PROBABILITY ON GRAPHS AND GROUPS: THEORY AND APPLICATIONS

Natalia Mosina

Advisor: Ioannis Karatzas

Submitted in partial fulfillment of the
Requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2009

Abstract

PROBABILITY ON GRAPHS AND GROUPS: THEORY AND APPLICATIONS

Natalia Mosina

We introduce the notion of the mean-set (expectation) of a graph- or group-valued random element. Using this concept, we prove a novel generalization of the strong law of large numbers on graphs and groups. Some other relevant results about configurations of mean-sets (or center-sets) in trees and free groups, which may be of independent interest, are discussed. We enhance our theory with other theoretical tools, such as an analogue of Chebyshev inequality for graphs and the notion of central order on graphs. Furthermore, we consider technical difficulties of computing sample mean-sets and some practical ways of dealing with this issue. Moreover, we provide results of actual experiments supporting many of our conclusions. In addition, we show that our generalized law of large numbers, as a new theoretical tool, provides a framework for motivating practical applications; namely, it has implications for group-based cryptanalysis. At the end of this exposition, we explain, among other things, how to analyze the security of a particular zero-knowledge, i.e., security preserving, group-based authentication protocol. Our analysis allows us to conclude that the security and reliability of a well-known authentication scheme in group-based cryptography proposed by Sibert is questionable. The present work provides a completely new direction of such analysis – it shows that there is a probabilistic approach to cryptographic problems, which are usually treated only from the algebraic point of view, and that this approach can be very effective.

Contents

1 Introduction **1**

1.1 Introduction 1

1.1.1 Some history and theoretical motivation 2

1.1.2 The core of our work 3

1.1.3 Further developments 6

1.1.4 Practical Motivation 8

1.2 Summary 13

2 Preliminaries on graphs and groups **17**

2.1 Graphs 17

2.2 Cayley graphs and groups 18

2.3 Free groups 20

2.4 Free abelian groups 22

2.5 A group of braids and its presentation 23

3 Mean-set and the strong law of large numbers – I. **27**

3.1 Mean of a group-valued random element 27

3.1.1 The mean set in a graph 28

3.1.2 Other possible definitions of $\mathbb{E}$ 32

3.1.3 Centers of Rubinshtein vs. mean-sets (center-sets). 34

3.2 Strong Law of Large Numbers (SLLN) 36

3.3 SLLN for multi-vertex mean-sets 44

3.3.1 Case of 2 vertices 44

3.3.2 Case of 3 and more vertices 46

3.3.3 The case when random walk is not genuinely $(k-1)$-dimensional. 54

3.4 Cones, subspaces, and monoids. 56

3.4.1 Properties of cones 56

3.4.2 Subspaces and monoids 58

3.5 Example of random walk for a cyclic graph. 60

4 Chebyshev's inequality on graphs 62

4.1 Introduction 62

4.2 Concentration of measure inequality on graphs (groups) 63

4.2.1 Singleton mean-set 64

4.2.2 Multi-vertex mean-set 68

5 Computations, applications, and experiments 69

5.1 Configurations of mean-sets 69

5.2 Computation of mean-sets 76

5.3 Experiments 79

5.3.1 Free group results 80

5.3.2 Free abelian group results 81

6 Refinements. Central order on $V(\Gamma)$. 88

6.1 Medians as mean-sets of class one 88

6.2 Expectation when $M^{(1)}$ is defined 92

6.2.1 Central order on Γ and general $\mathbb{E}(\xi)$ 93

6.2.2 SLLN and other results for central order 100

6.2.3 Mean sets and L^c-spaces 105

6.2.4 Examples of central order 108

7 Practical applications to cryptanalysis 112

7.1 Introduction 112

7.2 Zero-Knowledge (security-preservation) property. Some ideas. 114

7.2.1 Intuition 114

7.2.2 More precise realization 115

7.3 The protocol 116

7.3.1 Set up and description 117

7.3.2 Analysis 118

7.4 Effective computation of a mean-set 120

7.4.1 Length function for braids 121

7.5 The Mean-Set Attack 122

7.6 Experimental results 125

7.6.1 Platform group and parameters 125

7.6.2 Experiments 126

7.7 Defending against the Mean-Set Attack 128

7.7.1 Large mean-set 128

7.7.2 Undefined mean-set 129

7.7.3 Groups with no efficiently computable length functions 129

8 Further ideas. Problems and questions. 130

8.1 Centers and group geometry 130

8.1.1 Expectation and boundary of the graph 130

8.1.2 Some small problems about mean-sets and group geometry . . 132

8.2 Other goals. 133

8.3 Conclusion 134

Bibliography **134**

List of Figures

1.1 Labyrinth. Security-preserving scheme. 9

2.1 The Cayley graph of a free group F_2 22

2.2 A 4-strand braid. 24

2.3 Product of braids. 24

2.4 Isotopy of braids. 25

2.5 A trivial 4-strand braid. 25

2.6 Generators of B_4 and their inverses. 26

2.7 Inversion of a 4-strand braid. 26

2.8 Typical relations in braids. 26

3.1 $\mathbb{E}^{(2)}$ does not coincide with the classical mean in $\mathbf{R}^2$. 34

3.2 Cycle-4. 60

5.1 Impossible configuration of centers (gray vertices). 70

5.2 A graph with a cut point v_0. 70

5.3 Area of possible triangle side lengths. 72

5.4 Another example of configuration of centers (gray vertices) 75

5.5 Graph with local minimum of M, which is not a center (global minimum) 79

6.1 Example of central order relation $<^{(2)}$ in $\mathbf{Z}^2$. 109

List of Tables

3.1 Increments of the random walk 61

5.1 The results of experiment for F_2. 82

5.2 The results of experiment for F_4. 83

5.3 The results of experiment for F_6. 84

5.4 The results of experiment for A_2 - cut to fit. 85

5.5 The results of experiment for A_4. 86

5.6 The results of experiment for A_4 - continue. Cut to fit. 87

7.1 Experiments in B_5. 127

7.2 Experiments in B_{10}. 127

7.3 Experiments in B_{20}. 127

Acknowledgments

First and foremost, I would like to express my tremendous gratitude to my adviser, Ioannis Karatzas, for his generous time and commitment, for support and encouragement of all my endeavors, for open-mindedness, and for always being there for me not only in mathematics, but in everyday life for all the years of my Ph.D. studies. I consider it a privilege to have him as my mentor.

I would also like to extend my most sincere appreciation to my defense committee, David Bayer, Maria Chudnovsky, Ioannis Karatzas, Alexander Ushakov, and Mihalis Yannakakis, for their time, patience in reading this work, and their feedback on my dissertation. It is my big honor to defend my thesis in front of them.

My special thanks to Alexander Ushakov from Stevens Institute of Technology for introducing me to a wonderful and exciting field of cryptanalysis of group-based cryptographic protocols. He influenced the direction of my research, in terms of practical applications, to the great extent, and I found this collaboration very rewarding.

Another special and immeasurable gratitude that I would like to express goes to Panagiota Daskalopoulos. Her everyday support throughout my studies in Columbia University have always made me believe that everything is possible. She is greatly responsible for my cheerful state of mind, good spirits and hopes, everything that everybody needs when in graduate school.

I also owe a debt and a token of my deep appreciation to Frederick Gardiner, Raymond Gittings, Keith Harrow, George Shapiro, Wolfe Snow, and Noson Yanofsky from the City University of New York for encouraging me to go to a graduate school and believing in my ability to succeed. Their friendly advice, belief, and confidence helped me a lot throughout my undergraduate and graduate studies. I could always turn to them and know that they would be there for me, at all times, whether it is a despair or happiness. Thank you very much for that.

In addition, I wish to thank all my friends and colleagues who created a very friendly atmosphere around me throughout all the years that I spent in Columbia University. I am especially grateful to Tomoyuki Ichiba, Daniel Krasner, Adam Levine, Chen-Yun Lin, Qing Lu, and MingMin Shen. They have no relation to this work directly, but, as I said, their friendliness and thoughtfulness made it much easier for me to succeed.

Finally, my special thanks to my family for all the patience and understanding. My father, if he was still with us, would have been immensely proud to see his daughter getting a doctorate degree. He always encouraged me to go all the way up in my education. I dedicate this work to his memory.

My research was partially supported by the National Science Foundation (under grant NSF - DMS - 06 - 01774 and NSF - 0354639)

To the memory of my father.

Chapter 1

Introduction

"Mathematics compares the most diverse phenomena and discovers the secret analogies that unite them."

Josephe Fourier

1.1 Introduction

Random objects with values in groups and graphs are often dealt with in many areas of applied mathematics and theoretical computer science. Group-valued random elements would have a broader range of applications (supported by rigorous mathematical treatment) if we had such theoretical tools as the notion of the average (expectation) of a random element in a group, laws of large numbers with respect to this average, some results on the rate of convergence in these laws, and so forth. The purpose of this work is to start developing a necessary probability theory on graphs and groups that would serve as a theoretical framework for practical applications in different areas dealing with random objects in graphs and groups, in particular, in group-based cryptography, where these objects are very important.

1.1.1 Some history and theoretical motivation

One of the most profound and, at the same time, fundamental results studied in probability theory is undoubtedly *the strong law of large numbers* (SLLN). For independent and identically distributed (i.i.d.) real-valued random variables ξ_i, it states that

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \rightarrow \mathbb{E}(\xi_1) \tag{1.1}$$

almost surely (with probability one) as $n \rightarrow \infty$, provided that expectation $\mathbb{E}(\xi_1)$ is finite. Depending on the assumptions one is willing to make there are different versions of the result, but, nevertheless, they all are known under the general name of *laws of large numbers*. On the one hand, the assertion above may seem to be elementary in the sense that it is intuitively transparent that a sample (empirical) average should converge to a population (theoretical) average. On the other hand, the statement is indisputably deep because it entails the possibility to acquire precise information about randomly-occurring phenomena, or, quoting A. V. Skorokhod (35), "it allows us to make reliable conclusions from random premises," meaning that the average of a large number of random variables is "practically" non-random.

Starting from $1950's$, there have been ongoing attempts in the probabilistic literature to explore the existence of generalizations of the strong law of large numbers to general group-valued random variables. For example, works of R. Bellman (2) and H. Kesten (24) regarding the behavior of $Z_n := g_1g_2\dots g_n$ as $n \rightarrow \infty$ (where g_i are i.i.d. random variables in a general group G) date back to 1954 and 1959. The object of study for the desirable generalization was a random walk on groups. In 1960, the generalization of the law was obtained by Furstenberg and Kesten only for groups of matrices, which was further generalized in 1968. More detailed account of existing attempts to prove generalizations of the strong law of large numbers dealing with the objects of type $Z_n := g_1g_2\dots g_n$ for groups can be found in the recent work (year of 2006) by Anders Karlsson and François Ledrappier (23) where they present their version of a general law of large numbers for random walks on general groups. Nevertheless, even a minor touch on the history of the subject dealing with generalizations of SLLN to groups, employing the random walk Z_n, reveals some inherent difficulty of this long-term endeavor. This is from where we derive our theoretical motivation to look at the problem from another angle, and this is what we want to contrast with a drastically different approach our work is going to undertake.

