 0}} u(e) - \sum_{\substack{e \in E \\ u^- = 0}} u(e) = -1.$$

There is only one edge for which the initial vertex is 0, namely, $[0, 1]$. However, there is no vertex with terminal vertex 0. Therefore, the constant function $u = -1$ is the flow from 0. It's energy is

$$\langle u, u \rangle = \sum_{e \in E} r(e) = \sum_{n=1}^{\infty} r(e_n).$$

By the Theorem 2.39, the Markov chain is transient if and only if

$$\sum_{n=1}^{\infty} r(e_n) < \infty$$

and recurrent if and only if

$$\sum_{n=1}^{\infty} r(e_n) = \infty.$$

Note that this example implies that the network is transient if we can find a one-way together with a finite energy flow.

Definition 2.43. Let (X, P) be a reversible Markov chain with an invariant measure m and associated network $\mathcal{N} = (X, E, r)$. Let $(X_i)_{i\in\mathbb{N})}$ be a finite partition of X by finite subsets. The network $\mathcal{N}' = (\mathbb{N}, E, r')$ defined by the means of conductance

$$a'(i,j) = \sum_{x\in X_i}\sum_{y\in X_j} a(x,y)\, if\ \ i \neq j\ \ and\ \ 0\ \text{ otherwise.}$$

is called the *shortened network*.

Recall that conductance uniquely determines the transition probabilities (see the remarks after Definition 2.26). Thus, corresponding Markov chain $(\mathbb{N}, P')$ has the transition probabilities

$$p'(i,j) = \frac{a'(i,j)}{m'_i} = \frac{a'(i,j)}{\sum_{k\in\mathbb{N}} a'(i,k)}$$

for all $i, j \in \mathbb{N}$.

Observe that if (X, P) is irreducible, so is $(\mathbb{N}, P')$. Furthermore, we have $p'(i,i) = 0$ as $a'(i,i) = 0$ and

$$\begin{aligned} m'(i) &= \sum_{j\in\mathbb{N}} a'(i,j) = \sum_{j\in\mathbb{N}}\sum_{x\in X_i}\sum_{y\in X_j} a(x,y) \\ &= \sum_{x\in X_i}\sum_{j\in\mathbb{N}}\sum_{y\in X_j} a(x,y) \\ &= \sum_{x\in X_j} m(x) < \infty \end{aligned}$$

for all $i \in \mathbb{N}$.

Theorem 2.44. *Suppose that (X, P) is a reversible Markov chain with associated shortened Markov chain $(\mathbb{N}, P')$. If $(\mathbb{N}, P')$ is recurrent, then so is (X, P).*

Proof. Let $f \in \mathbb{R}^{\mathbb{N}}$ be in $\mathcal{D}(\mathcal{N}')$. Define $\overline{f} : X \to \mathbb{R}$ by

$$\overline{f}(x) = f(i)\ \text{ if }\ x \in X_i.$$

Observe that

$$\overline{1_{\mathbb{N}}} = 1_X.$$

We have $D(\overline{f}) = D(f)$:

$$\begin{aligned} D(\overline{f}) &= \frac{1}{2}\sum_{x,y\in X}(\overline{f}(x)-\overline{f}(y))^2 a(x,y) \\ &= \frac{1}{2}\sum_{i,j\in\mathbb{N}}\sum_{\substack{x\in X_i \\ y\in X_j}}(\overline{f}(x)-\overline{f}(y))^2 a(x,y) \\ &= \frac{1}{2}\sum_{i,j\in\mathbb{N}}(f(i)-f(j))^2\sum_{\substack{x\in X_i \\ y\in X_j}} a(x,y) \\ &= \frac{1}{2}\sum_{i,j\in\mathbb{N}}(f(i)-f(j))^2 a'(i,j) \\ &= D'(f). \end{aligned}$$

As a consequence, $f \in \mathcal{D}(\mathcal{N})$ if and only if $\overline{f} \in \mathcal{D}(\mathcal{N}')$.
One has $\overline{f} = \sum_{i\in supp(f)} f(i)1_{X_i} \in \ell_0(X)$ as $1_{X_i} \in \ell_0(X)$. Suppose that $(\mathbb{N}, P')$ is recurrent. Then $1_{\mathbb{N}_i} \in \ell_0(\mathbb{N})$ by the Theorem 2.39. Thus, we can find a sequence $(f_n)_{n\in\mathbb{N}}$ in $\ell_0(\mathbb{N})$ satisfying $\lim_{n\to\infty} D'(f_n - 1_{\mathbb{N}}) = 0$. Now consider the sequence $(\overline{f_n})_{n\in\mathbb{N}}$ in $\ell_0(X)$. We have

$$D'(f_n - 1_{\mathbb{N}}) = D(\overline{f_n - 1}) = D(\overline{f_n} - \overline{1_{\mathbb{N}}}) = D(\overline{f_n} - 1_X).$$

As a consequence, using the fact that $D(\overline{f}) = D(f)$ we get

$$\lim_{n\to\infty} D'(f_n - 1_{\mathbb{N}}) = \lim_{n\to\infty} D(\overline{f_n} - 1_X) = 0.$$

Hence, $1_X \in \mathcal{D}(\mathcal{N})$. The Markov chain (X, P) is recurrent by the virtue of Theorem 2.39(4). □

The following criterion is due to Crispin Nash-Williams.

Theorem 2.45. *Suppose that (X, P) is a reversible Markov chain with associated shortened Markov chain $(\mathbb{N}, P')$ such that*

$$a'(n,m) \neq 0 \ \textit{if} \ |i-j| \leq 1.$$

If one has

$$\sum_{n=1}^{\infty} \frac{1}{a'(i-1,i)} = \infty$$

then (X, P) is recurrent.

2.9 Polya's walk

Consider the simple random walk on the grid $\mathbb{Z}^d$. The problem which elevated the study random walks on groups is determining the expected number of visits to the starting point. It is now well-known that for $d = 1, 2$ the expected number is infinite, and for $d \geq 3$ is finite. In other words, the simple random walk on $\mathbb{Z}^d$ is recurrent for $d = 1, 2$ and transient for $d \leq 3$. This problem can be tackled with combinatorial approaches; however, the result can be derived from the Nash-Williams' criterion with significantly less of a workload.

Corollary 2.46. *The simple random walk on Z is recurrent*

Proof. Let $(\mathbb{Z}, P)$ be a simple random walk on Z. A simple random walk is always reversible with $m(\cdot) = 1$ and hence $a(x, y) = p(x, y) = \frac{1}{2}$ if $x \sim y$, or equivalently, $|x - y| = 1$, and 0 otherwise. Set $X_n = \{n, -n\}$. Then $\mathbb{Z}$ is a disjoint union of the finite sets X_n. Now let's calculate the conductances in the corresponding shortened network $(\mathbb{N}, P')$. Notice that $a'(i, j) = 0$ if $|i - j| \geq 2$:

$$a'(i,j) = \sum_{\substack{x \in X_i \\ y \in X_j}} p(x,y) = p(-i,-j) + p(-i,j) + p(i,-j) + p(i,j) = 0$$

as all the probabilities are zero. On the other hand, for $|i - j| = 1$ we have

$$\begin{aligned} a'(i,j) &= \sum_{\substack{x \in X_i \\ y \in X_j}} p(x,y) = p(-i,-j) + p(-i,j) + p(i,-j) + p(i,j) \\ &= p(-i,-j) + p(i,j) = \frac{1}{2} + \frac{1}{2} = 1. \end{aligned}$$

Then

$$\sum_{i=1}^{\infty} \frac{1}{a'(i-1,i)} = \sum_{i=1}^{\infty} 1 = \infty.$$

Hence, the Markov chain is recurrent by Nash-Williams' criterion. □

Let us also show that the simple random walk on $\mathbb{Z}^2$.

Corollary 2.47. *The simple random walk on $\mathbb{Z}^2$ is recurrent.*

Proof. Our usual choice for m is the constant function 1. However, it can be chosen to be the degree of the vertices so that all the edges have a conductance 1. Indeed, if $m(x) = deg(x)$, then

$$m(x)p(x,y) = deg(x)\frac{1}{deg(x)} = deg(y)\frac{1}{deg(y)} = m(y)p(y,x).$$

The set X_i, on the other hand, is equivalent to the set

$$\{(m,n) \in \mathbb{Z}^2 : |m| + |n| = i\}$$

for all $i \geq 0$.

Cardinality of the set X_i is equal to $4i$ for all $i \geq 1$. Moreover, $a'(i,j) = 0$ for $|i-j| \geq 2$ again. For $|i-j| = 1$, let $j = i-1$. We obtain

$$a'(i-1,i) = \sum_{\substack{x \in X_{i-1} \\ y \in X_i}} a(x,y) = \sum_{\substack{x \in X_{i-1} \\ y \in X_i \\ x \sim y}} 1.$$

Therefore, we have to calculate how many of the pairs $\{x,y\}$ are edges. Observe that the each of the nodes $(0, i-1), (0, -(i-1)), (i-1, 0), \quad \text{and} \quad (-(i-1), 0)$ from X_{i-1} are adjacent to exactly 3 vertices, and the remaining $4(i-1) - 4$ vertices are adjacent to exactly 2 vertices from X_i. Therefore, we have that

$$a'(i-1,i) = 8i - 4.$$

Thus,

$$\sum_{i=1}^{\infty} \frac{1}{a(i-1,i)} = \sum_{i=1}^{\infty} \frac{1}{8i-4} = \infty$$

□

One expects $a'(i-1,i)$ to be quadratic for $d = 3$ so that the simple random walk on $\mathbb{Z}^3$ is transient.

2.10 Growth rate of Groups

Let G be a finitely generated group by a symmetric subset $S \subset G$. Recall from the Definition 2.21 that the *word length* of an element $g \in G$ with respect to the generating set $S \subset G$ is the minimal integer $n \in \mathbb{N}$ such that g can be expressed as

product of n elements of the generating set.
Also recall that we defined a left-invariant word metric $d_S : G \times G \to \mathbb{N}$ by

$$d_S(g, h) = \ell_S(g^{-1}h)$$

for all $g, h \in G$.

Definition 2.48. Let $g \in G$ and $n \in \mathbb{N}$. The set

$$B_S(g, n) := \{h \in G : d_S(g, h) \leq n\}$$

is called the *ball of radius* n in G with the center $g \in G$ with respect to S.

Note that $B_S(1_G, n) = \{h \in G : \ell_S(h) \leq n\}$. By the left-invariance of d_S, we have that $|B_S(1_G, n)| = |B_S(g, n)|$ for all $g \in G$. In other words, $B_S(\cdot, n)$ is the set of all elements $h \in G$ whose word length is at most n.

Definition 2.49. The function $\alpha_S(n) : \mathbb{N} \to \mathbb{N}$ defined by setting

$$\alpha_S(n) = |B_S(1_G, n)| = |\{h \in G : \ell_S(h) \leq n\}|$$

for $n \in \mathbb{N}$ is called the *growth function* of G with respect to S.

Thus, the growth function measures the number of elements whose word length is at most n. It is easy to see that $B_S(1_G, 0) = 1$, $B_S(1_G, 1) = |S|$ and we have

$$B_S(1_G, 0) \subset B_S(1_G, 1) \subset ....$$

Therefore, the growth function is increasing.

Example 2.50. (a) Let $G = \mathbb{Z}$ and $S = \{-1, 1\}$. For $n \in \mathbb{N}$, the set of elements whose word length is less than or equal to n is

$$\{-n, ..., -1, 0, 1, ...n\}.$$

Thus, $\alpha_S(n) = 2n + 1$.

(b) Let $G = \mathbb{Z}$ again with $S = \{\pm 2 \pm 3\}$. There is only one element, namely 0, whose length is 1. For the words of length 2, we have the following elements $\alpha_S(1) = |\{\pm 2, \pm 3\}| = 4$. For $n = 2$, the set of elements

$$\{\pm 1, \pm 4, \pm 5, \pm 6\}$$

are of length 2. For $n \geq 3$, the only elements at a distance n are

$$\{\pm(3n-2), \pm(3n-1), \pm 3n\}.$$

Thus, $\alpha_S(n) = 6n + 1$.