As Anders Karlsson and François Ledrappier reasonably notice, it is not clear how exactly to formulate an assertion that would generalize the law of large numbers to groups. We think that this is precisely what causes the potential difficulty – the absence of clear intuitive grounds in dealing with $g_1g_2 \ldots g_n$ in a desirable generalization. Indeed, a mere attempt to mimic the left-hand side of (1.1) by writing $\sum_{i=1}^n \xi_i$ in a multiplicative form for group elements immediately eliminates its interpretation as an average, which is the heart of the matter of the SLLN (1.1). Keeping this in mind, we are not going to formulate the strong law of large numbers for group-valued random elements by considering a random walk on a general group that would lead to losing the information contained in the elements of a given group on the one side, as well as to hindering the idea behind the strong law of large numbers itself on the other side. Instead, we adhere to the basic principles in our approach. We remember that laws of large numbers establish that the average (or mean) of random variables converges to the average of their expectations (or just the expectation of a random variable in the i.i.d. case). In other words, the classical SLLN states that the sample mean should converge to the population mean with probability one. We want this fundamental idea to be reflected in our generalization of the strong law of large numbers for graphs and groups. We reach this goal in several steps.

1.1.2 The core of our work

Consider a locally finite graph $\Gamma = (V(\Gamma), E(\Gamma))$ (see Section 2 for basic graph- and group-theoretic preliminaries). First, we introduce the notion of the mean-set (expectation) $\mathbb{E}$ for graph-valued random elements $\xi : \Omega \to V(\Gamma)$ defined on a given probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Dealing with a random element $\xi(\omega)$, we find it convenient to work with a new (image) probability space $(V(\Gamma), \mathcal{S}, \mu)$ where μ is the atomic probability measure on $V(\Gamma)$ induced by ξ and defined by

$$\mu(g) = \mathbf{P}\Big(\{\omega \in \Omega \mid \xi(\omega) = g\}\Big),\ g \in V(\Gamma).$$

Next, we introduce a *weight function* $M_\xi : V(\Gamma) \to \mathbf{R}$ by

$$M_\xi(v) := \sum_{s \in V(\Gamma)} d^2(v,s)\mu(s),$$

where $d(v,s)$ is the distance between v and s in Γ, and prove that the domain of definition of $M_\xi(\cdot)$ is either the whole $V(\Gamma)$, in which case we say that M is *totally defined*, or $\emptyset$. In the case when $domain(M_\xi) = V(\Gamma)$, we define the mean-set of the graph-valued random element ξ to be

$$\mathbb{E}(\xi) := \{v \in V(\Gamma) \mid M_\xi(v) \leq M_\xi(u), \quad \forall u \in V(\Gamma)\}. \tag{1.2}$$

Observe the analogy with classical theory, where quadratic function $\mathbb{E}[(\xi_1 - c)^2]$ achieves its minimum at $c = \mathbb{E}(\xi)$ if $\xi_1, \xi_2, \ldots$ are i.i.d. L^2 real-valued random variables. The above definition of $\mathbb{E}(\xi)$, $\xi : \Omega \to V(\Gamma)$, provides the corresponding notion for groups via their Cayley graphs.

Next, we consider the empirical measure on $V(\Gamma)$, denoted by μ_n, of the sample of random graph elements $\xi_1(\omega), \ldots, \xi_n(\omega)$. In other words, for every $\omega \in \Omega$, $\mu_n(u;\omega) = \mu_n(u)$ (suppressing the second argument) is the *relative frequency* with which the value $u \in V(\Gamma)$ occurs in the sample above (see (3.6) further in the exposition), and $\mu_n \to \mu$ almost surely as $n \to \infty$. We let $M_n(v) := \sum_{i \in V(\Gamma)} d^2(v,i)\mu_n(i)$ be the random weight, corresponding to $v \in V(\Gamma)$, and $M_n(\cdot)$ the resulting random *sampling weight function*. Now, we define the *sample mean-set* relative to the sample $\{\xi_1, \ldots, \xi_n\}$ as the set of vertices

$$\mathbb{S}(\xi_1, \ldots, \xi_n) := \{v \in V(\Gamma) \mid M_n(v) \leq M_n(u), \quad \forall u \in V(\Gamma)\}. \tag{1.3}$$

The function $\mathbb{S}(\xi_1, \ldots, \xi_n)$ is an analogue of the average function $(x_1, \ldots, x_n) \mapsto (x_1 + \ldots + x_n)/n$ for $x_1, \ldots, x_n \in \mathbf{R}$. We let $\mathbb{S}_n = \mathbb{S}(\xi_1, \ldots, \xi_n)$. With these notions at hand, we prove a generalization of the strong law of large numbers on graphs and groups.

Theorem A. (SLLN for graph-valued random elements) *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If $M_{\xi_1}(\cdot)$ is totally defined and the mean set $\mathbb{E}(\xi_1) = \{v\}$ for some $v \in V(\Gamma)$, i.e., $\mathbb{E}(\xi_1)$ is a singleton set, then the following holds*

$$\mathbb{S}(\xi_1, \ldots, \xi_n) \xrightarrow{a.s.} \mathbb{E}(\xi_1) \text{ as } n \to \infty.$$

Note: Further in the exposition, we may find it convenient to write $\mathbb{E}(\mu)$, and to speak of the mean-set of the distribution μ induced by ξ on $V(\Gamma)$.

As we shall see (cf. Example 3.12 below), the classical *limit* of sets does not work in instances when mean-sets contain more than one vertex. Nevertheless, *limits superior (limsup's)* do the job in these cases. We discuss more general strong law of large numbers for multi-point mean-sets in Section 3.3 in Theorems 3.17, 3.20, 3.25, and Corollary 3.29.

It turns out that other definitions of mean-sets are possible. Given a random element ξ, we define a set

$$\mathbb{E}^{(c)}(\xi) := \{v \in V(\Gamma) \mid M_\xi^{(c)}(v) \leq M_\xi^{(c)}(u), \;\; \forall\, u \in V(\Gamma)\} \tag{1.4}$$

and call it a *mean-set of class c*, where

$$M_\xi^{(c)}(v) := \sum_{s \in V(\Gamma)} d^c(v, s)\mu(s),\ v \in V(\Gamma)$$

is the *weight function of class c.* Despite these different possibilities, we choose to work with (1.2), which is just a mean-set of class two, for reasons explained in Chapter 3, Section 3.1.2.

Once we have the notion of mean-set for a graph- and group-valued random elements, we notice that it satisfies the so-called "shift" property; namely,

$$\mathbb{E}(g\xi) = g\mathbb{E}(\xi), \forall g \in G. \tag{1.5}$$

This "shift" property proves to be very useful in practice because, together with the strong law of large numbers for groups, it allows our theory to be in accord with practical motivations and applications.

To enhance our theory with yet another tool, we prove an analogue of classical Chebyshev's inequality - the concentration of measure inequality for a graph- (group-)valued random element ξ.

Theorem B. (Chebyshev's inequality for graph-valued random elements) *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If the weight function $M_{\xi_1}(\cdot)$ is totally defined then there exists a constant $C = C(\Gamma, \xi_1) > 0$ such that*

$$\mathbf{P}\Big(\mathbb{S}(\xi_1, \ldots, \xi_n) \not\subseteq \mathbb{E}(\xi)\Big) \leq \frac{C}{n}.$$

1.1.3 Further developments

We go further than proving the novel strong law of large numbers for graph- and group-valued random elements; we make some observations about possible configurations of mean-sets in certain graphs that, in turn, lead to some implications about trees and free groups. This indicates that our work may also have applications in group theory. For example, we are able to conclude that if Γ is a tree and μ a probability measure on $V(\Gamma)$, then $|\mathbb{E}\mu| \leq 2$. Moreover, if $\mathbb{E}(\mu) = \{u, v\}$, then u and v are adjacent in Γ. This immediately implies that if μ is a probability distribution on a free group F, then the number of points in the mean-set of any F-valued random element ξ is at most two. In addition, we can draw some conclusions about the representation of center-sets for a free product of finitely-generated groups; namely, if G_1 and G_2 are finitely generated groups and $G = G_1 * G_2$ is the free product of G_1 and G_2, then, for any distribution μ on G, the set $\mathbb{E}\mu$ is a subset of elements of the forms gG_1 or gG_2 for some element $g \in G$. These developments indicate that our work is not restricted to the probabilistic domain only, but it goes beyond it by showing, yet another time, how intricately interconnected different areas of mathematics are.

In order to apply our results in practice, we have to be able to compute $\mathbb{S}_n$ efficiently, which may be technically hard. For that reason, we study special cases when $\mathbb{S}_n$ is easily computable and define an algorithm for computing it. We show that for concrete configurations of mean-sets in graphs and some local properties of $M_\xi(\cdot)$ we can achieve good results.

As we continue to build up our probability theory on graphs and groups, we introduce the notion of *median-set* on graphs that, as it turns out, possesses the same optimality property as medians in classical probability theory. Further developments lead us to the concept of *central order* on vertices of Γ (introduced in Definition 6.11) and, as a result, to the proposal of a more general definition of *mean-sets of class c relevant to the central order* (see Definition 6.14 below). This definition agrees with the definition of classical expectation on the real line when the weight function $M(\cdot)$ is finite, as well as with our first definition (see (1.2)) of mean-sets on graphs. However, to its great advantage, Definition 6.14 makes sense even when the weight function is infinite, due to the possibility of comparing the vertices of Γ using the binary relation of central order established in Definition 6.11. It turns out that if we use more general notion of expectation for graphs relevant to the central order, then our Strong Law of Large Numbers for graph- and group-valued random elements still holds and can be proved in a fashion similar to its original proof.

1.1.4 Practical Motivation

Any theory is worth much more if it actually finds applications to real world problems, or if it helps illuminate other areas within pure mathematics in unexpected ways. As indicated above, the theoretical motivation of our approach is rooted in the history of the subject, and on the conviction that the fundamental intuition about the law of large numbers should prevail. Our motivation is actually two-fold: theoretical and practical. Our theory can be used to analyze security (or reliability) of certain authentication protocols used in group-based cryptography. There are numerous sources that can help the reader unfamiliar with this area to get a quick grasp of the subject should he or she become interested; for example, surveys by Oded Goldreich (17) and by P. Dehornoy (10) are good available sources. One may also consult a book by A. G. Miasnikov, V. Shpilrain, and A. Ushakov on group-based cryptography (28). In addition, a good source on foundations of cryptography is (16). In a nutshell, modern cryptography is concerned with the construction of efficient schemes that are easy to operate but hard to foil. One of the major cryptographical problems is the *authentication problem*: the prover P wishes to prove his (her) identity to the verifier V via some secret information (such as password, for instance), but does not want V to discover anything about this secret. In other words, P wants to prove that he/she knows some private (secret) key without enabling an intruder (or eavesdropper, or cheating verifier) to obtain the private key. "Preservation of security" is at the heart of this problem. The existing authentication schemes (or protocols) use the so-called *zero-knowledge proofs* as a major tool for verifying the validity of secret-based actions of P, without revealing these secrets. One should intuitively think of *zero-knowledge* as another expression for "preservation of security."