(c) If $G = \mathbb{Z}^2$, it can be similarly shown that $\alpha_S(n) = 2n^2 + 2n + 1$

(d) Let $G = F_2$, where F_2 denotes the free group on two generators $\{a, b\}$. Set $S = \{a^{-1}, b^{-1}, a, b\}$. Then for $n \neq 0$, the number of words of the length 4 is $4 \cdot 3^{n-1}$ as we have 4 choice for the first letter and 3 for each vertex. Then we have

$$\alpha_S(n) = 1 + \sum_{k=1}^{n} 4 \cdot 3^{k-1} = 2 \cdot 3^n - 1.$$

One observes that the growth function depends on the generating set, as can be seen from the parts (a) and (b) of the above example. However, we will show that it does not "depend" much on the generating set. Notice that even if the growth functions of $\mathbb{Z}$ are different, they both are linear polynomials.

Lemma 2.51. *Let G be a finitely generated group with distinct symmetric generating subsets S and S'. Then there are real numbers c_1 and c_2 such that*

$$c_1 d_{S_1}(g, h) \leq d_S(g, h) \leq c_2 d_{S_2}(g, h)$$

for all $g, h \in G$.

Proof. Set $c_1 = \max\{l_{S'}(s) : s \in S\}$. Let $g \in G$. Then $g = s_1 \cdots s_n$ for some $s_i \in S,\ i = 1, ..., n$ such that $n \in \mathbb{N}$ is minimal. In other words, $\ell_S(g) = n$. Then we have

$$\ell_{S'}(g) = \ell_{S'}(s_1 \cdots s_n) \leq \sum_{i=1}^{n} \ell_{S'}(s_i) \leq c_1 n = c_1 \ell_S(g).$$

On the other hand, there are elements $s'_1, ..., s'_m \in S'$ such that $g = s'_1 \cdots s'_m$ such that $m \in \mathbb{N}$ is minimal. Set $\frac{1}{c_2} \max\{\ell_S(s') : s' \in S'\}$. Therefore,

$$\ell_S(g) = \sum_{i=1}^{m} l_S(s'_i) = \frac{1}{c_2} m = \frac{1}{c_2} \ell_{S'}(g)$$

Then we have

$$c_2 d_S(g, h) \leq d_{S'}(g, h) \leq c_1 d_S(g, h)$$

for all $g, h \in G$. □

Now we shall investigate when two functions grow in a similar rate.

Definition 2.52. A function $f : \mathbb{N} \to [0, \infty)$ is called a growth function if it is non-decreasing.

Suppose that $f, g : \mathbb{N} \to [0, \infty)$ are two growth functions. The function g is said to dominate f if there exists an integer $c \geq 1$ such that

$$f(n) \leq g(cn)$$

for all $n \geq 1$. In this case, we write $f \preccurlyeq g$.

We say that f and g are *equivalent* if and only if $f \preccurlyeq g$ and $f \succcurlyeq g$, and write $f \sim g$.

Proposition 2.53. *(1) The relation $\preccurlyeq$ is reflexive and transitive.*

(2) $\sim$ is an equivalence relation.

(3) let $f, g, f', g' : \mathbb{N} \to [0, \infty)$ be growth functions. If $f \sim f'$, $g \sim g'$, and $f \preccurlyeq g$, then $f' \preccurlyeq g'$.

Proof. (1) It is obvious that $\preccurlyeq$ is reflexive as f is an increasing function. Moreover, suppose that $f \preccurlyeq$ and $g \preccurlyeq h$. Then there are $c, d \in \mathbb{Z}$ such that

$$f(n) \leq g(cn) \text{ and } g(n) \leq h(dn)$$

for all $n \in \mathbb{N}$. Then we have

$$f(n) \leq g(cn) \leq h((cd)n)$$

implying that $f \preccurlyeq h$.

(2) It is clear that the relation is symmetric by the definition of $\sim$. Reflexivity and transitivity immediately follows from the part (1).

(3) By the hypothesis, we have that $f' \preccurlyeq f \preccurlyeq g \preccurlyeq g'$. Since $\preccurlyeq$ is transitive, we obtain $f' \preccurlyeq g'$.

□

Roughly speaking, $f \sim g$ tells us that f and g are growing at a same rate.

Example 2.54. (1) Let p and q be two polynomials. Then

$$p \preccurlyeq q \Longleftrightarrow deg(p) \leq deg(q)$$

and

$$p \sim q \Longleftrightarrow deg(p) = deg(q)$$

Note that a polynomial f is equivalent to n^d for some $d \geq 0$.

(2) For $a, b > 1$, $a^n \sim b^n$.

We shall denote the exponential growth by 2^n.

Corollary 2.55. *Let G be a finitely generated groups, and $S, S' \subset G$ be two finite symmetric generating subsets. Then the growth functions corresponding to S and S' are equivalent, i.e., $b_S \sim b_{S'}$.*

Proof. We have shown that for any $g \in G$,

$$\ell_{S'}(g) \leq c\ell_S(g)$$

in the proof of the Lemma 2.51. Thus,

$$B_S(1_G, n) = \{g \in G : \ell_S(g) \leq n\} \subset \{g \in G : \ell_{S'}(g) \leq cn\} = B_{S'}(1_G, cn).$$

Therefore, $b_S(n) = |B_S(1_G, n)| \leq |B_{S'}(1_G, cn)| = b_{S'}(n)$. Interchanging the roles, we have $b_S \sim b_{S'}$. □

Based on the previous corollary, all the growth functions of a finitely generated group G are in the same equivalence classes which is called the *growth type* of G. We denote the growth type by b^G. One says that the group G has an *exponential growth* if $b^G \sim 2^n$. If $b^G \not\sim 2^n$, then we say G has *sub-exponential growth* type.

Proposition 2.56. *Let H be a finitely generated subgroup of a finitely generated group G. Then*

$$b^H \preccurlyeq b^G.$$

Proof. Let G and H be generated by finite symmetric sets S and S', respectively. Then it is obvious that $S \cup S'$ is also a symmetric finite generating set for G. Then

$$B_{S'}(1_H, n) \subset B_S(1_G, n).$$

which yields

$$b_{S'}(n) \leq b_S(n)$$

for all $n \in \mathbb{N}$. Thus, $b^H \preccurlyeq b^G$. □

Proposition 2.57. *Let G_1, G_2 be two finitely generated groups generated by finite symmetric subsets S_1 and S_2. Then $b^{G_1 \times G_2} \sim b^{G_2} b^{G_2}$.*

Corollary 2.58. *The group $\mathbb{Z}^d$ has the growth type n^d.*

Proof. We know, from Example 2.50, that $n^{\mathbb{Z}} \sim n$. By the previous proposition, we deduce that $b^{\mathbb{Z}^d} \sim n^d$ for $d \geq 1$. □

2.11 Growth and Random walks

As one might expect, there is a relationship between the growth and a simple random walk on a group. Intuitively, growth of a group tells us how fast the Cayley graph of the group grows. Roughly speaking, if the graph of the group grows "too fast" then the simple random walk on the group may be expected to be transient.

In order to prove the proposition showing the relation between simple random walks and growth, we shall need a simple inequality.

Lemma 2.59. *Let $x_1, x_2, ..., x_n \in (0, \infty)$, and $n \geq 1$. Then one has*

$$\sum_{i=1}^{n} \frac{1}{x_i} \geq \frac{n^2}{\sum_{i=1}^{n} x_i}$$

Proof. Recall that for a convex function f, numbers $x_1, ..., x_n$ with weights a_i, Jensen's inequality can be stated as

$$f\left(\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i}\right) \leq \frac{\sum_{i=1}^{n} a_i f(x_i)}{\sum_{i=1}^{n} a_i}.$$

If all a_i are the same, then the inequality reduces to

$$f\left(\frac{\sum_{i=1}^n x_i}{n}\right) \leq \frac{\sum_{i=1}^n f(x_i)}{n}.$$

Then we have

$$\frac{n}{\sum_{i=1}^n x_i} = f\left(\frac{\sum_{i=1}^n x_i}{n}\right) \leq \frac{\sum_{i=1}^n f(x_i)}{n} = \frac{\sum_{i=1}^n \frac{1}{x_i}}{n}.$$

Rearranging the terms, we obtain

$$\sum_{i=1}^n \frac{1}{x_i} \geq \frac{n^2}{\sum_{i=1}^n x_i}.$$

□

Proposition 2.60. *Suppose that G is a finitely generated group with growth type at most quadratic. Then the simple random walk on the group is recurrent.*

Proof. Let $S \subset G$ be a symmetric finite generating subset. Let $X_n := \{g \in G : \ell_S(g) = n\}$. We may see X_n as a sphere of radius n centered at $1_g \in G$. Observe that G is disjoint union of X_n. Moreover, X_n is finite since $|X_n| = b_S(n) - b_S(n-1) < \infty$. Let (G, P) be a simple random walk on G with associated shortened Markov chain $(\mathbb{N}, P')$ defined by the means of the conductance

$$\begin{aligned} a'(i, i+1) &= \sum_{\substack{x \in X_i \\ y \in X_{i+1}}} a(x, y) \\ &= \sum_{\substack{x \in X_i \\ y \in X_{i+1}}} p(x, y) \\ &\leq \sum_{\substack{x \in X_i \\ s \in S}} p(x, xs) \\ &= \frac{1}{|S|}|S||X_i| = |X_i| \\ &= b_S(i) - b_S(i-1) \end{aligned} \tag{2.10}$$

Since G has at most quadratic growth, then there exist $c > 0$ and $N \in \mathbb{N}$ such that $b_S(n) \leq cn^2$ for all $n \geq N$. Then we have

$$\sum_{i=n+1}^{2n} \frac{1}{a'(i, i+1)} \geq \frac{n^2}{\sum_{i=n+1}^{2n} a'(i, i+1)}$$

by the previous lemma and (2.10)

$$\sum_{i=n+1}^{2n} \frac{1}{a'(i,i+1)} \geq \frac{n^2}{b_S(2n) - b_S(n)} \geq \frac{n^2}{b_S(2n)} \geq \frac{1}{4c}.$$

which shows that

$$\sum_{i=1}^{\infty} \frac{1}{a'(i,i+1)} = \infty.$$

By the virtue of the Nash-Williams' criterion (Proposition 2.45), we deduce the random walk is recurrent. □

CHAPTER 3

AMENABILITY

In a paper published in 1924, Banach and Tarski proved one of the most counter-intuitive results that a unit ball in $\mathbb{R}^3$ can be decomposed into finitely many disjoint pieces and reassembled into two unit balls using only translations and rotations of $\mathbb{R}^3$. In 1929, John Von Neumann discovered that these paradoxical decompositions rely on the existence of groups that do not satisfy a certain property, namely, Amenability. We will follow [7], [6], and [10].

Perhaps it is a good time to define what it means to "decompose into finitely many pieces and then reassembling together" formally.

Definition 3.1. Let G be a group acting on a set X. Two subsets $A, B \subset X$ are said to be *G-equidecomposable* if there exists sets $A_1, ..., A_n \subset A$ and $B_1, ...B_n \subset B$ such that

(1) $A = \bigcup_{i=1}^{n} A_i$ and $B = \bigcup_{i=1}^{n} B_i$

(2) $A_i \cap A_j = \emptyset = B_i \cap B_j$ for all $i \neq j$

(3) $g_i[A_i] = B_i$ for all $1 \leq i \leq n$.

Now we shall state the definition which characterizes the Banach-Tarski paradox.

Definition 3.2. A non-empty set $E \subset X$ is said to be *G-paradoxical* if for some positive integers n and m, there exist pairwise disjoint sets $A_1, ..., A_n, B_1, ..., B_m \subset E$ and group elements $g_1, ..., g_n, h_1, ..., h_m$ such that

$$E = \sum_{i=1}^{n} g_i[A_i] = \sum_{j=1}^{m} h_j[B_j].$$

In other words, E is G-paradoxical if there exists disjoint subsets $A, B \subset E$ such that A and B are G-equidecomposable with E.

Without getting into the details, using the above written terminology, we can state Banach-Tarski paradox as follows.

Theorem 3.3 (Banach-Tarski paradox). *The unit ball in $\mathbb{R}^3$ is G-paradoxical, where G is the subgroup of the isometry group of $\mathbb{R}^3$ generated by translations and rotations.*

Now let us turn our interest towards groups. Recall that an action is said to be free if and the only if the identity element is only element fixing all other elements. We know that every group acts on itself freely on itself by left multiplication.

Definition 3.4. A group G is said to be paradoxical if G-paradoxical with respect to its action on itself by left multiplication.