In order to make it easier for the uninitiated reader to understand the idea of the analysis, we shall provide, firstly, an oversimplified example that highlights the idea of zero-knowledge protocol (the idea of this example was borrowed from the paper "How to Explain Zero-Knowledge Protocols to Your Children" published by Jean-Jacques

Quisquater and others (32)). Secondly, we shall place ourselves in the more realistic frame of a group-based cryptography application and give some insight into how this practical problem motivates our developments. More advanced reader may skip the example and proceed further.

Example 1.1. Baby-Example of Zero-Knowledge (i.e., security-preserving) authentication scheme.

This is a simple story about the prover (or seller), who knows the secret and wants to sell it, and the verifier (or buyer), who wants to buy it. Let Peter be the prover and Victor the verifier. The story is as follows. Imagine that Peter knows the secret word that can serve as a secret key to open a door in a labyrinth. The labyrinth has only one entrance and two paths (labeled $p1$ and $p2$) that meet only once, and there is a magic door connecting the paths. We make an important assumption here:

> It is impossible to break the door open without the special word.

Victor is willing to pay Peter for his secret, but not until he is sure that Peter indeed knows it. Peter refuses to reveal the secret until he receives the money. They have to come up with a security-preserving scheme by which Peter can prove that he knows the secret word without telling it to Victor.

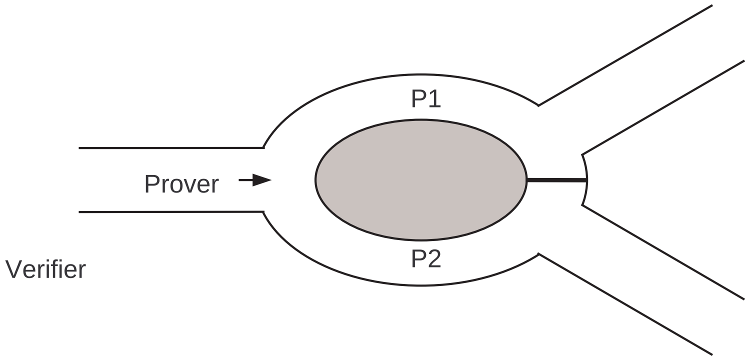

Figure 1.1: Labyrinth. Security-preserving scheme.

First, Victor waits outside of the labyrinth as Peter goes in and randomly takes either path $p1$ or $p2$; see Figure 1.1. Then Victor enters the labyrinth and shouts the name of the path he wants Peter to use to return, chosen at random. Given that Peter really does know the magic word, he opens the door (if needed) and returns along the required path. Observe that Victor does not know which path Peter had gone down initially.

If Peter is not honest and does not know the word, then he can only return by the required path if Victor gives the name of the same path Peter entered by. Since Victor chooses $p1$ or $p2$ randomly, Peter has 50 percent chance of guessing correctly. If they repeat the scheme many times, say 20 or 30, in a row, the probability of successfully anticipating all of Victor's requests gets negligibly small, and Victor should be convinced that Peter, indeed, knows the secret. In that case, the transaction will take place.

This naive example actually provides a very good intuition about what the zero-knowledge authentication protocols are. We devote the entire Chapter 7 to applications of our theory to the cryptanalysis of group-based authentication schemes, where we describe one of the existing authentication protocols, analyze it, and conclude, using our theory, that its security is questionable.

Now, let us just briefly imagine the following situation in cryptography. Let G be a group and let the Prover's private key be an element $s \in G$ and his public key a pair (w, t), where w is an arbitrary element of the group G and $t = s^{-1}ws$. In 2003, several identification protocols (authentication schemes) were proposed by Sibert, Dehornoy, and Girault in (34) (see also (10)), where the authors were claiming that the schemes enable the Verifier to check that the Prover knows the secret key, while ensuring the confidentiality of the private key, and, thus, meeting necessary security compliance. Security of the protocols is based on the complexity of certain algebraic problem(s) in G, for instance, conjugacy search problem (this complexity is an analogue of the assumption in the simple example above about the magic door being unbreakable). We present one of such schemes in detail later, in Chapter 7. For now, we just say that during the *authentication phase* of the protocol, the Prover sends to the Verifier a sequence of random elements of 2 types (depending of the value of random bit c, chosen by the Verifier): r and sr, where r is a randomly generated element and s is a secretly chosen fixed element. The process is repeated many times. Right now it is not important exactly how the scheme goes. What is important to observe, is that any eavesdropper (or intruder, or cheating verifier) can obtain two strings of elements:

$$R_1 = \{r_{i_1}, \ldots, r_{i_k}\}$$

and

$$R_2 = \{sr_{j_1}, \ldots, sr_{j_{n-k}}\}.$$

The Eavesdropper's goal is to recover the element s based on the intercepted sets above. The important assumption is:

> The random elements $r_{i_1}, \ldots, r_{i_k}$ and $r_{j_1}, \ldots, r_{j_{n-k}}$ have the same distribution, i.e., all these elements are generated by the same random generator.

To explain the idea of how we can, in fact, obtain the secret key using our theory, assume for a moment that the group G is an infinite cyclic group $\mathbf{Z}$. (Of course this is never the case in cryptography, since conjugation does not make much sense in an abelian group, but nevertheless lets make that assumption for simplicity.) In that case we can rewrite the elements of R_2 in additive notation as $\{s + r_{j_1}, \ldots, s + r_{j_{n-k}}\}$. Then we can compute the average

$$r_1 = \frac{1}{k} \sum_{m=1}^{k} r_{i_m}$$

of the elements in $R_1 \subset \mathbf{Z}$ and the average

$$r_2 = \frac{1}{n-k} \sum_{l=1}^{n-k} (s + r_{j_l}) = s + \frac{1}{n-k} \sum_{l=1}^{k} r_{j_l}$$

of the elements in $R_2 \subset \mathbf{Z}$. By the strong law of large numbers for real-valued random variables the larger the string R_1 is, the closer the value of r_1 to the mean $\mathbb{E}(\mu)$ of the distribution μ on $\mathbf{Z}$, induced by r_{i_1}. Similarly, the larger the string R_2 is, the closer the value of r_2 is to the number $s + \mathbb{E}(\mu)$. Therefore, subtracting r_1 from r_2, we can obtain a good guess of what s is, depending on sizes of the sets R_1 and R_2.

Let us observe three crucial properties that allow us to compute the secret element in the commutative case:

(SL) (Strong law of large numbers for real-valued random variables, as in (1.1).)

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i \overset{a.s.}{\to} \mathbb{E}\xi_1.$$

(LP) (Linearity property) For any real-valued random variable ξ, we have

$$\mathbb{E}(\xi + c) = \mathbb{E}(\xi) + c.$$

(EC) (Effective/efficient computations) The average value $\displaystyle\frac{1}{n}\sum_{i=1}^{n} \xi_i$ is efficiently computable.

Now it is clear what our practical motivation is, and how it affects our approach to the formulation of the generalized strong law of large numbers for graphs and groups, as in Theorem A. We generalize the notion of a sample average $\frac{1}{n}\sum_{i=1}^{n} \xi_i$ for real-valued random variables to any finitely generated group and locally finite graph, as in (1.3), so that the properties above are satisfied (with linearity property being replaced by the "shift“ property of (1.5)). With this theory at hand, the Eavesdropper breaks the scheme by computing the set

$$\mathbb{S}(sr_{j_1}, \ldots, sr_{j_{n-k}}) \cdot [\mathbb{S}(r_{i_1}, \ldots, r_{i_k})]^{-1}.$$

When n is sufficiently large, this set contains the private key s, or rather, a very good guess of what s is. We conclude that the proposed zero-knowledge authentication (security-preserving) protocol is not reliable. The detailed definition of the scheme, as well as its analysis, attack, and supporting experiments are presented in Chapter 7 in the sequel.

Remark 1.2. One of the contributions and novelties of this work is that it provides a completely new approach to the security analysis of some identification protocols existing in group-based cryptography. Such analysis is usually based on the complexity of certain algebraic problem(s) in the platform group G. The present work shows that there is a probabilistic approach to the cryptographic problem that is usually treated only from the algebraic point of view, and that this approach can be very effective.

1.2 Summary

Chapter 2 reviews some necessary graph- and group-theoretic preliminaries that constitute the setting of our work. In particular, we review the notions of a graph Γ, X-digraph, distance in a graph, and a Cayley graph of a group relative to a given generating set X. In addition, we recall definitions of a free group and free abelian group since we deal with these concepts in Chapter 5, Section 5.1, and use them to conduct experiments supporting our SLLN in Section 5.3. At the end of Chapter 2, we briefly speak about braid groups, which are often used in group-based cryptography as we shall see in Chapter 7. Next, in Chapter 3, we prepare the ground for the main result of our work by introducing the notion of (graph-) group-valued random element $\xi : \Omega \to V(\Gamma)$, weight function $v \mapsto M_\xi(v)$, and mean-set of graph-valued random element as the set of vertices in Γ that minimize the weight function M_ξ, as in (1.2). We prove a series of results relevant to the newly defined objects. In particular, we show that if $\xi : \Omega \to V(\Gamma)$ is a random element with values in a connected locally finite graph Γ with totally defined weight function $M_\xi(\cdot)$, then the mean-set $\mathbb{E}(\xi)$ is non-empty and finite. Next, we prove the so-called "shift" property (1.5) of the expectation on groups that is so useful in practical applications. Finally, in Section 3.1.2, we consider other possible definitions of $\mathbb{E}$; namely, mean-sets of class c, as in (1.4).

Next, we turn to the formulation of the strong law of large numbers for graph- and group-valued random elements. We give a careful proof of the law in the case of a singleton mean-set. We also consider cases of multi-point center-sets and generalize the law of large numbers to these situation. These tasks are carried out in Chapter 3, Section 3.2. Chapter 4 is devoted to the analogue of the Chebyshev inequality on graphs and groups.

In Chapter 5, Section 5.1, we consider configurations of center-sets in graphs. We start with the observation that it is impossible for certain combinations of vertices to comprise center-sets of some graphs. This leads to the notion of a so-called cut-point for a metric space, in general, and for a graph (Γ, d), in particular. It turns out that existence of a cut-point in Γ affects possible configurations of mean-sets dramatically, and this is a subject of a Cut-Point Lemma that we prove. The Lemma is followed by a series of Corollaries featuring applications to trees and free groups. More specifically, we prove that if Γ is a tree and μ is a probability measure on $V(\Gamma)$, then $|\mathbb{E}\mu| \leq 2$. From this, we deduce that if μ is a probability distribution on a free group F, then the number of elements in its mean-set cannot exceed two points, for any F-valued random element ξ. Moreover, the representation of a center-set for a free product of finitely generated groups becomes apparent.

Section 5.2 deals with computational problems and methods of computing of $\mathbb{S}_n = \mathbb{S}(\xi_1, \ldots, \xi_n)$. We propose an algorithm (Direct Descent) that can be used to compute the minimum of a given real-valued function $f : V(\Gamma) \to \mathbf{R}$. We show that if a function in question (weight function) satisfies certain local monotonicity properties, then we can achieve good results. In particular, we prove that our algorithm finds a central point for trees.

Further, in Section 5.3, we demonstrate how the technique of computing mean-sets, employing the Direct Descent Algorithm, works in practice. We perform series of experiments in which we compute the sample mean-sets of randomly generated samples of n random elements and observe the convergence of the sample mean-set to the actual mean. The results are presented in the tables for free and free abelian groups with a uniform distribution μ_L on a *sphere* of radius L, defined as $S_L := \{w \in F(X) \mid |w| = L\}$, in Section 5.3. From the tables, we clearly see that our experimental results support the strong law of large numbers for graphs and groups; namely, the law works in practice, indeed.

Section 6.1, in Chapter 6, is devoted to the mean-set of class one, i.e., $\mathbb{E}^{(1)}(\xi)$, as in (1.4) with $c = 1$. We interpret this set as the median-set of the distribution induced by ξ on Γ by proving Proposition 6.1 featuring the connection of $\mathbb{E}^{(1)}(\xi)$ with the classical median of a given distribution on the real line $\mathbf{R}$. Further in this section, we introduce a function $\rho^{(1)}(u,v) := \sum_{s\in\Gamma}(d(u,s) - d(v,s))\mu(s)$, for $u, v \in \Gamma$, which allows us to define a binary relation $<^{(1)}$ on the vertices of Γ by

$$u <^{(1)} v \quad \Leftrightarrow \quad \rho^{(1)}(u,v) < 0.$$

This new development permits us to eliminate the problem of dependence of median-sets on finiteness of the weight function of class one, i.e., $M^{(1)}(\cdot)$; it also allows us to define a median-set in Γ relative to μ by $\mathbb{E}^{(1)}(\mu) = \{v \in \Gamma \mid u \not<^{(1)} v, \forall u \in \Gamma\}$ which, as we prove in Proposition 6.8, is always finite and non-empty, despite the finiteness of $M^{(1)}(\cdot)$. Towards the end of this section, we remark on an optimality property of median-sets and their interpretation in terms of L^p-spaces, pointing out the fact that our median-sets on graphs correspond to L^1 settings, as well as the classical ones.

In Section 6.2, we introduce a notion of central order on Γ via a certain binary relation (denoted by $<^{(c)}$), which allows us to compare vertices sometimes; similar to $<^{(1)}$ of Section 6.1, but more general. This idea leads us to the proposal of a new, more general, definition of *mean-sets on graphs and groups relevant to the central order*, which coincides with the original definition of $\mathbb{E}$ in the case when weight function, for a given distribution μ, is finite. The advantage of the new treatment of $\mathbb{E}$ via the binary relation of central order though is that mean-sets relevant to this central order make sense even if the weight function is infinite. Moreover, it allows us to relax our assumptions in several major results of the sequel. In particular, we prove that

SLLN for graph- and group-valued random elements with singleton mean-sets relevant to $<^{(c)}$ holds with the assumption of $M^{(1)} < \infty$, as opposed to the requirement of finiteness of $M = M^{(2)}$ in Section 3.2. We continue Section 6.2 by bringing forth the connection between mean-sets relevant to $<^{(c)}$ and L^p-spaces. Finally, we conclude the theme of ordering of the vertices of Γ by giving several examples of central order, illuminating the idea of comparability of vertices, in general, and, hopefully, facilitating a better perception of the reader, in particular.

At the end of the exposition, we indicate how our theory works in practice. We provide careful analysis of an existing zero-knowledge group-based authentication scheme in cryptography. The analysis is supported by real experiments, providing an illustration of how the private element s gets revealed. We demonstrate that the analysis of a security feature of the well-known protocol leads to a rather disappointing (to the authors of the protocol) conclusion - the scheme is not reliable, and is far from secure. The secret information can be retrieved, using the strong law of large numbers for groups as the theoretical rationale of the retrieving procedure. This is the subject of the Chapter 7.

We conclude the present work with Chapter 8, where we state several problems that may be of interest to group geometry and its connection with our theory. In addition, we indicate our future goals and possible directions of research.

Chapter 2

Preliminaries on graphs and groups

Let us briefly recall some elementary notions and ideas from group and graph theory that will constitute the setting of our work. The reader may skip this chapter and consult it later if necessary. For a better insight into graph theory, the reader is referred to (40), while (26) can serve as a good introduction into group theory.

2.1 Graphs

An *undirected graph* Γ is a pair of sets (V, E) where:

- $V = V(\Gamma)$ is called the *vertex set*;

- $E = E(\Gamma)$ is a set of unordered pairs $(v_1, v_2) \in V \times V$ called the *edge set*.

Elements in V are called *vertices* and elements in E are called *edges*. If $e = (v_1, v_2) \in E$ then we say that v_1 and v_2 are *adjacent* in Γ. The set of all vertices adjacent to v_1 in Γ is denoted by $Adj(v_1)$. The number of vertices adjacent to v is called the *degree* of v and is denoted by $deg(v)$. We say that the graph Γ is *locally finite* if every vertex has finite degree (see (41)).

A *directed graph* Γ is a pair of sets (V, E) where $E = E(\Gamma)$ is a set of ordered pairs $(v_1, v_2) \in V \times V$. If $e = (v_1, v_2) \in E$ then we say that v_1 is the *origin* of the edge e

and v_2 is the *terminus* of e. For an edge $e = (v_1, v_2) \in E$, we denote by $o(e)$ its origin v_1 and by $t(e)$ its terminus v_2. Observe that any undirected graph can be viewed as a directed graph in which a pair $(v_1, v_2) \in E$ serves as two edges (v_1, v_2) and (v_2, v_1).

A *path* p in a directed graph Γ is a finite sequence of edges $e_1, \ldots, e_n$ such that $t(e_i) = o(e_{i+1})$. The vertex $o(e_1)$ is called the *origin* of the path p and is denoted by $o(p)$. The vertex $t(e_n)$ is called the *terminus* of the path p and is denoted by $t(p)$. The number n is called the *length* of the path p and is denoted by $|p|$. We say that two vertices $v_1, v_2 \in V(\Gamma)$ are *connected*, if there exists a path from v_1 to v_2 in Γ. The graph Γ is *connected* if every pair of vertices is connected.

The *distance* between v_1 and v_2 in a graph Γ is the length of a shortest path between v_1 and v_2; if v_1 and v_2 are disconnected, then we say that the distance is infinite. The distance between v_1 and v_2 is denoted by $d(v_1, v_2)$. We say that a path $p = e_1, \ldots, e_n$ from v_1 to v_2 is *geodesic* in a graph Γ if $d(o(p), t(p)) = d(v_1, v_2) = n$, i.e., if p is a shortest path from v_1 to v_2.

A path $p = e_1, \ldots, e_n$ in a graph Γ is *closed*, if $o(p) = t(p)$. In this event we say that p is a *cycle* in Γ. A path p is *simple*, if no proper segment of p is a cycle. The graph Γ is a *tree* if it does not contain a simple cycle.

2.2 Cayley graphs and groups

Consider a finite set, also called *alphabet*, $X = \{x_1, \ldots, x_n\}$, and let X^{-1} be the set of formal inverses $\{x_1^{-1}, \ldots, x_n^{-1}\}$ of elements in X. This defines an involution ${}^{-1}$ on the set $X^{\pm 1} := X \cup X^{-1}$ which maps every symbol $x \in X$ to its formal inverse $x^{-1} \in X^{-1}$ and every symbol $x^{-1} \in X^{-1}$ to the original $x \in X$. An alphabet X is called a *group alphabet* if X contains x^{-1} corresponding to each $x \in X$, i.e., $X^{-1} \subseteq X$, and there is an involution which maps elements of X to their inverses. An X-*digraph* is a graph (V, E) with edges labeled by elements in $X^{\pm 1} = X \cup X^{-1}$ such that for any edge $e = u \xrightarrow{x} v$ there exists an edge $v \xrightarrow{x^{-1}} u$, which is called the inverse of e and is denoted by e^{-1}. Let Γ and Δ be X-digraphs. A mapping $\varphi : \Gamma \to \Delta$ is called an *X-digraph morphism*, if it preserves connectedness and labeling on edges.

We say that an X-digraph is *non-folded*, if it contains a pair of distinct edges $e_1 = u \xrightarrow{x} v_1$ and $e_2 = u \xrightarrow{x} v_2$ equally labeled and having the same origin. Otherwise, an X-digraph is called *folded*.

Now we recall some definitions from the theory of groups. Let G be a set equipped with a binary operation $\cdot : G \times G \to G$. A pair $(G, \cdot)$ is called a *group* if the following conditions are satisfied:

(G1) The operation $\cdot$ is *associative*, i.e., for any $a, b, c \in G$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

(G2) G contains an element e such that for any $g \in G$, $g \cdot e = e \cdot g = g$. Such an element e is called the *identity* of the group G.

(G3) For every $g \in G$, there exists an element $h \in G$ such that $g \cdot h = h \cdot g = e$. In this event the element h is called the *inverse* of g and is denoted by g^{-1}.

Since $\cdot$ is the only operation defined on a group, the sign $\cdot$ is usually omitted in notation. Let $X \subset G$ be a set of generators for G, i.e. $G = \langle X \rangle$. Assume that X is closed under inversion, i.e., $X = X^{\pm 1}$. The *Cayley graph* $C_G(X)$ of G relative to X is a labeled graph (V, E), where the vertex set is $V = G$, and the edge set E contains only edges of the form $g_1 \xrightarrow{x} g_2$ where $g_1, g_2 \in G$, $x \in X$ and $g_2 = g_1 x$. The *distance* between elements $g_1, g_2 \in G$ relative to the generating set X is the distance in the graph $C_G(X)$ between the vertices g_1 and g_2 or, equivalently,

$$d_X(g_1, g_2) = \min\{n \mid g_1 x_1^{\varepsilon_1} x_2^{\varepsilon_2} \ldots x_n^{\varepsilon_n} = g_2 \text{ for some } x_i \in X, \varepsilon_i = \pm 1, i = 1, \ldots, n\}$$

Notice that if we change the generating set X, the distance d_X changes. Nevertheless, if X is fixed, we write $d(\cdot, \cdot)$ instead of $d_X(\cdot, \cdot)$. When we work with groups and Cayley graphs related to them, we always assume that X is fixed.

Lemma 2.1. *Let X be a group alphabet and $\Gamma = (V, E)$ be a folded X-digraph. The graph Γ is the Cayley graph of some group G, if and only if*

(1) Γ *is connected, and*

(2) for any two vertices $u, v \in V$ *there exists an* X*-digraph isomorphism* $\varphi : \Gamma \to \Gamma$ *such that* $\varphi(u) = v$*.*

Proof. Clearly, if an X-digraph Γ is a Cayley graph then it satisfies both properties.