Proposition 3.5. *Free group F_2 on two generators is paradoxical.*

Proof. Let a, b be generators of F_2. Denote the set of all reduced words starting with i by $X(i)$. Clearly, we have

$$F_2 = 1_{F_2} \cup X(a) \cup X(b) \cup X(a^{-1}) \cup X(b^{-1}).$$

Observe that $aX(a^{-1})$ contains all the words starting with b and b^{-1}. Similarly, the set $bX(b^{-1)})$ contains all the words starting with a and a^{-1}. Then, we have

$$X(a) \cup aX(a^{-1}) = F_2 = X(b) \cup bX(b^{-1})$$

□

There are many equivalent but seemingly unrelated definitions of amenability. The following is the original definition due to John von Neumann in 1929, in response to the Banach-Tarski paradox.

Definition 3.6. Let G be a countable group. Then G is said to be amenable if there exists a finitely generated additive G-invariant probability measure on the power set $\mathcal{P}(G)$, that is, a map $\mu : \mathcal{P}(G) \to [0, 1]$ such that

(1) $\mu(G) = 1$

(2) $\mu(A \cup B) = \mu(A) + \mu(B)$ for all disjoint sets $A, B \subset G$.

(3) $\mu(gA) = \mu(A)$ for all $g \in G$ and $A \subset G$.

Example 3.7. Every finite group G is amenable with finitely additive G-invariant probability measure

$$\mu(A) = \frac{|A|}{|G|}.$$

Now let us show that amenable group are not paradoxical.

Proposition 3.8. *Let G be countable group. If G is amenable, then G is not paradoxical.*

Proof. Let G be a countable amenable group. Then by definition, there exists a finitely additive G-invariant probability measure μ. Suppose that G is a paradoxical group, i.e,. there are disjoint sets $A_1, A_2, ..., A_n, B_1, ..., B_m \subset G$ such that

$$G = \bigcup_{i=1}^{n} g_i[A_i] = \bigcup_{j=1}^{m} h_j[B_j]$$

Then since μ is invariant and $\mu(G) = 1$, we have

$$\begin{aligned} 1 = \mu(G) &\geq \mu\left(\bigcup_{i=1}^{n} A_i\right) + \mu\left(\bigcup_{j=1}^{m} B_j\right) \\ &= \sum_{i=1}^{n} \mu(A_i) + \sum_{j=1}^{n} \mu(B_j) \\ &= \sum_{i=1}^{n} \mu(g_i[A_i]) + \sum_{j=1}^{m} \mu(h_j[B_j]) \\ &\geq \mu\left(\bigcup_{i=1}^{n} g_i[A_i]\right) + \mu\left(\bigcup_{j=1}^{m} h_j[B_j]\right) = 2 \end{aligned}$$

□

The converse of the proposition also holds and therefore shows that exactly non-amenable groups the paradoxical groups. One can check [6], Theorem 4.9.1 for the details.

Now it immediately follows that the free group F_2 is not amenable. It was conjectured by John von Neumann that every non-amenable group contains an isomorphic copy of F_2. The conjecture was disproven by Ol'shanskii. On the other hand, the conjecture is true for many important class of groups such as linear groups.

Consider the real vector space of all bounded functions $f : X \to \mathbb{R}$ which is denoted by $\ell^\infty(X)$. It is a Banach space when equipped with the norm

$$||f||_\infty = \sup_{x \in X} |f(x)|.$$

Definition 3.9. A mean on X is a linear map $m : \ell^\infty \to \mathbb{R}$ such that

(1) $m(1) = 1$,

(2) $x \geq 0 \Rightarrow m(x) \geq 0$

for all $x \in \ell^\infty(X)$

It can be proven that there is a bijection between the space of finitely additive G-invariant probability measures and G-invariant means.

Definition 3.10. A group G is called amenable if there exists a G-invariant mean on $\ell^\infty(G, \mathbb{R})$, i.e.,

$$m(g \cdot f) = m(f)$$

for all $g \in G$ and all $f \in \ell^\infty(G, \mathbb{R})$.

Remark. Due to the above definition, it now makes sense why such groups are called amenable. In 1949, Mahlon M. Day introduced the term amenable as a pun on *"mean"*.

In order to prove the Kesten's theorem (Theorem (4.1)), we shall use the following formulation of amenability. We shall follow the notation from [7], chapter 14.

Definition 3.11. Let G be a group, X and Ω two subsets of G. The *X-external boundary* of Ω is defined as

$$\partial_X(\Omega) := X\Omega - \Omega = \bigcup_{x \in X} (x\Omega - \Omega).$$

The group is G is said to be amenable if for every finite $X \subset G$ and for every $\varepsilon > 0$, there exists a finite subset $\Omega \subset G$ such that

$$|\partial_X(\Omega)| < \varepsilon|\Omega|.$$

Observe that we have

$$\Omega \cup X\Omega = X \cup (X\Omega - \Omega).$$

Thus, $|\partial_X(\Omega)| < \varepsilon|\Omega|$ is equivalent to

$$|\Omega \cup X\Omega| < (1+\varepsilon)|\Omega|.$$

Example 3.12. Every finite group is amenable. For every $X \subset G$ and $\varepsilon > 0$, we can take $\Omega = G$. Then clearly

$$0 = |\partial_X(G)| < \varepsilon|G|.$$

Example 3.13. The group $\mathbb{Z}$ is amenable. Let $X \subset \mathbb{Z}$ be any subset, and $\varepsilon > 0$ be any positive real number. Let $k = \max\{|x| : x \in X\}$ and set $\Omega = [-n, n]$ for some $n \in \mathbb{N}$ that will be chosen later. Then it is clear that $X\Omega \subset [-n-k, n+k]$. Then we have

$$|\partial_X(\Omega)| = |\Omega \cup X\Omega| - |\Omega| \leq 2n + 2k - 2n = 2k$$

Now choose $n \in \mathbb{N}$ such that $2k \leq \varepsilon(2n+1)$. We obtain

$$|\partial_X(\Omega)| \leq 2k < \varepsilon(2n+1) = \varepsilon|\Omega|.$$

Example 3.14. Every group with subexponential growth is amenable.

Proof. Let us assume that G is non-amenable for the sake of contradiction. Then there is a finite set $X \subset G$ and $\varepsilon > 0$ such that, for any

$$|\Omega \cup X\Omega| \geq (1+\varepsilon)|\Omega|$$

for every $\Omega \subset G$. Let H be the subgroup generated by X. Set $\Omega := \bigcup_{k=0}^{n-1} X^k$, $n \geq 1$. Notice that $X\Omega = \bigcup_{k=0}^{n} X^k$. Thus,

$$|\bigcup_{k=0}^{n} X^k| - |\bigcup_{k=0}^{n-1} X^k| \geq \varepsilon|\bigcup_{k=0}^{n-1} X^k|.$$

One observes that Ω is the set of elements of H at most length n which equals to $b_X^H(n)$. Therefore, we have

$$b_X^H(m) - b_X^H(m-1) \geq \varepsilon b_X^H(m-1).$$

which implies that b^H is an exponential growth. By the Proposition (2.56), we have that G is also a group of exponential growth, a contradiction $\square$

Definition 3.15. Let G be a group and X, Ω be two finite subsets of G. Define X-*internal boundary* of Ω to be the set

$$\delta_X(G) = \{g \in \Omega : \text{ there is an } x \in X \text{ such that } xg \notin \Omega\}.$$

For the sake of simplicity, we shall denote the $\{x\}$-external boundary by $\partial_x(\Omega)$ of Ω.

Proposition 3.16. *Let G be a group. The following conditions are equivalent:*

(1) G is amenable

(2) For every $\varepsilon > 0$ and finite subset X, there exists a finite subset Ω such that

$$|\partial_x(\Omega)| < \varepsilon|\Omega|$$

for every $x \in X$

(3) For every $\varepsilon > 0$ and finite subset X there exists a finite subset $\Omega \subset G$ such that

$$|\delta_X(\Omega)| < \varepsilon|\Omega|.$$

Proof. $(1) \Leftrightarrow (2)$
It is clear that $(1) \Rightarrow (2)$. For the converse, we first observe that

$$\partial_{X_1 \cup X_2}(\Omega) = (X_1 \cup X_2)\Omega - \Omega \subset ((X_1\Omega) - \Omega) \cup (X_2\Omega) - \Omega) = \partial_{X_1}(\Omega) \cup \partial_{X_2}(\Omega).$$

which implies that

$$|\partial_{X_1 \cup X_2}(\Omega)| \leq |\partial_{X_1}(\Omega)| + |\partial_{X_2}(\Omega)|.$$

By the above-observation, we have

$$\partial_X(\Omega) = \sum_{x \in X} \partial_x(\Omega) \leq |X|\varepsilon|\Omega|$$

from which we can deduce that $(2) \Rightarrow (1)$.

$(1) \Leftrightarrow (3)$
Let $g \in \delta_X(\Omega)$. By definition, of $delta_X(\Omega)$ there exists an $x \in X$ such that $xg \notin \Omega$. There are at most $|X|$ choice for x such that $xg \notin \Omega$. Therefore, we have

$$|\partial_X(\Omega)| \leq |X||\delta_X(\Omega)|$$

from which $(1) \Rightarrow (3)$ follows. On the other hand, for every $g \in$ there is at least one x such that $xg \notin \Omega$. In other words, there is at least one element in $\bigcup_{x \in X} \partial_x(\Omega)$. Then, we obtain

$$|\delta_X(\Omega)| \leq \sum_{x \in X} |\partial_x(\Omega)| \leq |X| \max_{x \in X} |\partial_x(\Omega)| \leq |X| \cdot |\partial_X(\Omega)|$$

from which we deduce that $(3) \Rightarrow (1)$.

The following lemma will be crucial for proving Kesten's Theorem 4.1

□

Lemma 3.17. *Let G be a group with finite generating set $S \subset G$. Then the following are equivalent:*

(i) G is amenable.

(ii) There exists a finite nonempty subset $\Omega \subset G$ such that

$$\frac{|\Omega - s\Omega|}{|\Omega|} < \varepsilon$$

for every $\varepsilon > 0$ and for all $s \in S$.

Proof. It is clear that (i) implies (ii) by definition of amenability. Now suppose that (ii) holds true. We can assume that $1_G \in S$. Let X be any finite subset of G. Then it is clear that there exists an $n \in \mathbb{N}$ such that $X \subset S^n$. In order words, there are generating elements $s_1, ..., s_n$ such that $g = s_1 \cdots s_n$. It is clear that one has $g(A - B) = g(A) - g(B)$ for all subsets $A, B \subset G$ and $g \in G$. Moreover, we have

$|A - B| = |A - C| + |C - B|$ for any sets $A, B, C \subset G$. Then

$$\begin{aligned}
|\Omega - s\Omega| &= |\Omega - s_1 \cdots s_n\Omega| \leq |\Omega - s_1\Omega| + |s_1\Omega - s_1 \cdots s_n\Omega| \\
&\leq |\Omega - s_1\Omega| + |s_1\Omega - s_1 s_2\Omega| + |s_1 s_2\Omega - s_1 s_2 \cdots s_n\Omega| \\
&\leq |\Omega - s_1\Omega| + |\Omega - s_2\Omega| + |s_1 s_2\Omega - s_1 s_2 s_3\Omega| + |s_1 s_2 s_3\Omega - s_1 \cdots s_n\Omega| \\
&\leq |\Omega - s_1\Omega| + |\Omega - s_2\Omega| + ... + |\Omega - s_n\Omega| \\
&\leq n\varepsilon|\Omega| = \varepsilon^{'}|\Omega|
\end{aligned}$$

where $\varepsilon^{'} = n\varepsilon$. This completes the proof. □

CHAPTER 4

KESTEN'S AMENABILITY CRITERION FOR FINITELY GENERATED GROUPS

In his seminal paper [8], Harry Kesten introduced a probabilistic criterion for Amenability. We know that Cayley graphs carry out geometric information about the groups and recall that we can define simple random walks on groups through their Cayley graphs. The quantity $\sqrt[2n]{p^{(2n)}(e,e)}$ measures how fast the decaying sequence of returning probabilities $p^{(2n)}(e,e)$ decreases as $n \to \infty$. Kesten showed that a finitely generated group is amenable if and only if the quantity $\rho(G,S) = \rho(Cay(G,S)) = \lim_{n\to\infty} \sqrt[n]{p^{(2n)}(e,e)} = 1$. Roughly speaking, this shows that the Cayley graph of an amenable group is not too "wild", that is, the probability of return doesn't decrease steeply. Thus, in this sense, amenable groups are the "nicer" infinite groups. The main sources we will follow in this chapter are [8], [7], and [2].