Conversely, assume that Γ satisfies properties (1) and (2). We directly construct a group G and choose a generating set $X \subset G$ such that Γ is the Cayley graph of G relative to X. Put $G = V(\Gamma)$. To define multiplication on $V(\Gamma)$ fix some vertex $v_0 \in \Gamma$. For each $v \in \Gamma$ choose any path from v_0 to v and denote its label by w_v. Now, for $u, v \in V(\Gamma)$ put $u \cdot v$ to be the endpoint of the path starting at v_0 labeled with $w_u \circ w_v$. It is straightforward to check that $V(\Gamma)$ with so defined product satisfies properties of a group. Furthermore, choose as generators the endpoints of the paths starting at v_0 labeled by the letters X. Clearly, Γ is the Cayley graph of G relative to the chosen generating set.

□

Let G_1 and G_2 be groups. We define a new group called the direct or Cartesian product of G_1 and G_2, denoted by $G_1 \times G_2$, as follows. The group $G_1 \times G_2$ is the set of all pairs (g_1, g_2) where $g_1 \in G_1$ and $g_2 \in G_2$ with multiplication defined coordinatewise. If $X_1 \subset G_1$ generates the group G_1 and $X_2 \subset G_2$ generates the group G_2, then the set

$$X = \{(g_1, e_2) \mid g_1 \in X_1\} \cup \{(e_1, g_2) \mid g_2 \in X_2\}$$

generates the group $G_1 \times G_2$ and the length function on $G_1 \times G_2$ is usually defined relative to X.

2.3 Free groups

Since some of the results in the sequel are concerned with trees, which are Cayley graphs of free groups, we collect basic definitions and notions about this specific kind of group in a separate subsection for the convenience of the reader. Recall that, abstractly, a group G is called *free* if it contains a subset X such that any non-trivial product of elements from X is a non-trivial element in G. In this event, we say that X is a *free basis* for G. The cardinality $|X|$ of X is called the rank of G. It is a fact from group theory that any two free groups of equal ranks are isomorphic and can be considered the same. Therefore, a free group of rank $n \in \mathbb{N}$ is denoted by F_n. We work with free groups of finite ranks only.

The definition of a free group given above is one of the several possible abstract definitions. The efficient realization of F_n is as follows. Let X be a set of distinct symbols. We can consider the corresponding group alphabet $X^{\pm 1}$ which contains X and the formal inverses of the elements in X. Consider the set of all words over $X^{\pm 1}$. We say that a word w is *reduced*, if it does not contain a subword of the type xx^{-1} or $x^{-1}x$ for some $x \in X$. The process of removal of subwords of the type xx^{-1} and $x^{-1}x$ is called *reduction*. It can be shown that any sequence of removing xx^{-1} and $x^{-1}x$ in a non-reduced word w always results in the same word, which we denote by $\overline{w}$.

The set of all reduced words is denoted by $F(X)$. We can define multiplication on $F(X)$ as follows. For $u, v \in F(X)$ we define

$$u \cdot v = \overline{u \circ v}$$

where $u \circ v$ is the concatenation of u and v. It is easy to check that the so-defined pair $(F(X), \cdot)$ is a group; moreover, it is a free group of rank $|X|$, and X is a free basis for $F(X)$.

Since every element in $F(X)$ has a unique representation as a reduced word over $X^{\pm 1}$, it follows that the length $d_X(w)$ of an element

$$w = x_1^{\varepsilon_1} \ldots x_n^{\varepsilon_n} \in F(X),$$

where $\varepsilon_i = \pm 1$, is the total number of symbols involved in writing of w which is n.

The Cayley graph of the free group of rank n relative to its free basis is a regular infinite tree in which every vertex has degree $2n$. For example, Figure 2.1 depicts

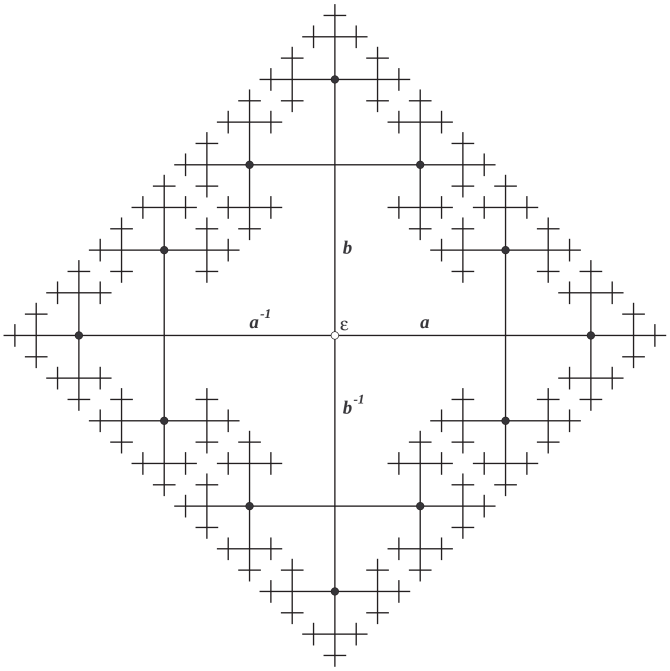

Figure 2.1: The Cayley graph of a free group F_2

the Cayley graph of a free group $F_2 =< a, b >$. Black dots on the picture indicate all elements of length 2 (all words of length 2); they comprise the so-called sphere or radius 2, S_2, which we will encounter later, in Chapter 5.

2.4 Free abelian groups

Another important class of groups is that of free abelian groups. Recall that a group G is called *abelian* if for any two elements $a, b \in G$ the equality $ab = ba$ holds. An abelian group G is called a free abelian group with *free abelian basis* $X \in G$ if for any abelian group H and any mapping $\varphi : X \to H$ there exists a homomorphism φ^* such that the diagram below commutes.

$$\begin{tikzcd}
X \arrow[r, hook, "\varphi"] \arrow[dr, "\varphi"'] & G \arrow[d, "\varphi^*"] \\
& H
\end{tikzcd}$$

The defined property is called the universal property of the free abelian groups. The cardinality of the subset X is called the rank of the group G and it is easy to show that any two free abelian groups of equal ranks are isomorphic. In fact, the free abelian group of rank n is isomorphic to the direct product of n infinite cyclic groups

$$A_n = \mathbf{Z} \times \mathbf{Z} \times \ldots \times \mathbf{Z}.$$

Hence, we may think of the elements of A_n as n-tuples of integers, with the binary operation denoted by + given by

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n).$$

The standard generating set X for $\mathbf{Z}^n$ consists of n tuples $e_1, \ldots, e_n$ where each $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ contains 0's everywhere except the ith position, where it contains 1. The length of an element $a = (a_1, \ldots, a_n)$ relative to X is given by

$$|a| = |a|_X = \sum_{i=1}^{n} |a_i|.$$

The Cayley graph of a free abelian group of rank n is an infinite n-dimensional grid.

2.5 A group of braids and its presentation

Practical applications of our theory are concerned with cryptanalysis of group-based protocols that employ braid groups as their platform. We collect some preliminaries on these groups in this section, following the exposition of (11) and (28). To make it clear, we choose to give an intuitive view of this group, illuminating it with many pictures.

A braid is obtained by laying down a number of parallel pieces of string and intertwining them, without loosing track of the fact that they run essentially in the same direction. In our pictures the direction is horizontal. We number strands at each horizontal position from the top down. See Figure 2.2 for example.

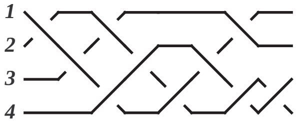

Figure 2.2: A 4-strand braid.

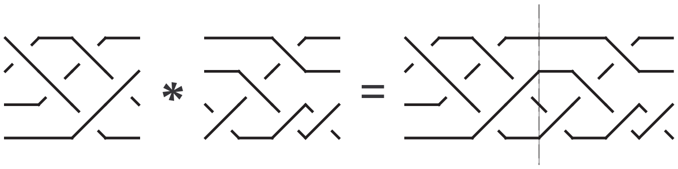

Figure 2.3: Product of braids.

If we put down two braids u and v in a row so that the end of u matches the beginning of v we get another braid denoted by uv, i.e., concatenation of n-strand braids is a product (see Figure 2.3).

We consider two braids equivalent if there exists an isotopy between them, i.e., if it is possible to move the strands of one of the braids in space (without moving the endpoints of strands and moving strands through each other) to get the other braid. See Figure 2.4 to visualize it. We distinguish a special n-strand braid which contains no crossings and call it a trivial braid (Figure 2.5).

Clearly the trivial braid behaves as left and right identity relative to the defined multiplication. The set B_n of isotopy classes of n-strand braids has a group structure because if we concatenate a braid with its mirror image in a vertical plane the result

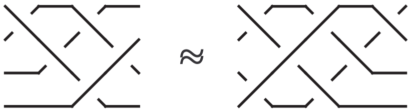

Figure 2.4: Isotopy of braids.

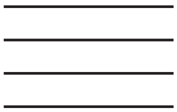

Figure 2.5: A trivial 4-strand braid.

is isotopic to the trivial braid. See Figure 2.7 to visualize an inversion of a 4-strand braid.

Basically, each braid is a sequence of strand crossings. A crossing is called positive if the front strand has a negative slope, otherwise it is called negative. There are exactly $n-1$ crossing types for n-strand braids, we denote them by $\sigma_1, \ldots, \sigma_{n-1}$, where σ_i is a positive crossing of ith and $(i+1)$st strands. See Figure 2.6 for an example for B_4.

Since, as we mentioned above, any braid is a sequence of crossings the set $\{\sigma_1, \ldots, \sigma_{n-1}\}$ generates B_n. It is easy to see that crossings $\sigma_1, \ldots, \sigma_{n-1}$ are subject to the relations

$$[\sigma_i, \sigma_j] = 1$$

for every i, j such that $|i-j| > 1$ and

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$$

for every i such that $1 \leq i \leq n-2$. The corresponding braid configurations are shown in Figure 2.8.

It is more difficult to prove that these two types of relations actually describe the equivalence on braids, i.e., the braid group B_n has the following (Artin) presentation

$$B_n = \left\langle \ \sigma_1, \ldots, \sigma_{n-1} \ \middle| \ \begin{array}{ll} \sigma_i \sigma_j \sigma_i = \sigma_j \sigma_i \sigma_j & \text{if } |i-j| = 1 \\ \sigma_i \sigma_j = \sigma_j \sigma_i & \text{if } |i-j| > 1 \end{array} \right\rangle.$$

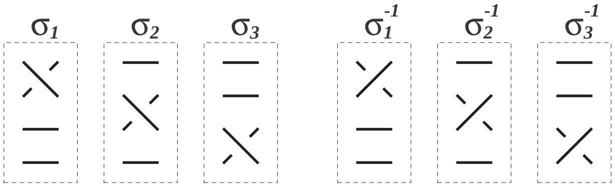

Figure 2.6: Generators of B_4 and their inverses.