Let G be a group with a finite symmetric generating set $S \subset G$. Let (G,P) be a simple random walk on G. Recall that for all $x, y \in X$, $z \in \mathbb{R}$, we defined the *Green function*

$$G(x,y|z) = \sum_{n=0}^{\infty} p^{(n)}(x,y)z^n.$$

We called the quantity

$$\rho(G,S) := \limsup_{n\to\infty} \sqrt[n]{p^{(n)}(e,e)} = \frac{1}{R}$$

spectral radius, where R is the radius of convergence for $G(x,y|z)$ (recall that if (G,P) is equipped with irreducibility, then the Green function has the same radius of convergence for all $x, y \in X$, $z \in \mathbb{R}$).

Theorem 4.1 (**Kesten's amenability criterion for finitely generated groups**). *A finitely generated group G is amenable if and only if $\rho(G,S) = 1$.*

In order to prove the theorem, we will employ operator theoretical concepts and results. We refer to [4], [7], [1] for the rest of the chapter.

Let $T : \ell^2(G) \to \ell^2(G)$ be a linear operator. Recall that its *norm* is the quantity

$$||T|| = \sup_{\substack{f \in \ell^2(G) \\ ||f|| \leq 1}} ||Tf||_2 = \sup_{\substack{f \in \ell^2(G) \\ f \neq 0}} \frac{||Tf||_2}{||f||_2}.$$

It is a standard fact that a linear operator is continuous if and only if $||T|| < \infty$. Let us denote the space of all continuous linear maps by $\mathcal{C}(\ell^2(G))$. We shall denote, as usual, the identity operator by $I : \ell^2(G) \to \ell^2(G)$.

An operator $T^* \in \mathcal{C}(\ell^2(G))$ satisfying

$$\langle Tf_1, f_2 \rangle = \langle f_1, T^* f_2 \rangle$$

for all $f_1, f_2 \in \ell^2(G)$ is called *adjoint* of T. Note that this operator exists and its unique by Riesz' representation theorem (see [1], Theorem 11.9, page 85). Moreover, $||T|| = ||T^*||$. One can show that the adjoint is involution, i.e., $(A^*)^* = A$ Indeed, we have

$$\langle Tf_1, f_2 \rangle = \langle f_1, T^* f_2 \rangle = \langle T^* f_2, f_1 \rangle = \langle f_2, (T^*)^* f_1 \rangle$$

which implies that

$$\langle f_2, Tf_1 \rangle = \langle f_2, (T^*)^* f_1 \rangle .$$

By uniqueness of the adjoint, we deduce that $(T^*)^* = T$. We say that T is *self-adjoint* if $T = T^*$.

Let $T \in \mathcal{C}(\ell^2(G))$ be a self-adjoint linear operator. Then the set

$$\begin{aligned} \sigma(T) : &= \{\lambda \in \mathbb{R} : (T - \lambda I) \text{ is not bijective}\} \\ &= \{\lambda \in \mathbb{R} : (T - \lambda I) \text{ has no bounded inverse}\} \end{aligned}$$

is called the spectrum of T.

We call the quantity

$$\rho(T) := \sup_{\lambda \in \sigma(T)} |\lambda| = \max_{\lambda \in \sigma(T)} |\lambda|$$

the *spectral radius* of T. Last equality follows from the fact that the spectrum is a compact set (see [4, p. 22]) such that

$$\sigma(T) \subset [-||T||, ||T||].$$

Moreover, we have

$$\rho(T) = ||T|| = \lim_{n\to\infty} ||T^n||^{1/n}.$$

One can see [4, p. 22] for details of the fact above.

Let $g \in G$. Define a linear operator $T_g : \ell^2(G) \to \ell^2(G)$ by setting

$$T_g f(h) = f(hg)$$

for all $h \in G$ and $f \in \ell^2(G)$.

Proposition 4.2. *The linear operator T_g is an isometry, that is, $||T_g|| = 1$.*

Proof. We have

$$\begin{aligned} ||T_g f||^2 = \langle T_g f, T_g f\rangle &= \sum_{h\in G} T_g f(h)^2 \\ &= \sum_{h\in G} f(gh)^2 = \sum_{k\in G} f(k)^2 \\ &= ||f||_2^2 \end{aligned}$$

for all $f \in \ell^2(G)$. □

As a consequence, T_g is bounded. Hence, $T_g \in \mathcal{C}(\ell^2(G))$.

Proposition 4.3. *The linear operator T_g has an inverse $(T_g)^* = T_{g^{-1}}$.*

Proof. It is clear that $T_{g^{-1}}$ is the inverse of T_g. On the other hand, letting $k = gh$, we have

$$\langle T_g f_1, f_2\rangle = \sum_{h\in G} f_1(gh) f_2(h) = \sum_{k\in G} f_1(k) f_2(g^{-1}k) = \langle f_1, T_{g^{-1}} f_2\rangle$$

for all $f_1, f_2 \in \ell^2(G)$. By the uniqueness of the adjoint, we obtain $(T_g)^* = T_{g^{-1}}$. □

Now suppose G is a finitely generated group, and S is a finite symmetric generating subset. We define the linear operator $M_S : \ell^2(G) \to \ell^2(G)$ by

$$M_S = \frac{1}{|S|} \sum_{s\in S} T_s. \tag{M}$$

for all $f \in \ell^2(G)$ and $g \in G$. The operator M_S is called the *Markov operator* with respect to S.

Proposition 4.4. *M_S is a self-adjoint linear operator with*

$$||M_S|| \leq 1.$$

Proof. M_S is linear and continuous as $T_s \in \mathcal{C}(\ell^2(G)$. We have

$$||M_S||^2 = ||\sum_{s\in S} \frac{1}{|S|} T_s|| \leq \frac{1}{|S|} \sum_{s\in S} ||T_s|| \leq 1.$$

Thus, $||M_S|| \leq 1$. On the other hand,

$$(M_S)^* = \left(\frac{1}{|S|}\sum_{s\in S} T_s\right)^* = \frac{1}{|S|}\sum_{s\in S} T_s^* = \frac{1}{|S|}\sum_{s\in S} T_{s^{-1}} = \frac{1}{|S|}\sum_{s\in S} T_s = M_S$$

where we used the fact that generating set is symmetric. □

Proposition 4.5. *Let M_S be the Markov operator. We claim that*

$$(M_S^n \delta_g)(h) = p^{(n)}(g, h).$$

for all $g, h \in G$ and $n \geq 0$.

Proof. For $n = 0$,

$$(M_S^0 \delta_g)(h) = \delta_g(h) = p^{(0)}(g, h).$$

Furthermore,

$$\begin{aligned}(M_S^{n+1}\delta_g)(h) &= (M_S(M_S^n\delta_g)(h)) \\ &= \frac{1}{|S|}\sum_{s\in S} M_S^n\delta_g(hs) = \frac{1}{|S|}\sum_{s\in S} p^{(n)}(g, hs).\end{aligned}$$

Let $hs = k$. Then,

$$p^{(n)}(g, hs) = \sum_{k\in G} p^{(n)}(g, k)\delta_g(h^{-1}k) = \sum_{k\in G} p^{(n)}(g, k)p(k, h) = p^{(n+1)}(g, h).$$

Thus, we have obtained

$$(M_S^{n+1}\delta_g)(h) = \frac{1}{|S|}\sum_{s\in S} p^{(n+1)}(g, h).$$

□

As a consequence, we obtain

Corollary 4.6. *Let $g, h \in G$ and $n \geq 0$. Then, we have*

$$\langle M_S^n \delta_g, \delta_h \rangle = p^{(n)}(g, h).$$

We refer to [7, p. 349] for the following proposition.

Proposition 4.7. *One has*

$$\rho(G, S) = ||M_S|| = \rho(M_S).$$

Proof of Theorem 4.1. ($\Rightarrow$) Let G be a finitely generated amenable group. We shall show that $\rho(G, S) = 1$.
Let us make the following crucial observation. We have shown that $||M_S|| \leq 1$. Recall that $\sigma(M_S)$ is a compact subset such that

$$\sigma(M_S) \subset [-||M_S||, ||M_S||].$$

Then it follows that showing $\rho(G, S) = 1$ is equivalent to proving that $1 \in \sigma(M_S)$. Indeed, if $1 \in \sigma(M_S)$, then $||M_S|| = 1$ by above remarks, and hence we would have $\rho(G, S) = ||M_S|| = 1$. By the Proposition (3.16), there is a finite subset $\Omega \subset G$ satisfying

$$|\partial_s(\Omega)| < \frac{\varepsilon^2}{2}|\Omega|$$

for all $s \in S$. Define the map $f : \ell^2(G) \to \mathbb{R}$ as $f := \dfrac{1}{\sqrt{|\Omega|}} \mathbb{1}_{\Omega^{-1}}$.

Lemma 4.8. *We have*

$$||f - T_s f||_2 < \varepsilon.$$

for every $s \in S$.

Proof. For any subsets $A, B \subset G$ we have $||\mathbb{1}_A - \mathbb{1}_B||_2^2 = |A - B| + |B - A|$. Indeed,

$$\begin{aligned} ||\mathbb{1}_A - \mathbb{1}_B||_2^2 = \sum_{g \in G} (\mathbb{1}_A - \mathbb{1}_B)^2 &= \sum_{\substack{g \in G: \\ \mathbb{1}_A - \mathbb{1}_B = 1}} (\mathbb{1}_A - \mathbb{1}_B)^2 + \sum_{\substack{g \in G: \\ \mathbb{1}_A - \mathbb{1}_B = -1}} (\mathbb{1}_A - \mathbb{1}_B)^2 \\ &= |A - B| + |B - A|. \end{aligned}$$

Moreover, note that $T_s \mathbb{1}_{\Omega^{-1}}(h) = \mathbb{1}_{\Omega^{-1}}(hs) = 1$ if $hs = \omega^{-1}$ for some $\omega \in \Omega$ and 0 otherwise. Then $hs \in \Omega^{-1} \Leftrightarrow h \in \Omega^{-1} s^{-1}$. Thus, $T_s \mathbb{1}_{\Omega^{-1}} = \mathbb{1}_{\Omega^{-1} s^{-1}}$.

Consequently,

$$\begin{aligned}||f - T_s f||_2^2 &= \sum_{h\in G}(f - T_s f)^2 = \frac{1}{|\Omega|}\sum_{h\in G}(\mathbb{1}_{\Omega^{-1}} - T_s\mathbb{1}_{\Omega^{-1}})^2 \\ &= \frac{1}{|\Omega|}\sum_{h\in G}(\mathbb{1}_{\Omega^{-1}} - \mathbb{1}_{\Omega^{-1}s^{-1}})^2 \\ &= \frac{1}{|\Omega|}\left(|\Omega^{-1} - \Omega^{-1}s^{-1}| + |\Omega^{-1}s^{-1} - \Omega^{-1}|\right) \\ &= \frac{1}{|\Omega|}\left(|\Omega - \Omega s| + |\Omega s - \Omega|\right) \\ &= 2\frac{\partial_s(\Omega)}{|\Omega|} < \varepsilon.\end{aligned}$$

□

As a consequence to the lemma above, we have

$$\begin{aligned}||(I - M_S)f||_2 &= \left\| f - \frac{1}{|S|}\sum_{s\in S} T_s f \right\|_2 \\ &= \frac{1}{|S|}\left\| f|S| - \sum_{s\in S} T_s f \right\|_2 \\ &= \frac{1}{|S|}\left\| \sum_{s\in S}(f - T_s f) \right\|_2 \\ &\leq \frac{1}{|S|}\sum_{s\in S}\|f - T_s f\|_2 \\ &< \frac{1}{|S|}\sum_{s\in S}\varepsilon = \varepsilon.\end{aligned}$$

Now suppose that $I - M_S$ has an inverse T. Then,

$$1 = \|T(I - M_s)f\|_2 \leq \|T\|_2 \|(I - M_S)f\|_2 < \|T\|_2 \varepsilon.$$

for all f such that $\|f\|_2 = 1$. Thus, we have $\|T\| > \frac{1}{\varepsilon}$ for every $\varepsilon > 0$. Hence, there is no bounded inverse of $I - M_S$ i.e. $1 \in \sigma(M_S)$. □

Let $f \in \mathbb{R}^G$. Recall that its support is the set

$$supp(f) := \{g \in G : f(g) \neq 0\}.$$

The vector subspace of all finitely supported functions $f : G \to \mathbb{R}$ is denoted by $\mathbb{R}[G]$.