From this description, one can see that there are many pairs of commuting subgroups in B_n, which makes it possible to use B_n as the platform group for cryptographic protocols such as Ko, Lee et al. (25).

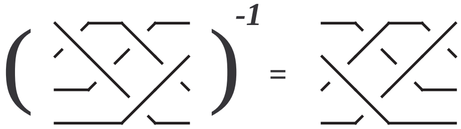

Figure 2.7: Inversion of a 4-strand braid.

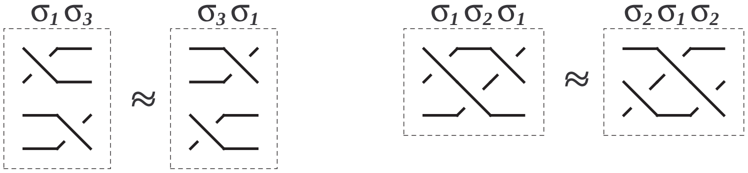

Figure 2.8: Typical relations in braids.

Chapter 3

Mean-set and the strong law of large numbers – I.

3.1 Mean of a group-valued random element

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $\xi : \Omega \to G$ a random variable taking values in the group G. The goal of this section is to introduce a G-valued functional $\mathbb{E}$ on the space of random variables $\xi : \Omega \to G$ satisfying the property

$$\mathbb{E}(g\xi) = g\mathbb{E}(\xi),$$

which would lead to practical applications.

In addition to that, the required definition of $\mathbb{E}$ must incorporate the geometric structure of the group, or, more precisely, the structure of its Cayley graph. In Subsection 3.1.1 we propose a definition of $\mathbb{E}$ for graph-valued random elements. The same definition will hold for random elements on finitely-generated groups, because every such group is associated with its *Cayley* graph.

Recall that if $\xi_1, \xi_2, \ldots$ are i.i.d. L^2 real-valued random variables, then a quadratic function $\mathbb{E}[(\xi_1 - c)^2]$ achieves its minimum at $c = \mathbb{E}(\xi)$. We can think about the sample mean c_n in a similar way, considering $\sum_{i=1}^{n}[\xi_i - c_n]^2$. We specialize this situation to graphs and groups, make it precise, prove the strong law of large numbers, and support it by the actual experiments with free groups. This goal is reached in several steps.

3.1.1 The mean set in a graph

Consider a measurable mapping ξ from $(\Omega, \mathcal{F})$ to a discrete measurable space $(V, \mathcal{S})$. This mapping is called a *random element* in V defined on a given probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We can think of V as the state- (or phase-) space of the element. Let $\Gamma = (V(\Gamma), E(\Gamma))$ be a graph and consider a random element $\xi : \Omega \to V(\Gamma)$ with values in the set of vertices of Γ. This ξ induces the atomic probability measure μ on $V(\Gamma)$ defined by

$$\mu(g) = \mathbf{P}\Big(\{\omega \in \Omega \mid \xi(\omega) = g\}\Big) \tag{3.1}$$

for each $g \in V(\Gamma)$. If we want to emphasize the random element ξ, we write $\mu_\xi(g)$; otherwise, we suppress the subindex. Naturally, dealing with a random element $\xi : \Omega \to V(\Gamma)$, we may (and often will) work on the induced probability space $(V(\Gamma), \mathcal{S}, \mu)$. For each vertex $v \in V(\Gamma)$ we define the value

$$M_\xi(v) := \sum_{i \in V(\Gamma)} d^2(v, i)\mu(i) \tag{3.2}$$

called the *weight* of v relative to the measure μ, where $d(v, i)$ is the distance between v and i in Γ. If $M_\xi(v)$ is finite, we say that the *weight function* $M_\xi : V(\Gamma) \to [0, \infty]$ is defined at v. Observe that the *weight function* is not always finite. The case of interest of course, is when $M(v)$ is *totally defined*, meaning that $M(v)$ is finite on the whole set $V(\Gamma)$.

Definition 3.1. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $\xi : \Omega \to V(\Gamma)$ a random element. We denote the set of vertices in Γ that minimize the weight function $v \mapsto M_\xi(v)$ by

$$\mathbb{E}(\xi) := \{v \in V(\Gamma) \mid M_\xi(v) \le M_\xi(u), \quad \forall u \in V(\Gamma)\}, \tag{3.3}$$

and refer to it as the *mean-set* (or the *center-set*) of the random element $\xi : \Omega \to V(\Gamma)$.

Remark 3.2. Very often we leave the random element ξ in the background to shorten the notation and write $M(v)$ instead of $M_\xi(v)$ if this causes no confusion. Moreover, we may write $\mathbb{E}(\mu)$ instead of $\mathbb{E}(\xi)$ sometimes and speak of the mean set of distribution μ induced by ξ on $V(\Gamma)$.

Lemma 3.3. *Let Γ be a connected graph and u, v be adjacent vertices in Γ. If the value $M(u)$ is defined, then $M(v)$ is defined.*

Proof. Since $d(u, v) = 1$, by the triangle inequality for $v, u, i \in V(\Gamma)$ we have $d(v, i) \leq d(u, i) + d(u, v) = d(u, i) + 1$ and

$$M(v) = \sum_{i \in V(\Gamma)} d^2(v,i)\mu(i) \leq \sum_{i \in V(\Gamma)} [d(u,i)+1]^2 \mu(i) \leq$$

$$\leq \sum_{i \in V(\Gamma),\ d(u,i)=0} \mu(i) + \sum_{i \in V(\Gamma),\ d(u,i)\geq 1} [d(u,i)+1]^2 \mu(i) \leq$$

$$\leq 1 + 4 \sum_{i \in V(\Gamma),\ d(u,i)\geq 1} d^2(u,i)\mu(i) = 1 + 4M(u).$$

□

Lemma 3.3 immediately implies that the weight function on a connected graph Γ is either defined on the whole set $V(\Gamma)$ or undefined on the whole set $V(\Gamma)$; we formulate this below as a corollary.

Corollary 3.4. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, Γ a connected graph, and $\xi : \Omega \to V(\Gamma)$ a random element in Γ. Then either domain$(M) = V(\Gamma)$ or domain$(M) = \emptyset$.*

Lemma 3.5. *Let $\xi : \Omega \to V(\Gamma)$ be a Γ-valued random variable, where Γ is a connected locally finite graph, with totally defined weight function $M_\xi(\cdot)$. Then the mean set $\mathbb{E}(\xi)$ is non-empty and finite.*

Proof. Let μ be a measure of (3.1) induced on Γ by ξ. For an arbitrary but fixed vertex $v \in \Gamma$, the weight function

$$M(v) = \sum_{i \in V(\Gamma)} d^2(v,i)\mu(i) = \sum_{n=0}^{\infty} \left(n^2 \sum_{i \in V(\Gamma), d(v,i)=n} \mu(i) \right)$$

is defined at v by assumption. Choose $r \in \mathbb{N}$ such that

$$\frac{1}{2}M(v) \leq \sum_{n=0}^{r} \left(n^2 \sum_{i \in V(\Gamma), d(v,i)=n} \mu(i) \right) = \sum_{i \in B_v(r)} d^2(v,i)\mu(i),$$

where

$$B_v(r) := \{i \in V(\Gamma) \mid d(v,i) \leq r\} \tag{3.4}$$

is the *ball* in Γ of radius r centered at v.

If we take a vertex u such that $d(u,v) \geq 3r$, then using the triangle inequality we obtain the following lower bound:

$$M(u) = \sum_{i \in V(\Gamma)} d^2(u,i)\mu(i) = \sum_{i \in B_v(r)} d^2(u,i)\mu(i) + \sum_{i \notin B_v(r)} d^2(u,i)\mu(i) \geq$$

$$\geq \sum_{i \in B_v(r)} [2r]^2 \mu(i) + \sum_{i \notin B_v(r)} d^2(u,i)\mu(i) \geq \sum_{i \in B_v(r)} [2r]^2 \mu(i) \geq 4 \sum_{i \in B_v(r)} d^2(v,i)\mu(i) \geq 2M(v).$$

Thus, $d(v,u) \geq 3r$ implies $u \notin \mathbb{E}(\xi)$ and, hence, $\mathbb{E}(\xi) \subseteq B_v(3r)$. Since the graph Γ is locally finite, it follows that the sets $B_v(3r)$ and $\mathbb{E}(\xi)$ are finite. This implies that the function M attains its minimal value in $B_v(3r)$ and hence $\mathbb{E}(\xi) \neq \emptyset$.

□

Let G be a group and $X \subseteq G$ its generating set. It follows from the definition of the distance d_X that for any $a, b, s \in G$ the equality

$$d_X(a,b) = d_X(sa, sb) \tag{3.5}$$

holds. This equality implies that $\mathbb{E}(\xi)$ possesses the desirable property $\mathbb{E}(g\xi) = g\mathbb{E}(\xi)$, as the following proposition shows.

Proposition 3.6. (“Shift” Property) Let G be a group and $g \in G$. Suppose that $(\Omega, \mathcal{F}, \mathbf{P})$ is a given probability space and $\xi : \Omega \to G$ a G-valued random element on Ω. Then for the random element ξ_g defined by $\xi_g(\omega) := g\xi(\omega)$ we have

$$\mathbb{E}(\xi_g) = g\mathbb{E}(\xi).$$

Proof. Let μ_{ξ_g} be the measure induced on G by ξ_g, in the manner of (3.1). It follows from the definition of ξ_g that for any $h \in G$

$$\mu_{\xi_g}(h) = \mathbf{P}\Big(\{\omega \mid \xi_g(\omega) = h\}\Big) = \mathbf{P}\Big(\{\omega \mid g\xi(\omega) = h\}\Big) =$$

$$= \mathbf{P}\Big(\{\omega \mid \xi(\omega) = g^{-1}h\}\Big) = \mu_\xi(g^{-1}h).$$

This, together with (3.5), implies that for any $h \in G$

$$M_{\xi_g}(h) = \sum_{i \in G} d^2(h,i)\mu_{\xi_g}(i) = \sum_{i \in G} d^2(g^{-1}h, g^{-1}i)\mu_\xi(g^{-1}i) = \sum_{i \in G} d^2(g^{-1}h, i)\mu_\xi(i) = M_\xi(g^{-1}h).$$

Hence, the equality $M_{\xi_g}(h) = M_\xi(g^{-1}h)$ holds for any random variable ξ and elements $g, h \in G$. Therefore, for any $h_1, h_2 \in G$

$$M_{\xi_g}(h_1) < M_{\xi_g}(h_2) \Leftrightarrow M_\xi(g^{-1}h_1) < M_\xi(g^{-1}h_2)$$

and

$$\mathbb{E}(\xi_g) = \Big\{h \in G \mid M_{\xi_g}(h) \le M_{\xi_g}(f), \ \ \forall f \in G\Big\} = \Big\{h \in G \mid M_\xi(g^{-1}h) \le M_\xi(g^{-1}f), \ \ \forall f \in G\Big\} =$$

$$= \Big\{h \in G \mid M_\xi(g^{-1}h) \le M_\xi(f), \ \ \forall f \in G\Big\} = \Big\{gh \in G \mid M_\xi(h) \le M_\xi(f), \ \ \forall f \in G\Big\} = g\mathbb{E}(\xi).$$

□

The equality $d_X(a,b) = d_X(as,bs)$ does not hold for a general group $G = \langle X \rangle$. It holds in abelian groups. As a corollary we have the following result, proved in essentially the same way as Proposition 3.6.