It is clear that the subspace $\mathbb{R}^G$ is dense in $\ell^2(G)$.

Since $\mathbb{R}^G$ is dense in $\ell^2(G)$, we have that

$$\|T\| = \sup_{\substack{f\in\mathbb{R}[G]\\ \|f\|_2\leq 1}} \|Tf\|_2 = \sup_{\substack{f\in\mathbb{R}[G]\\ f\neq 0}} \frac{\|Tf\|_2}{\|f\|_2}$$

for all $T \in \mathcal{B}(\ell^2(G))$. We set

$$\|f\|_1 := \sum_{g\in G} |f(g)|.$$

for all $f \in \mathbb{R}[G]$. We have $supp(T_g f) = supp(f)g^{-1}$ for all $f \in \mathbb{R}[G]$. Indeed,

$$\begin{aligned} \textit{supp}(T_g f) &= \{h \in G : f(hg) \neq 0\} \\ &= \{kg^{-1} \in G : f(k) \neq 0\} \\ &= \{k \in G : f(k) \neq 0\}g^{-1} \\ &= \textit{supp}(f)g^{-1} \end{aligned}$$

where $k = hg$. As a consequence, $T_g f, M_S f \in \mathbb{R}[G]$ for all $f \in \mathbb{R}[G]$. Moreover, $\|Tf\|_1 = \|f\|_1$ for all $f \in \mathbb{R}[G]$ since

$$\sum_{h\in G} |T_g f(h)| = \sum_{h\in G} |f(hg)| = \sum_{h\in G} |f(h)|.$$

Lemma 4.9. *Let $f \in \mathbb{R}[G]$ be a non-negative function such that $\|f\|_1 = 1$. Then there exits $n \in \mathbb{N}$, nonempty subsets $A_i \subset G$ and real numbers $c_i > 0$, $i = 1, 2, ..., n$ satisfying $A_n \subset ... \subset A_1$ and $c_1 + ... + c_n = 1$ such that*

$$f = \sum_{i=1}^{n} c_i \frac{\mathbb{1}_{A_i}}{|A_i|}.$$

Proof. Let $0 < y_1 < ... < y_n$ denote finite number of values attained by f. Let us set

$$A_i = \{g \in G : f(g) \geq y_i\}$$

for every $i = 1, ..., n$. It is clear that $A_1, ...A_n$ is a decreasing sequence of subsets of G. Moreover, we have

$$f = y_1\mathbb{1}_{A_1} + (y_2 - y_1)\mathbb{1}_{A_2} + ... + (y_n - y_{n-1})\mathbb{1}_{A_n}$$

Now for each $2 \leq i \leq n$, set $c_i = (y_i - y_{i-1})|A_i|$ and let $c_1 = y_1|A_1|$. It is clear that c_i are positive for all $1 \leq i \leq n$ as desired. Furthermore, we have

$$\begin{aligned}\sum_{i=1}^{n} c_i &= y_1(|A_1| - |A_2|) + y_2(|A_2| - |A_3|) + ... + y_n|A_n| \\ &= y_1|A_1 - A_2| + y_2|A_2 - A_3| + ... + y_{n-1}|A_{n-1} - A_n| + y_n|A_n| \\ &= \sum_{g \in G} f(g) = \|f\|_1 = 1.\end{aligned}$$

□

The lemma above characterizes finitely supported non-negative functions.

Lemma 4.10. *For every $\varepsilon > 0$, there exists a non-negative $f \in \mathbb{R}[G]$ such that $\|f\|_2 = 1$ and*

$$\|M_S f\|_2 \geq 1 - \varepsilon.$$

Proof. By the Proposition 4.7, we have $\|M_S\|_2 = 1$. Recall that

$$\|M_S\|_2 = \sup_{\substack{g \in \mathbb{R}[G] \\ \|g\|_2 = 1}} \|M_S g\|_2 .$$

Since 1 is the supremum, there exists $g \in \mathbb{R}[G]$ such that

$$\|M_S g\|_2 \geq 1 - \varepsilon.$$

Notice that, however, we claimed that there is a non-negative function $f \in \mathbb{R}[G]$ of the norm 1 which satisfies the above inequality. To achieve that, we shall show that the modulus of g also satisfies the inequality. Let $f = |g|$. Then,

$$\begin{aligned}\|M_S g\|_2^2 &= \left\| \frac{1}{|S|} \sum_{s \in S} T_s g \right\|_2^2 \\ &= \sum_{h \in G} \left(\frac{1}{|S|} \sum_{s \in S} g(hs) \right)^2 \\ &\leq \sum_{h \in G} \left(\frac{1}{|S|} \sum_{s \in S} |g(hs)| \right)^2 \\ &= \|M_S|g|\|_2^2 = \|M_S f\|_2^2\end{aligned}$$

□

In order to prove a crucial lemma, we need the following result for Hilbert spaces.

Lemma 4.11. *Let $\mathcal{H}$ be a Hilbert space. Let $u, v \in \mathcal{H}$ such that $\|u\| \leq 1$, $\|v\| \leq 1$. For every $0 < \varepsilon \leq 2$ such that $\|u - v\| \geq \varepsilon$, there exists a $\delta > 0$ such that*

$$\left\| \frac{u+v}{2} \right\| \leq 1 - \delta.$$

Proof. Using the parallelogram law, we have

$$\begin{aligned} \|u+v\|^2 &= \|u+v\|^2 + \|u-v\|^2 - \|u-v\|^2 \\ &= 2\|u\|^2 + 2\|v\|^2 - \|u-v\|^2 \\ &\leq 4 - \varepsilon^2 \end{aligned}$$

Now put $\delta = 1 - \frac{1}{2}\sqrt{4-\varepsilon^2}$. Then we obtain the desired result. □

Geometrically, this means that whenever the vectors are far enough from each other, the midpoint of the line segment lies inside the unit ball.

Lemma 4.12. *For every $\varepsilon > 0$ there is a non-negative function $f_\varepsilon \in \mathbb{R}[G]$ such that $\|f_\varepsilon\|_2 = 1$ and*

$$\|f_\varepsilon - T_s f_\varepsilon\|_2 < \varepsilon$$

for all $s \in S$.

Proof. Suppose the converse is true, i.e., there exists $\varepsilon_0 > 0$ such that for all $f \in \mathbb{R}[G]$ with $\|f\|_2 = 1$ there exists $s \in S$ such that

$$\|f_{\varepsilon_0} - T_s f_{\varepsilon_0}\| \geq \varepsilon_0.$$

Since $\ell^2(G)$ is a Hilbert space, the above lemma applies. In other words, there exists $\delta > 0$ such that

$$\left\| \frac{f + T_{s_0} f}{2} \right\|_2 \leq 1 - \delta$$

for all $f \in \mathbb{R}[G]$.

Then we have

$$\begin{aligned}
\|M_S f\|_2 &= \left\| \frac{1}{|S|} \sum_{s \in S} T_s f \right\|_2 \\
&= \left\| \frac{1}{|S|}(f + T_{s_0} f) + \frac{1}{|S|} \sum_{s \in S - \{1_G, s_0\}} T_s f \right\|_2 \\
&\leq \frac{2}{|S|} \left\| \frac{f + T_{s_0} f}{2} \right\|_2 \frac{1}{|S|} + \sum_{s \in S - \{1_G, s_0\}} \|T_s f\|_2 \\
&= \frac{2}{|S|}(1 - \delta) + \frac{|S| - 2}{|S|} = \frac{|S| - 2\delta}{|S|} = 1 - \frac{2\delta}{|S|}
\end{aligned}$$

which is a contradiction to the Lemma 4.10. □

Now, let $F := f_{\frac{\varepsilon}{2|S|}}$ as in Lemma 4.12. Therefore, $F \geq 0$, $\|F\|_2 = 1$, and $\|F - T_s F\|_2 \leq \frac{\varepsilon}{2|S|}$ for all $s \in S$. Now let $f = F^2 \in \mathbb{R}[G]$. Then $f \geq 0$ and $\|f\|_1 = 1$ since

$$\|f\|_1 = \sum_{g \in G} F^2(g) = \|F\|_2 .$$

Proposition 4.13. *Let g be a function as defined above. Then*

$$\|f - T_s f\|_1 \leq \frac{\varepsilon}{|S|}$$

for all $s \in S$.

Proof. Firstly, note that $T_s F^2 = (T_s F)^2$ for all $s \in S$. Now the proposition follows from direct computation of left hand side and Cauchy-Schwarz inequality:

$$\begin{aligned}
\|f - T_s f\|_1 &= \left\|F^2 - T_s F^2\right\|_1 = \left\|F^2 - (T_s F)^2\right\|_1 \\
&= \|(F - T_s F)(F + T_s F)\|_1 \\
&= \sum_{g \in G} |F(g) - T_s F(g)| \cdot |F(g) + T_s F(g)| \\
&= \langle F - T_s F, F + T_s F \rangle \\
&\leq \|F - T_s F\|_2 \cdot \|F + T_s F\|_2 \quad \text{(Cauchy-Schwarz)} \\
&\leq \|F - T_s F\|_2 \cdot (\|F\|_2 + \|T_s F\|_2) \\
&= 2\,\|F - T_s F\|_2 \\
&< \frac{\varepsilon}{|S|}
\end{aligned}$$

□

We can sense that the statement of the next Lemma 4.14 has the same flavor with the definition of Amenability. What we do, in essence, is expressing the "shifts" in subsets of a group in the operator theoretic setting.

Lemma 4.14. *Let $f \in \mathbb{R}[G]$ such that $f \geq 0$ and $\|f\|_1 = 1$. Then*

$$\|f - T_g f\|_1 = 2\sum_{i=1}^{n} c_i \frac{|A_i - A_i g|}{|A_i|}$$

for all $g \in G$.

Proof. We have shown that $T_g \mathbb{1}_A = \mathbb{1}_{Ag^{-1}}$ for all $g \in G$ and $A \subset G$. Then

$$\begin{aligned} f - T_g f &= \sum_{i=1}^{n} c_i \frac{\mathbb{1}_{A_i} - T_g \mathbb{1}_{A_i}}{|A_i|} \\ &= \sum_{i=1}^{n} c_i \frac{\mathbb{1}_{A_i} - \mathbb{1}_{A_i g^{-1}}}{|A_i|}. \end{aligned}$$

Now let

$$B = \bigcup_{1 \leq i \leq n} (A_i - A_i g^{-1}) \quad \text{and} \quad C = \bigcup_{1 \leq i \leq n} (A_i g^{-1} - A_i).$$

Then we have

$$\begin{aligned} \|f - T_g f\|_1 &= \sum_{g \in G} |(f - T_g f)(g)| \\ &= \sum_{g \in G} \left| \sum_{i=1}^{n} \frac{(\mathbb{1}_{A_i} - \mathbb{1}_{A_i g^{-1}})(h)}{|A_i|} \right| \\ &= \sum_{g \in B} \left| \sum_{i=1}^{n} \frac{(\mathbb{1}_{A_i} - \mathbb{1}_{A_i g^{-1}})(h)}{|A_i|} \right| + \sum_{g \in C} \left| \sum_{i=1}^{n} \frac{(\mathbb{1}_{A_i} - \mathbb{1}_{A_i g^{-1}})(h)}{|A_i|} \right| \\ &= \sum_{i=1}^{n} c_i \frac{|A_i - A_i g^{-1}|}{|A_i|} + \sum_{i=1}^{n} c_i \frac{|A_i g^{-1} - A_i|}{|A_i|} \\ &= \sum_{i=1}^{n} \frac{|A_i g - A_i|}{|A_i|} + \sum_{i=1}^{n} \frac{|A_i - A_i g|}{|A_i|} \\ &= 2\sum_{i=1}^{n} c_i \frac{|A_i - A_i g|}{|A_i|}. \end{aligned}$$

□

Recall that one has

$$f = \sum_{i=i}^{n} c_i \frac{\mathbb{1}_{A_i}}{|A_i|}$$

for c_i and A_i as in the Lemma 4.9.

Let $\mathcal{D} = \{1, 2, ..., n\}$ and consider the probability measure μ on $\mathcal{D}$ defined by $\mu(\{i\}) = c_i$ for every $i \in \mathcal{D}$. Moreover, for each $g \in G$, define a subset $\mathcal{D}_g$ of $\mathcal{D}$ by setting

$$\mathcal{D}_g = \left\{ i \in \mathcal{D} : \frac{|A_i - A_i g|}{|A_i|} \geq \varepsilon \right\}.$$

By the previous Lemma, we have

$$\begin{aligned} \|f - T_g f\|_1 &= 2 \sum_{i \in \mathcal{D}} c_i \frac{|A_i - A_i g|}{|A_i|} \\ &\geq 2 \sum_{i \in \mathcal{D}_g} c_i \frac{|A_i - A_i g|}{|A_i|} \\ &\geq 2\varepsilon \sum_{i \in \mathcal{D}_g} c_i \\ &= 2\varepsilon \mu(\mathcal{D}_g). \end{aligned}$$

Therefore, we have the inequality

$$\mu(\mathcal{D}_g) \leq \frac{\|f - T_g f\|_1}{2\varepsilon}$$

for all $g \in G$.

Using the Proposition 4.13, we obtain

$$\mu(\mathcal{D}_s) \leq \frac{\|f - T_s f\|_1}{2\varepsilon} \leq \frac{1}{2\varepsilon} \frac{\varepsilon}{|S|} = \frac{1}{2|S|} < \frac{1}{|S|}.$$

for all $s \in S$. As a consequence, we have

$$\mu\left(\cup_{s \in S} \mathcal{D}_s\right) \leq \sum_{s \in S} \mu(\mathcal{D}_s) < 1.$$

Therefore,

$$\cup_{s \in S} \mathcal{D}_s \neq \mathcal{D}$$

which implies that there exists $i \in \mathcal{D}$ such that

$$\frac{|A_i - A_i s|}{|A_i|} < \varepsilon$$

for all $s \in S$ which concludes the proof Kesten's theorem by the virtue of the Lemma 3.17.

The following corollary, roughly speaking, shows how an amenable subgroup H nicely sits inside the group G and mimics a finite group in the graph theoretic setting. Evidently, the following result points out why amenability is sometimes referred as a "finiteness" property.

Corollary 4.15. *Let G be a group generated by a finite symmetric set $S \subset G$, and let H be an amenable subgroup of G. Then we have*

$$\rho\left((Cay(G,S)\right) = \rho\left(Sch(G/H,S)\right).$$

Proof. Let $W(x,y,n)$ be the set of all walks of length n in the Cayley graph $Cay(G,S)$, starting at $x \in G$ and terminating in $y \in G$. Similarly, let $W(Hx,Hy,n)$ be the set of all walks of length n in the Schreier graph $Sch(G/H,S)$, with initial and terminal vertices Hx and Hy, respectively. We will denote the set of all walks starting at $x \in G$ by $W(x,n)$ for the Cayley graph. In the same fashion, let $W(Hx,n)$ be the set of all walks of length n in the Schreier graph. Using these notations, the returning probabilities are

$$p^{(2n)}(e,e) = \frac{|W(e,e,2n)|}{|W(e,2n)|} \text{ and } q^{(2n)}(H,H) = \frac{|W(H,H,2n)|}{|W(H,2n)|}$$

for the random walks (G,P) and $(G/H,Q)$ on Cayley and Schreier graphs, respectively. It is clear that $|W(x,2n)| = |W(H,2n)|$ for all $n \geq 0$. We can observe that the number of returning walks in the Schreier graph is more than that of the Cayley graph. Let $(e,s_1,...,s_1s_2\cdots s_{2n})$ be a returning walk in the Cayley graph. In other words, $s_1\cdots s_{2n} = e$. On the other hand, $Hs_1\cdots s_{2n} = He = H$. Therefore, every returning walk in the Cayley graph, corresponds to a returning walk in the Schreier graph. Hence, we have

$$q^{(2n)}(H,H) \geq p^{(2n)}(e,e).$$

This implies that

$$\rho(G/H,S) = \limsup_{n\to\infty} \sqrt[2n]{q^{(2n)}(H,H)} \geq \limsup_{n\to\infty} \sqrt[2n]{p^{(2n)}(e,e)} = \rho(G,S)$$

Now let R_n be the terminal point of the simple random walk on G starting at the identity. Then $Q[R_{2n} \in H]$ is equal to the returning probability for the simple random

walk on G/H starting at H. By the definition of the spectral radius, for every $\varepsilon > 0$ there exists a natural number N such that for all $n \geq N$ we have

$$q_{2n} := q^{(2n)}(H, H) \geq ((1 - \varepsilon)\rho))^{2n}$$

where $\rho = \rho(Sch(G/H, S))$. Let n be any natural number such that $n > N$, and for $h \in H$ define

$$\mu(h) = \frac{Q[R_{2n} = h]}{q_{2n}}.$$

It is obvious that μ is a probability distribution on H. Then μ induces a random walk (H, P') on H by virtue of the Proposition 2.25. Since the group H is amenable, the spectral radius of the random walk induced by μ is equal to 1 by Kesten's theorem. Hence, again by the definition of the spectral radius, there exists an integer m such that

$$p'^{(2m)}(e, e) \geq (1 - \varepsilon)^{(2m)}.$$

Now we will derive the relationship between the two random walks mentioned above. For $x, y \in H$, we have

$$p'(x, y) = \mu(x^{-1}y) = \frac{Q[R_{2n} = x^{-1}y]}{q_{2n}} = \frac{q^{(2n)}(e, x^{-1}y)}{q_{2n}}$$

We first claim that

$$p^{(4kn)}(x, y) \geq q_{2n}^{2k} p'^{(k)}(x, y)$$

for all $x, y \in H$ and any natural number k. We will do induction on $k \in \mathbb{N}$. Let us prove the claim for $k = 1$, i.e, $p^{(4n)}(x, y) \geq q_{2n}^2 p'(2)(x, y)$.

$$\begin{aligned}
p'^{(2)}(x, y) &= \sum_{z \in H} p'(x, z)p'(z, y) \\
&= \frac{1}{q_{2n}^2} \sum_{z \in H} p^{(2n)}(e, x^{-1}z)p^{(2n)}(e, z^{-1}y) \\
&= \frac{1}{q_{2n}^2} \sum_{z \in H} p^{(2n)}(x, z)p^{(2n)}(z, y) \\
&\leq \frac{1}{q_{2n}^2} \sum_{z \in G} p^{(2n)}(x, z)p^{(2n)}(z, y) \\
&= \frac{1}{q_{2n}^2} p^{4n}(x, y)
\end{aligned}$$

Suppose the inequality holds for k, then

$$\begin{aligned} p'^{(2(k+1))}(x,y) &= \sum_{z\in H} p'^{(2k)}(x,y)p'^{(2)}(x,y) \\ &\leq \frac{1}{q_{2n}^{2k}}\frac{1}{q_{2n}^{2}}\sum_{z\in G} p^{(4kn)}(x,z)p^{(4n)}(z,y) \\ &= \frac{1}{q_{2n}^{2(k+1)}}p^{(4n(k+1))}(x,y) \end{aligned}$$

Thus the claim is proven. In particular, it holds for m defined as in above. Since x and y were arbitrary, we have

$$p^{(4mn)}(e,e) \geq q_{2n}^{2m}p'^{(m)}(e,e) \geq \rho^{4mn}(1-\varepsilon)^{2m(2n+1)}.$$

Taking $4mn$-th roots and taking limit as $n \to \infty$ yields

$$\rho\left(Cay(G,S)\right) \geq (1-\varepsilon)\rho(Sch(G/H,S)).$$

□

CHAPTER 5

INVARIANT RANDOM SUBGROUPS AND KESTEN'S THEOREM

In this chapter, we will present Kesten's theorem for *Invariant Random Subgroups*(IRS) which is a Borel probability measure on the space of subgroups of a given countable group G such that measure is invariant under the group's action by conjugation. This concept helps us to study distribution and properties of subgroups within a large one in a probabilistic sense. Essentially, it provides a way to consider *random subgroups* while ensuring that the probabilistic setting respects the structure of the group. In other words, we can define an invariant random subgroup as a random subgroup whose distribution is invariant under conjugation of G. We will refer to [2], [5] in this chapter.

5.1 Invariant Random Subgroups

Let G be a countable group. Consider the power set of G, i.e. set of all subsets of of G. We know that there is a bijection between the power set of G and the set of all maps from G to $\{0,1\}$ (denoted by $\{0,1\}^G$). We can endow product topology on $\{0.1\}^G$ by considering product topologies of the discrete topology on $\{0,1\}$. Now consider

$$Sub(G) = \{H \subset \mid H \leq G\}.$$

Naturally, there is a topology on $Sub(G)$ induced by the product topology on $\{0,1\}^G$. In this topology, a sequence of subgroups H_n converges if for all $g \in G$, there exists $N \in \mathbb{N}$ such that either $g \in H_n$ or $g \notin H_n$ for all $n \geq N$. It is clear that all the limit points of sequences of subgroups will also be a subgroup. Therefore, $Sub(G)$ is a closed subset of $\{0,1\}^G$. By the Tychonoff Theorem, $\{0,1\}^G$ is a compact space

as G is countable. Thus, $Sub(G)$ is also compact. Moreover, G acts continuously on $Sub(G)$ by conjugation.

Definition 5.1. An *invariant random subgroup* (IRS) of G is a Borel probabaility measure on $Sub(G)$ that is invariant under the conjugation action of G on $Sub(G)$.

Example 5.2. Let N be a normal subgroup of G, then the Dirac measure δ_N is an invariant random subgroup on G.

Thus, in a sense, invariant random subgroup is a generalization of the normal subgroup.

Example 5.3. Let G be a group and H a subgroup such that it has only finitely many conjugates. Let N be the number of conjugates of H. Then the uniform measure μ assigning $1/N$ to each conjugate is an IRS.

The following example illustrates how we can obtain more IRS. Check [5], Proposition 6 for the details.

Proposition 5.4. *Suppose G acts on the Borel probability space (X, μ) by measure preserving maps. Then the maps $St : X \to Sub(G)$ given by $x \mapsto St_G(x)$ is Borel. Moreover, the push-forward of μ under the stabilizer map measure $\nu = St_*(\mu)$ is an IRS supported on $\{St_G(x) \mid x \in X\}$.*

It turns out every IRS arises in this way as it has been proven by Miklós Abért, Yair Glasner, and Bálint Virág ([2], Proposition 13). In other words, for every IRS there exists a measure-preserving action of G on Borel probability space (X, μ) such that the IRS is the stabilizer of μ-random point of X in G.

5.2 Kesten's theorem for invariant random subgroups

Now we will provide some general lemmas that will be used later. All results of this section are from [2].

Lemma 5.5. *Let G be a group generated by $S \subset G$, and let $n \geq 2$ be a even natural number. Then for the Cayley graph $Cay(G, S)$, we have*

$$|W(x, y, n)| \leq |W(x, x, n)| \leq |S|^2 |W(x, x, n-2)|.$$

Proof. Let $A = |S|M$, where M is the Markov operator defined as in the Chapter 4 (see page 63, (M)). Let δ_x be the characteristic function of x. Observe that $A^{n/2}\delta_x(z)$ gives us that the number of paths from z to x in $n/2$ steps. Indeed,

$$A^{n/2}\delta_x(y) = \sum_{s_1,\ldots,s_{n/2}\in S} \delta_x(zs_1\cdots s_{n/2})$$

gives the number of walks from x to z as $zs_1\cdots s_{n/2}$ corresponds to walks of length $n/2$ from z. Then

$$\sum_{g\in G} A^{n/2}\delta_x(g)$$

gives us the number of all walks of length $n/2$ to the vertex x. Hence, the number of all walks from x to y in n steps is

$$\sum_{g\in G}\left(\sum_{s_1,\ldots s_{n/2}\in S}\delta_x(gs_1\cdots s_{n/2})\right)\left(\sum_{s_1,\ldots,s_{n/2}\in S}\delta_y(gs_1\cdots s_{n/2})\right)$$

which is equivalent to

$$\langle A^{n/2}\delta_x, A^{n/2}\delta_y\rangle.$$

Thus, Cauchy-Schwarz inequality yields

$$\begin{aligned}
|W(x,y,n)| &= \langle A^{n/2}\delta_x, A^{n/2}\delta_y\rangle \\
&\leq \sqrt{\langle A^{n/2}\delta_x, A^{n/2}\delta_x\rangle}\sqrt{\langle A^{n/2}\delta_y, A^{n/2}\delta_y\rangle} \\
&= \langle A^{n/2}\delta_x, A^{n/2}\delta_x\rangle \\
&= |W(x,x,n)|
\end{aligned}$$

since we have $A^{n/2}\delta_x = A^{n/2}\delta_y$ for the Cayley graph is regular and homogeneous. For the second inequality, notice that

$$W(x,y,n) = \bigcup_{y,z \text{ are neighbors of x}} W(y,z,n-2).$$

Then using the first inequality and again taking into account that the Cayley graph is homogeneous and $|S|$-regular, we get

$$|W(x,x,n) = |S|^2|W(y,z,n-2)| \leq |S|^2|W(y,y,n-2)| = |S|^2W(x,x,n-2)|$$

where y, z are neighbors of x. □

Let G be a group and let $g \in G$ be a G-valued random variable. We define *Markov operator* $M(g)$ on $\ell^2(G)$ by means of

$$M(g)f(x) = \sum_{h \in G} \mathbf{P}[g = h] f(xh)$$

for $x \in G$.