Proposition 3.7. Let G be an abelian group and $g \in G$. Suppose that $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space and $\xi : \Omega \to G$ a G-valued random element on Ω. Then for the random element ξ_g defined by $\xi_g(\omega) := \xi(\omega)g$ we have

$$\mathbb{E}(\xi_g) = (\mathbb{E}(\xi))g.$$

3.1.2 Other possible definitions of $\mathbb{E}$

There are many other possible definitions of $\mathbb{E}$ for which the statement of Proposition 3.6 holds. Let c be a positive integer. By analogy to the function $M_\xi(v)$, define a *weight function* $M_\xi^{(c)}(v)$ of class c by

$$M_\xi^{(c)}(v) := \sum_{i \in V(\Gamma)} d^c(v,i)\mu(i).$$

Definition 3.8. We define the mean-set $\mathbb{E}^{(c)}(\xi)$ of class c by

$$\mathbb{E}^{(c)}(\xi) := \{v \in V(\Gamma) \mid M^{(c)}(v) \leq M^{(c)}(u), \quad \forall u \in V(\Gamma)\}.$$

The weight function $M_\xi(\cdot)$ and the mean-set $\mathbb{E}(\xi)$ are special cases of $M_\xi^{(c)}(\cdot)$ and $\mathbb{E}^{(c)}(\xi)$, namely,

$$M_\xi = M_\xi^{(2)} \text{ and } \mathbb{E} = \mathbb{E}^{(2)}.$$

It is straightforward to check that all the statements of the previous section hold for $M_\xi^{(c)}(\cdot)$ and $\mathbb{E}^{(c)}(\xi)$. All proofs work with minimal modifications. Nevertheless, we choose to work with $\mathbb{E} = \mathbb{E}^{(2)}$, which is in better agreement with the classical case. It is easy to observe that our definition of $\mathbb{E}$ agrees with the classical definition of the expectation on $\mathbf{R}$, in the following sense.

Proposition 3.9. Let $\xi : \Omega \to \mathbf{Z}$ be an integer-valued random variable with classical expectation

$$\mathfrak{m} = \sum_{n \in \mathbf{Z}} n\mathbf{P}(\xi = n).$$

Assume that $M \equiv M_\xi^{(2)}$ is defined on $\mathbf{Z}$. Then $1 \leq |\mathbb{E}^{(2)}(\xi)| \leq 2$ and for any $v \in \mathbb{E}^{(2)}(\xi)$, we have $|\mathfrak{m} - v| \leq \frac{1}{2}$.

Proof. We can naturally extend the function $M^{(2)}(\cdot)$ from $\mathbf{Z}$ to $\mathbf{R}$ by defining $M^{(2)}(v) = \sum_{n \in \mathbf{Z}} |v-n|^2 \mathbf{P}(\xi = n)$ for any $v \in \mathbf{R}$. For the function $M^{(2)} : \mathbf{R} \to [0, \infty)$, we have:

$$[M^{(2)}(v)]' = 2\sum_{n \in \mathbf{Z}} (v - n)\mathbf{P}(\xi = n).$$

It is easy to check that the function $M^{(2)}(v)$ attains its minimum at

$$\frac{\sum_{n\in\mathbf{Z}} n\mathbf{P}(\xi = n)}{\sum_{n\in\mathbf{Z}} \mathbf{P}(\xi = n)} = \sum_{n\in\mathbf{Z}} n\mathbf{P}(\xi = n) = \mathfrak{m}$$

and is strictly decreasing on the interval $(-\infty, \mathfrak{m})$ and strictly increasing on the interval $(\mathfrak{m}, \infty)$. Now, we recall that

$$\mathbb{E}^{(2)}(\xi) = \{v \in \mathbf{Z} \mid M^{(2)}(v) \le M^{(2)}(u), \quad \forall u \in \mathbf{Z}\},$$

and it becomes clear that $\mathbb{E}^{(2)}(\xi)$ can contain the integers $\lfloor \mathfrak{m} \rfloor$ and $\lceil \mathfrak{m} \rceil$ only. Finally, since $M^{(2)}(v)$ is a quadratic function, it follows that it is symmetric relative to its vertex point $\mathfrak{m}$. Therefore, $M^{(2)}(v)$ restricted to $\mathbf{Z}$ attains its minimum at the integer points closest to $\mathfrak{m}$. In other words, $\lfloor m \rfloor \in \mathbb{E}^{(2)}(\xi)$ if and only if $\mathfrak{m} - \lfloor \mathfrak{m} \rfloor \le 1/2$; and $\lceil \mathfrak{m} \rceil \in \mathbb{E}^{(2)}(\xi)$ if and only if $\lceil \mathfrak{m} \rceil - \mathfrak{m} \le 1/2$. Hence the result.

□

Remark 3.10. Unfortunately, $\mathbb{E}^{(2)}$ does not coincide with the classical mean in $\mathbf{R}^2$. Recall that the classical mean in $\mathbf{R}^2$ is defined coordinate-wise, i.e., the mean of $(x_1, y_1), \ldots, (x_n, y_n)$ is a point in $\mathbf{R}^2$ defined by

$$(\mathbb{E}X, \mathbb{E}Y).$$

Now consider the distribution on $\mathbf{Z}^2$ such that $\mu(0,0) = \mu(0,3) = \mu(3,0) = 1/3$ and for all other points $\mu = 0$. Then the classical mean defined by the formula above is the point $(1,1)$ and the mean $\mathbb{E}^{(2)}$ is the point $(0,0)$. See Figure 3.1 (each vertex marked by gray has probability $1/3$, others have probability 0. The classical mean is at the point v, while the mean defined by $\mathbb{E}^{(2)}$ is at the point O).

At last, it is worth noticing that, in some cases, $\mathbb{E}^{(1)}$ contradicts our intuition of where the average should be:

- Let μ be a distribution on $\mathbf{R}$ such that $\mu(-1) = \mu(1) = 1/2$. Then $\mathbb{E}^{(1)}(\mu) = [-1, 1]$ the whole interval between -1 and 1, which is counterintuitive.

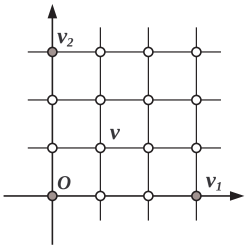

Figure 3.1: $\mathbb{E}^{(2)}$ does not coincide with the classical mean in $\mathbf{R}^2$.

- Let $0 < \varepsilon < 1/2$ and $\mu(-1) = 1/2 - \varepsilon$ and $\mu(1) = 1/2 + \varepsilon$. Then $\mathbb{E}^{(1)}(\mu) = \{1\}$. This means that even very small perturbations of the distribution change the mean-set dramatically.

In spite of the fact that we refrain from employing $\mathbb{E}^{(c)}$ of class $c = 1$ as a mean-set in our theory, $\mathbb{E}^{(1)}$ deserves special attention and will be treated separately in Section 6.1 below, where we interpret it as a median-set of Γ.

3.1.3 Centers of Rubinshtein vs. mean-sets (center-sets).

It turns out that class one is a particular class of mean-sets that have already found its use in probabilistic literature as a possible notion of a center of a distribution. Let us briefly indicate what it was about. The notion of measure, and, in particular, of normalized measure, lies at the foundation of the whole probability theory. Availability of various characteristics of such measures on different spaces facilitates the productiveness of the ongoing research in theoretical and applied probability. We introduced the notion of the mean-set of a (graph-)group-valued random element ξ – one of the possible discrete characteristics of probability measure on state spaces, such as graphs or groups. In the process of our research, we discovered that one of the well-known mathematicians, G. Sh. Rubinstein (a student and a co-author of the outstanding scientist of the 20th century, L. V. Kantorovich) was also interested in the discrete characteristics of normalized measures, but, of course, with the different goals in mind. In 1995, in his article ”On Multiple-point centers of normalized measures on locally compact metric spaces“ (see (33)), G. Sh. Rubinstein introduces the notion of one-point centers and multiple-point centers, which constitute the families of discrete characteristics of distributions on the function space $\Phi(E)$. This is the space of probability distributions on an arbitrary unbounded locally compact metric space $E = (E, \rho)$ with the corresponding σ-algebra of Borel sets $\mathcal{B}$. In other words, $\Phi(E)$ consists of non-negative normalized σ-additive functions φ, defined on $\mathcal{B}$ with the finite first integral moments

$$m_1(u, \varphi) = \int_E \rho(x, u)\varphi(de_x), \ \forall u \in E.$$

Rubinshtein's one-point centers $u^* \in E$ are characterized by the fact that the first moments $m_1(u^*, \varphi)$ corresponding to them are minimal, i.e., the moments coincide with

$$\mu_1^* = \inf_{u \in E} m_1(u, \varphi).$$

These one-centers are just a special case of Rubinshtein's multiple-point centers that are defined using a certain partition of the original metric space $E = (E, \rho)$ (see (33) for details). While one-centers of Rubinshtein correspond to our mean-sets of class one, namely, to $\mathbb{E}^{(1)}(\mu)$, his k-centers of (33) have no analogues in the present work. As indicated in Subsection 3.1.2, $\mathbb{E}^{(1)}(\mu)$, as a possible candidate for the mean-set on graphs, does not suit our purposes, even though it is easier to work with it, especially when it comes to computations. On the contrary, choosing to minimize the first integral moment is in accord with the settings of Rubinshtein's article (33) that relates his k-centers directly to the generalizations of the well-known studies of Kantorovich, Monge, and Appell on optimal volume-preserving transportation of masses over metric compacta. In short, G. Sh. Rubinshtein characterizes his k-centers in terms of the famous Monge-Appell-Kantorovich metric (or just ”transportation metric“ – the optimal value of the objective functional in the mass transportation problem). The reader is referred to (33), (21), (22), and Chapter VIII of (20) for more details, while we feel especially appreciative of the opportunity to pay a tribute of respect to the scientific legacy of L. V. Kantorovich and his famous transportation problem. The reason for introducing of Rubinshtein's k-centers goes back to the general theory of the famous metric. The theory was presented in 1958, in the joint article of L. V. Kantorovich and G. Sh. Rubinshtein (see (22)), which shows how the transportation metric can be used for introducing a norm in the space of measures, and how Kantorovich's optimality criterion becomes a theorem on the duality of the space of measures with the Kantorovich metric and the space of Lipschitz functions. Before 1958, it was not known whether the space of Lipschitz functions is conjugate to any Banach space. The interested reader may consult the above mentioned sources plus recent (2005, 2007) surveys on history of the above metric by A. Vershik ((39), (38)).