Lemma 5.6. *Let $g_1, ..., g_n$ be independent random variables. Let $g = g_1 \cdots g_n$. One has*

$$\boldsymbol{P}[g = e] \leq \prod_{i=1}^{n} \|M(g_i)\| .$$

Proof. Let g_1, g_2 be independent G-valued random variables. Then $\mathbf{P}[g_1 g_2 = h] = \sum_{k \in G} \mathbf{P}[g_1 = k]\mathbf{P}[g_2 = k^{-1}h]$ by the independence of G-valued random variables. Then we have

$$\begin{aligned} M(g_1 g_2)f(x) &= \sum_{h \in G} \sum_{k \in G} \mathbf{P}[g_1 = k]\mathbf{P}[g_2 = k^{-1}h] f(xh) \\ &= \sum_{k \in G} \mathbf{P}[g_1 = k] \sum_{h \in G} \mathbf{P}[g_2 = k^{-1}h] f(xh) \\ &= \sum_{k \in G} \mathbf{P}[g_1 = k] \sum_{l \in G} \mathbf{P}[g_2 = l] f(xkl) \\ &= \sum_{k \in G} M(g_2)f(xk) \\ &= M(g_1)M(g_2)f(x). \end{aligned}$$

Hence, for independent $g_1, ..., g_n$ we have

$$M(g_1 \cdots g_n) = \prod_{i=1}^{n} M(g_i).$$

Now let δ_e be the characteristic function of identity in $\ell^2(G)$. Then

$$\begin{aligned} \mathbf{P}[g = e] &= \sum_{h \in G} \mathbf{P}[g = h]\delta_e(h) = \langle M(g)\delta_e, \delta_e \rangle \\ &\leq ||M(g)\delta_e|| ||\delta_e|| \leq ||M(g)|| \\ &\leq \prod_{i=1}^{n} ||M(g_i)||. \end{aligned}$$

□

The next lemma is modification of the Proposition 4.7.

Lemma 5.7. *Let G be a group and $S \subset G$ be a finite symmetric subset. Let $K = \langle S \rangle$. Suppose that g is the random element of S. Then*

$$||M(g)|| = \rho\left(Cay(K, S)\right)$$

where $M(g)$ is the Markov operator on $\ell^2(G)$.

Let G be a group and S be a finite symmetric multiset. Suppose that H is a subgroup of G. Let S^n be the set of n-tuples. For $a = (a_1, ..., a_n) \in S^n$, we will set $[a] = a_0 \cdots a_{n-1}$. Define

$$A_H(S, n) = \{a \in S^n : [a] \in H\}.$$

For the sake of simplicity, we will denote $A_{\{e\}}(S, n)$ by simple $A(S, n)$. We immediately notice that while $A_H(S, n)$ corresponds to the set of returning walks in the Schreier graph $Sch(G/H, S)$, the set $A(S, n)$ corresponds to that of the Cayley graph $Cay(G, S)$.

Lemma 5.8. *Let $(a_0, ..., a_{n-1})$ be a uniform random element of $A(S, n)$. Then the distribution of the segment $(a_t, ..., a_{t+k-1}) \in S^k$ for any $1 \leq k$ and $0 \leq t \leq n-1$ is independent of t, where the indices are understood modulo n.*

Proof. Let t, u be integers such that $0 \leq t, u \leq n-1$. We are claiming that the distribution of $(a_t, ..., a_{t+k-1}) \in S^k$ and $(a_u, ..., a_{u+k-1}) \in S^k$ are the same. The uniform distributions are exactly

$$\frac{|\{(b_t, ..., b_{t+k-1}) \in S^k \mid \exists (a_0, ..., a_{n-1}) \in A(S, n), b_0 = a_t, ..., b_{k-1} = a_{t+k-1}\}|}{|S|^k}$$

and

$$\frac{|\{(b_u, ..., b_{u+k-1}) \in S^k \mid \exists (a_0, ..., a_{n-1}) \in A(S, n), b_0 = a_u, ..., b_{k-1} = a_{u+k-1}\}|}{|S|^k}.$$

Without loss of generality, suppose that $u = t + m$ for some natural number m. For every $(a_0, ..., a_{n-1}) \in A(S, n)$, we have $a_0 \cdots a_{n-1} = e$. Observe that $A(S, n)$ is invariant under cyclic rotation, i.e. $a_t a_{t+1} \cdots a_{n-1} a_0 \cdots a_{t-1} = e$. For every $(a_0, ..., a_{n-1}) \in A(S, n)$ such that $b_0 = a_t, ..., b_{k-1} = a_{t+k-1}$, there is $(a_{t+m}, a_{t+m+1}, ..., a_{t+m-1})$ corresponding to $(b_u, ..., b_{u+k-1})$ and vice verse. □

Lemma 5.9. *Let $k \in \mathbb{N}$ and $(a_0, ..., a_{k-1})$ be a uniform random element of $A(S, n)$ such that $n > 2k$. Then we have*

$$\boldsymbol{P}\left[(a_0, ..., a_{k-1} = b)\right] \geq |S|^{-2k}$$

for any $b \in S^k$.

Proof. Suppose that $b = (b_0, ..., b_{k-1}) \in S^k$. For any $(c_0, ..., c_{n-2k-1}) \in A(S, n-2k)$,

$$(b_b, ..., b_{k-1}, b_{k-1}^{-1}, ..., b_0^1, c_0, c_1, ..., c_{2n-k-1-})$$

is clearly in $A(S, n)$. Therefore, we have

$$\mathbf{P}\left[(a_0, ..., a_{k-1}) = b)\right] \geq \frac{|A(S, n-2k)|}{|A(S, n)|}.$$

If we divide and multiply the numerator by $|W(e, e, n-2k)|$, we get $p^{(n-2k)}(e, e)|W(e, e, n-2k)|$. Similarly, we get $p^{(2n)}(e, e)|W(e, e, 2n)|$ in the denominator. Hence, we have

$$\mathbf{P}\left[(a_0, ..., a_{k-1} = b)\right] \geq \frac{p^{(n-2k)}(e, e)}{p^{(2n)}(e, e)} \frac{|W(e, e, n-2k|}{|W(e, e, 2n)|} = \frac{p^{(n-2k)}(e, e)}{p^{(2n)}(e, e)} |S|^{-2k}$$

Now using the Lemma 5.5, we will obtain the desired inequality by showing that $p^{(n-2k)}(e, e) \leq p^{(2n)}(e, e)$: By applying second inequality $2k$ times, we obtain

$$|W(e, e, 2n) \leq |S|^{2k} W(e, e, n-2k).$$

Hence,

$$\frac{|W(e, e, n-2k|}{|W(e, e, 2n)|} |S|^{2k} \geq 1.$$

Again by doing algebraic manipulations, we obtain that

$$\frac{p^{(n-2k)}(e, e)}{p^{(2n)}(e, e)} \frac{|W(e, e, n-2k)|}{|W(e, e, 2n)} |S|^{2k} \geq 1.$$

On the other hand, $\dfrac{|W(e, e, n-2k)|}{|W(e, e, 2n)|} = |S|^{-2k}$. Hence,

$$\mathbf{P}\left[(a_0, ..., a_{k-1}) = b)\right] \geq |S|^{-2k}.$$

□

For $a \in S^n$, consider the event that the walks corresponding to a in $Sch(G/H, S)$ is at the same vertex at times $t-1$ and t, i.e.

$$Ha_0 \cdots a_{t-1} = Ha_0 \cdots a_t.$$

The last equality is equivalent to

$$a_t \in H^{a_0 \cdots a_{t-1}}.$$

Let us fix $a \in S^n$ and define the index set

$$I(S, H, a) = \{t \in \{0, ..., n-1\} \mid a_t \in H^{a_0 \cdots a_{t-1}}\}.$$

Definition 5.10. We call $a, b \in S^n$ H-equivalent if

$$Ha_0 \cdots a_t = Hb_0 \cdots b_t \ \ (0 \le t \le n-1) \ \text{ and } \ a_t = b_t \ \ (t \notin I(S, H, a)).$$

It is easy to check that this indeed a and b are in equivalence relation. Let us denote the equivalence class of a by $C(a)$.

Lemma 5.11. *Let G be a group, and $S \subset G$ be a finite symmetric multiset. Suppose that H is a subgroup of G. Then*

$$|A_H(S, n)| \ge \sum_{a \in A(S,n)} [p(S, H, a)]^{-1},$$

where $n\mathbb{N}$ and $p(S, H, a) = \prod_{t \in I(S,H,a)} \rho(Cay(\langle S \cap H^{a_0 \cdots a_{t-1}} \rangle, S \cap H^{a_0 \cdots a_{t-1}})$

Proof. For $0 \le t \le n-1$, let

$$X_t = S \cap H^{a_0 \cdots a_{t-1}}.$$

Since X_t is formed by intersecting elements of symmetric multiset S and the subgroup $H^{a_0 \cdots a_{t-1}}$, it is also a symmetric multiset. Now let $a \in A_H(S, n)$, that is, $Ha_0 \cdots a_{n-1} = H$. Then by the definition of H-equivalence $Hb_0 \cdots b_{n-1} = H$ for every $b = (b_0, ..., b_{n-1}) \in C(a)$. Thus, $C(a) \subset A_H(S, n)$. This implies that $A_H(S, n)$ partitions into disjoint union of its H-equivalence classes. Let x be any element of $C(a)$. Then it is of the form $(x_1, ..., x_n)$, where $x_t = a_t$ for $t \notin I(S, H, a)$, and x_t is

a uniform random element of X_t for $t \in I(S, H, a)$. By the virtue of Lemma 5.6 and Lemma 5.7, we obtain that

$$\begin{aligned}
\mathbf{P}[x = e] &= \mathbf{P}[x \in A(S, n)] \\
&\leq \prod_{t=1}^{n} \rho(Cay(\langle S \cap H^{a_0 \cdots a_{t-1}} \rangle, S \cap H^{a_0 \cdots a_{t-1}}) \\
&\leq \prod_{t \in I(S,H,a)} \rho(Cay(\langle S \cap H^{a_0 \cdots a_{t-1}} \rangle, S \cap H^{a_0 \cdots a_{t-1}}) \\
&= p(S, H, a).
\end{aligned}$$

where the last inequality follows from the fact that the spectral radius is less than or equal to 1. Thus, we have

$$\sum_{a \in A(S,n)} [p(S, H, a)]^{-1} \leq \sum_{a \in A(S,n)} \mathbf{P}[x \in A(S, n)]^{-1}.$$

We will naturally conclude the proof that the right-hand side of the inequality is less than or equal to $|A_H(S, n)|$. Let us note that

$$\mathbf{P}[(x \in A(S, n)] = \frac{|A(S, n) \cap C(a)|}{|C(a)|}.$$

We have

$$\begin{aligned}
\sum_{a \in A(S,n)} \mathbf{P}[x \in A(S, n)]^{-1} &= \sum_{a \in A(S,n)} \frac{|C(a)|}{|A(S, n) \cap C(a)|} \\
&\leq \sum_{a \in A(S,n)} \frac{|C(a)|}{|A(S, n) \cap C(a) \cap A_H(S, n)|} \\
&= \sum_{a \in A(S,n)} \frac{|C(a)|}{|A_H(S, n)|} \\
&\leq \sum_{a \in A(S,n)} \frac{|A_H(S, n)|}{|A_H(S, n)|} \\
&= |A(S, n)| \\
&\leq |A_H(S, n)|.
\end{aligned}$$

□

Let us denote

$$[S^k] = \{a_0 \cdots a_{k-1} | a_i \in S^k, 0 \leq i \leq k - 1\}$$

where we see $[S^k]$ as a multiset in G. Let $\{S^k\}$ be the set of elements in $[S^k]$, the set which we obtain by removing the duplicates of the elements.

Lemma 5.12. *Let G be a group and $S \subset G$ be its finite symmetric generating set. Let H be an non-amenable invariant random subgroup with positive probability. Then there are integers $p, k, r > 0$ such that*

$$\rho(Cay(H, S^{2k} \cap H)) < 1 - r$$

with probability at least p.

Proof. For any $k \geq 0$, we have $\{S^k\} \subset \{S^{k+2}\}$ as we can add ss^{-1} to the words in $\{S^k\}$. Define

$$N = \bigcup_{k=0}^{\infty} \{S^{2k}\}.$$

The set N is the set of all words of even length obtained by multiplying the elements of the generating set S. We can see that N is actually a subgroup of G. Indeed, it contains the identity element and the product of two words of even length is again even. Moreover, the inverse of a word of even length is again of the even length. Furthermore, we claim that N has index at most 2. If $S \subset N$, we have $\langle S \rangle \subset N$ which implies that $N = G$. If $S \not\subset N$, then there exists $s \in S$ such that $s \notin N$. Thus, we have $sN \neq N$ which implies that $G \neq N$. Then for any $g \in G \setminus N$, g is of the form $s_1 \cdots s_{2m+1}$ for some $m \in \mathbb{N}$. Multiplying both sides by s^{-1}, we get

$$s^{-1}g = s^{-1}s_1 \cdots s_{2k+1}.$$

Since $s^{-1}g$ is of an even length, we get that $s^{-1}g \in N$ which is equivalent to $sN = gN$. Therefore, there can only be two cosets of N in G.

Now let H_{2k} be the subgroup generated by $\{S^{2k}\} \cap H$. Observe that union of the subgroups H_{2k} is just $H \cap N$. If N is the whole group G, then the index of $H \cap N$ in H is 1. If N has an index 2, then it is well known that N is normal in G. Thus, we can apply the second isomorphism theorem and obtain that

$$[H : H \cap N] = [HN : N] \leq [G : N] = 2$$

Thus, $\bigcup_{k=0}^{\infty} H_{2k} = H \cap N$ has in index at most 2 in H. If all the subgroups H_{2k} were amenable, then the group H itself would be amenable since taking countable

union preserves amenability. Hence, there exists k such that H_{2k} is non-amenable with positive probability since

$$\{H \leq G \mid H \text{ is non-amenable}\} = \{H \leq G \mid \exists k \geq 0, \ H_{2k} \text{ is non-amenable}\}.$$

On the other hand, by Kesten's Theorem 4.1, we have

$$\begin{aligned}&\{H \leq G \mid \exists k > 0, \ H_{2k} \text{ is non-amenable}\} = \\ &= \left\{H \leq G \mid \exists k \leq 0, \rho(H_{2k}, [S^{2k}] \cap H)) < 1\right\}\end{aligned}$$

Now we can stratify the last set as follows:

$$\begin{aligned}&\left\{H \leq G \mid \exists k > 0, \rho(H_{2k}, [S^{2k}] \cap H)) < 1\right\} = \\ &= \bigcup_{r \in \mathbb{Q}^+} \bigcup_{k>0} \left\{H \leq G \mid \exists k > 0, \rho(Cay(H_{2k}, [S^{2k}] \cap H)) < 1 - r\right\}\end{aligned}$$

where $\mathbb{Q}^+$ denotes the set of positive rational numbers. Thus, there exists a positive natural number $k > 0$ and $r > 0$ such that

$$\mathbf{P}\left[\{H \leq G \mid \rho(Cay(H, [S^2 k] \cap H)) < 1 - r\}\right] \geq p.$$

□

We will include a standard inequality which will be used in the proof main result.

Lemma 5.13. *Let X be a random variable such that $0 \leq X \leq R$. Then*

$$\boldsymbol{P}\left[X \geq \frac{E[X]}{2}\right] \geq \frac{\boldsymbol{E}[X]/R}{2 - \boldsymbol{E}[X]/R}$$

where $\boldsymbol{E}[X]$ denotes the expectation of the random variable X.

Theorem 5.14 (Kesten's Theorem for Invariant Random Subgroups)**.** *Let G be a group generated by a finite symmetric set subset S, and let H be an invariant random subgroup of G. Then the following are equivalent*

(1) $\rho(Sch(G/H, S)) = \rho(Cay(G, S))$ almost surely;

(2) H is amenable almost surely.

Proof. (1) By the Corollary 4.15, amenability of H almost surely implies that

$$\rho(Sch(G/H, S)) = \rho(Cay(G, S))$$

almost surely.

Now assume that H is non-amenable with positive probability. Then by the Corollary 5.12, there are $k, r > 0$ such that

$$\rho(Cay(H, [S^k] \cap H)) < 1 - r$$

with probability at least $p > 0$.

Let $T = [S^k]$ and fix a natural number $n \in \mathbb{N}$. Define

$$J(T, H, a) = \{t \in I(T, H, a) \mid \rho(\langle T \cap H^{a_0 \cdots a_{t-1}} \rangle, T \cap H^{a_0 \cdots a_{t-1}}) < 1 - r\}$$

Then we have

$$\begin{aligned} p(T, H, a) &= \prod_{t \in I(T,H,a)} \rho(\langle T \cap H^{a_0 \cdots a_{t-1}} \rangle, T \cap H^{a_0 \cdots a_{t-1}}) \\ &\leq \prod_{t \in J(T,H,a)} \rho(\langle T \cap H^{a_0 \cdots a_{t-1}} \rangle, T \cap H^{a_0 \cdots a_{t-1}}) \\ &\leq (1 - r)^{|J(T,H,a)|}. \end{aligned}$$

taking into account that the spectral radius is less than or equal to 1. Then the Lemma 5.11 yields

$$|A_H(T, n)| > (1 - r)^{-|J(T,H,a)|} \tag{A}$$

Now For a fixed $a \in A(T, n)$ and $0 \leq t \leq n - 1$ we obtain

$$\begin{aligned} &\mathbf{P}[t \in J(T, H, a)] = \\ &\qquad = \mathbf{P}[a_t \in H^{a_0 \cdots a_{t-1}} \text{ and } \rho(\langle T \cap H^{a_0 \cdots a_{t-1}} \rangle, T \cap H^{a_0 \cdots a_{t-1}}) < 1 - r] \\ &\qquad = \mathbf{P}[a_t \in H \text{ and } \rho(Cay(H, T \cap H)) < 1 - r] \end{aligned}$$

by the invariance of H. Using the fact that events written above are independent, we obtain the expected value of $|J(T, H, a)|$ over H.

$$\mathbf{E}[|J(T, H, a)|] = \sum_{t=0}^{n-1} \mathbf{P}[a_t \in H \text{ and } \rho(Cay(H, T \cap H)) < 1 - r]. \tag{B}$$

For a fixed subgroup H and $0 < r < 1$ such that $\rho(Cay(H, T \cap H)) < 1 - r$, we have $T \cap H \neq \emptyset$ as otherwise the spectral radius would be 1. Furs such H's and $0 \leq n - 1$,

using Lemma 5.8 and Lemma 5.9, we have

$$\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}\mathbf{P}[a_t\in H]\geq\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}\mathbf{P}[a_t=b]\geq|T|^{-2}.$$

Now using the independency of the events $\mathbf{P}[a_t\in H]$ and $\mathbf{p}[\rho(Cay(H,T\cap H))<1-r]$, we obtain

$$\begin{aligned}&\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}\mathbf{P}[a_t\in H\text{ and }\rho(Cay(H,T\cap H))<1-r)]=\\&=\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}\mathbf{P}[a_t\in H]\mathbf{P}[\rho(Cay(H,T\cap H))<1-r)]\\&\geq p|T|^{-2}\end{aligned}$$

using the fact that

$$\mathbf{P}[\rho(Cay(H,T\cap H))<1-r)]]\geq p.$$

Therefore, summing with respect to t and using (B), we have

$$\mathbf{E}\left[\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}|J(T,H,a)|\right]\geq p|T|^{-2}n.$$

On the other hand,

$$\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}|J(T,H,a)|\leq\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}n\leq n.$$

by the definition of $J(T,H,a)$.

Hence, using Lemma 5.13 and monotonicity of the expected value, we obtain

$$\mathbf{P}\left[\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}|J(T,H,a)|\geq\frac{1}{2}p|T|^{-2}n\right]\geq\frac{p|T|^{-2}}{2-p|T|^{-2}}.\qquad\text{(C)}$$

The last inequality shows that there is an invariant random subgroup H which satisfying (C).

Now let H be a subgroup satisfying (C). By the inequality on arithmetic and geometric means, we have

$$\begin{aligned}\frac{1}{|A(T,n)|}\sum_{a\in A(T,n)}(1-r)^{-|J(T,H,a)|}&\geq(1-r)^{(-1/|A(T,n)|\sum_{a\in A(T,n)}|J(T,H,a)|}\\&\geq(1-r)^{(-p/2)|T|^{-2}n}.\end{aligned}$$

Now (A) implies

$$|A_H(T,n)| > \sum_{a \in A(T,n)} (1-r)^{-|J(T,H,a)|} \geq |A(T,n)|(1-r)^{(-p/2)|T|^{-2}n}.$$

Since $T = [S^k]$, get $|A_H(T,n)|$ and $|A(T,n)| = |A(S,kn)|$. Recall that $|A_H(S,kn)|$ and $|A(S,kn)|$ correspond to the number of returning walks in the Schreier and Cayley graphs. Denote the return probability on the Schreier graph and Cayley graphs, respectively, by $q^{(kn)}(H,H)$ and $p^{(kn)}(e,e)$ in kn steps. Thus, dividing by $|S|^{kn}$ and taking *nk*th roots, we get that for every natural number $n > 1$ an invariant random subgroup H with the probability at least $p|T|^{-2}/(2-p|T|^{-2})$ satisfies

$$q^{(nk)}(H,H)^{1/nk} > p^{(nk)}(H,H)^{1/nk}(1-r)^{-(1/2k)p|T|^{-2}}.$$

Note that $(1-r)^{-(1/2k)p|T|^{-2}}$ is a constant which is greater than 1. Then taking limits as n tends to ∞, we obtain

$$\rho(Sch(G/H,S)) \geq \rho(Cay(G,S)(1-r)^{-(1/2k)p|T|^{-2}} > \rho(Cay(G,S))$$

with probability at least $p|T|^{-2}/(2-p|T|^{-2})$. □

REFERENCES

[1] *An Introduction to Functional Analysis*, pages 259–290. Springer New York, New York, NY, 1998.

[2] M. Abért, Y. Glasner, and B. Virág. Kesten's theorem for invariant random subgroups. *Duke Math. J.*, 163(3):465–488, 2014.

[3] T. A. Banach Stefan. Sur la decomposition des ensembles de points en parties respectivement congruentes. *Fundamenta Mathematicae*, 6(1):244–277, 1924.

[4] B. Beauzamy. *Introduction to operator theory and invariant subspaces*, volume 42 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 1988.

[5] M. G. Benli, R. Grigorchuk, and T. Nagnibeda. Universal groups of intermediate growth and their invariant random subgroups. *Funct. Anal. Appl.*, 49(3):159–174, 2015. Translation of Funktsional. Anal. i Prilozhen. **49** (2015), no. 3, 1–21.

[6] T. Ceccherini-Silberstein and M. Coornaert. *Cellular automata and groups*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2010.

[7] T. Ceccherini-Silberstein and M. D'Adderio. *Topics in groups and geometry—growth, amenability, and random walks*. Springer Monographs in Mathematics. Springer, Cham, [2021] ©2021. With a foreword by Efim Zelmanov.

[8] H. Kesten. Symmetric random walks on groups. *Transactions of the American Mathematical Society*, 92(2):336–354, 1959.

[9] C. Löh. *Geometric group theory*. Universitext. Springer, Cham, 2017. An introduction.

[10] J. Neumann. Zur allgemeinen theorie des masses. *Fundamenta Mathematicae*, 13(1):73–116, 1929.

[11] W. Woess. *Random walks on infinite graphs and groups*, volume 138 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2000.

[12] W. Woess. *Denumerable Markov Chains*. EMS Textbooks in Mathematics. European Mathematical Society, Switzerland, 1 edition, 2009. XVII, 351 S.