The purpose of this digression to the domain of transportation problem is to indicate how similar, in some sense, notions and ideas can serve completely different purposes. The settings in which Rubinshtein's centers were introduced and the goals that were pursued clearly diverge from ours. We deal with the (graph-)group-valued random elements and prove the generalization of the Strong Law of Large Numbers for such elements. Our mean-set (or center-set) is the expectation that we use in the formulation of the law and other results. Our findings lead to some unexpected corollaries and applications, while, at the same time, interconnecting different areas of mathematics (probability theory, graph theory, group theory), cryptanalysis and theoretical computer science.

3.2 Strong Law of Large Numbers (SLLN)

Let $\xi_1, \ldots, \xi_n$ be a sample of independent and identically distributed graph-valued random elements $\xi_i : \Omega \to V(\Gamma)$ defined on a given probability space $(\Omega, \mathcal{F}, \mathbf{P})$. For every $\omega \in \Omega$, let $\mu_n(u;\omega)$ be the relative frequency

$$\mu_n(u;\omega) := \frac{|\{i \mid \xi_i(\omega) = u, \quad 1 \le i \le n\}|}{n} \tag{3.6}$$

with which the value $u \in V(\Gamma)$ occurs in the random sample $\xi_1(\omega), \ldots, \xi_n(\omega)$. We shall suppress the argument $\omega \in \Omega$ to ease notation, and let

$$M_n(v) := \sum_{s \in V(\Gamma)} d^2(v,s)\mu_n(s)$$

be the random weight, called the *sampling weight*, corresponding to $v \in V(\Gamma)$, and $M_n(\cdot)$ the resulting random *sampling weight function.*

Definition 3.11. The set of vertices

$$\mathbb{S}_n = \mathbb{S}(\xi_1, \ldots, \xi_n) := \{v \in V(\Gamma) \mid M_n(v) \le M_n(u), \quad \forall u \in V(\Gamma)\}$$

is called the *sample mean-set* (or *sample center-set*) relative to ξ.

Our goal is to prove that our (empirical) sample mean-set $\mathbb{S}_n$ converges, in some sense, to the (theoretical) mean-set $\mathbb{E}(\xi)$ as $n \to \infty$. To achieve this goal, we need to consider the notion of limit for a sequence of subsets of vertices in $V(\Gamma)$, in a context that would make our theory work mathematically and, at the same time, would not contradict practical considerations. It turns out that the limit-superior *limsup* does the job. One can easily see from the simple Example 3.12 below that it may happen, in cases when the mean-sets in graphs contain more than one vertex, that $\limsup_{n\to\infty} \mathbb{S}_n = \mathbb{E}(\xi)$, while $\liminf_{n\to\infty} \mathbb{S}_n = \emptyset$. This implies that we cannot formulate the desired law in terms of the classical limit for a sequence of sets for multi-vertex mean-sets, since that would require perfect agreement of the limit-superior *limsup* and the limit-inferior *liminf.*

Let $\{V_n\}_{n=1}^{\infty}$ be a sequence of subsets of V. Recall that

$$\limsup_{n\to\infty} V_n = \{v \mid v \in V_{n_k}, \quad k = 1, 2, \ldots\}$$

for some subsequence $\{n_k\}$ depending on v. We write that $\limsup_{n\to\infty} V_n = \{v \mid v \in V_n,\ \text{i.o.}\}$, where i.o. stands for "infinitely often." Similarly,

$$\liminf_{n\to\infty} V_n = \{v : v \in V_n \text{ for all } n \text{ except for a finite number } \} =$$

$$= \{v : v \in V_n \text{ for all } n \geq n_0(v)\}.$$

Properties of limits of sets can be found in numerous sources in the literature, (3) in particular.

Example 3.12. Consider a graph $\Gamma = (V, E)$ where $V = \{v_1, v_2\}$ and $E = \{(v_1, v_2)\}$, i.e., the graph Γ is the connected graph on 2 vertices. Let μ be the probability measure on $V(\Gamma)$ induced by some random element $\xi_1 : \Omega \to V$ and defined by $\mu(v_1) = \mu(v_2) = 1/2$. In that case $M(v_1) = M(v_2) = 1/2$ and, consequently, $\mathbb{E}(\xi) = \{v_1, v_2\}$.

Consider a sequence $\xi_1, \xi_2, \ldots$ of such random elements (independent and identically distributed). In other words, we just deal with a sequence of randomly generated vertices v_1, v_2 of the graph Γ. By definition,

$$M_n(v_1) = \frac{1}{n}|\{i \mid \xi_i = v_2, 1 \leq i \leq n\}| \ \text{ and } \ M_n(v_2) = \frac{1}{n}|\{i \mid \xi_i = v_1, 1 \leq i \leq n\}|.$$

Hence,

$$v_1 \in \mathbb{S}_n \ \Leftrightarrow \ M_n(v_1) \leq M_n(v_2) \ \Leftrightarrow \ |\{i \mid \xi_i = v_2, 1 \leq i \leq n\}| \leq |\{i \mid \xi_i = v_1, 1 \leq i \leq n\}|$$

and, similarly,

$$v_2 \in \mathbb{S}_n \ \Leftrightarrow \ M_n(v_2) \leq M_n(v_1) \ \Leftrightarrow \ |\{i \mid \xi_i = v_1, 1 \leq i \leq n\}| \leq |\{i \mid \xi_i = v_2, 1 \leq i \leq n\}|.$$

Let

$$R(n) := nM_n(v_1) - nM_n(v_2) = |\{i \mid \xi_i = v_2, 1 \leq i \leq n\}| - |\{i \mid \xi_i = v_1, 1 \leq i \leq n\}|.$$

Observe that we can think of $R(n)$ as a simple symmetric random walk on $\mathbf{Z}$ starting at 0 and

$$\begin{cases} R(n+1) = R(n) - 1, & \text{if } \xi_{n+1} = v_1; \\ R(n+1) = R(n) + 1, & \text{if } \xi_{n+1} = v_2. \end{cases}$$

In other words, $R(0) = 0$, $R(n) = \sum_{i=1}^{\infty} \zeta_i$ where $\zeta_1, \zeta_2, \ldots$ are i.i.d. random variables such that

$$\zeta_1 = \begin{cases} 1, & \text{with probability } 1/2; \\ -1, & \text{with probability } 1/2. \end{cases}$$

Moreover, we have

$$\{v_1 \in \mathbb{S}_n\} = \{M_n(v_1) \le M_n(v_2)\} = \{R(n) \le 0\}$$

and

$$\{v_2 \in \mathbb{S}_n\} = \{M_n(v_1) \ge M_n(v_2)\} = \{R(n) \ge 0\}.$$

Now, since a simple symmetric random walk on $\mathbf{Z}$ is recurrent, it follows that for every $i = 1, 2$, $\mathbf{P}\{v_i \in \mathbb{S}_n, \ \text{ i.o.}\} = 1$ Hence, almost always we have

$$\limsup_{n\to\infty} \mathbb{S}_n = \{v_1, v_2\} = \mathbb{E}\xi.$$

$$\liminf_{n\to\infty} \mathbb{S}_n = \emptyset,$$

and $\lim\limits_{n\to\infty} \mathbb{S}_n$ does not exist.

□

Lemma 3.13. *Let Γ be a locally-finite connected graph, $v \in V(\Gamma)$, and $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$ such that $M_{\xi_1}(v)$ is defined. Then*

$$\mathbf{P}\Big(M_n(v) \to M(v) \ \text{as } n \to \infty\Big) = 1. \tag{3.7}$$

Proof. For every $v \in V(\Gamma)$, the value $M(v) = \sum_{u \in V(\Gamma)} d^2(v, u)\mu(u)$ is equal to the expectation $\mathbb{E}(d^2(v, \xi_1))$ of the random variable $d^2(v, \xi_1)$. Hence, by the strong law of large numbers for i.i.d. random variables $\{d^2(v, \xi_i)\}_{i=1}^{\infty}$, we have the required a.s. convergence $M_n(v)$ to $M(v)$. □

It is important to notice that in general the convergence in Lemma 3.13 is not uniform in a sense that, for some distribution μ on a locally finite (infinite) graph Γ and some $\varepsilon > 0$, it is possible that

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N \ \forall v \in V(\Gamma), \ \ |M_n(v) - M(v)| < \varepsilon\Big) \ne 1.$$

In other words, the convergence for every vertex, as in Lemma 3.13, is insufficient to prove the strong law of large numbers, stated in introduction. Next lemma is a key tool in the proof of our strong law of large numbers.

Lemma 3.14 (Separation Lemma)**.** *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If the weight function $M_{\xi_1}(\cdot)$ is totally defined, then*

$$\mathbf{P}\Big(\exists N \ s.t. \ \forall n > N, \ \max_{v \in \mathbb{E}(\xi_1)} M_n(v) < \inf_{u \in V(\Gamma) \setminus \mathbb{E}(\xi_1)} M_n(u)\Big) = 1.$$

Proof. Our goal is to prove that for some $\delta > 0$

$$\mathbf{P}\Big(\exists N \ \forall n > N \ \forall v \in \mathbb{E}(\xi_1), \ \forall u \in V(\Gamma) \setminus \mathbb{E}(\xi_1), \quad M_n(u) - M_n(v) \geq \delta\Big) = 1. \quad (3.8)$$

We prove the formula above in two stages. In the first stage we show that for some fixed $v_0 \in \mathbb{E}(\xi_1)$ and for sufficiently large number $m > 0$ the following holds

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N \ \forall v \in \mathbb{E}(\xi_1), \ \forall u \in V(\Gamma) \setminus B_{v_0}(m), \quad M_n(u) - M_n(v) \geq \delta\Big) = 1 \quad (3.9)$$

in the notation of (3.4). In the second stage we prove that

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N \ \forall v \in \mathbb{E}(\xi_1), \ \forall u \in B_{v_0}(m) \setminus \mathbb{E}(\xi_1), \quad M_n(u) - M_n(v) \geq \delta\Big) = 1 \quad (3.10)$$

Having the formulae above proved we immediately deduce that (3.8) holds using σ-additivity of measure.

Let $v_0 \in \mathbb{E}(\xi_1)$ and μ be the probability measure on Γ induced by ξ_1, as in (3.1). As described in the beginning of Section 3.1.1, we can naturally put ourselves on the probability space $(V(\Gamma), \mathcal{S}, \mu)$, which is the image of the original probability space under the mapping $\xi : \Omega \to V(\Gamma)$. Since the weight function $M(\cdot)$ is defined at v_0, we can choose $r \in \mathbf{R}$ as in Lemma 3.5, such that $\frac{1}{2}M(v_0) \leq \sum_{i \in B_{v_0}(r)} d^2(v_0, i)\mu(i)$. Put $m = 3r$. In Lemma 3.5 we proved that, if a vertex u is such that $d(u, v_0) \geq 3r$, then

$$M(u) = \sum_{i \in V(\Gamma)} d^2(u, i)\mu(i) \geq 4 \sum_{i \in B_{v_0}(r)} d^2(u, i)\mu(i) \geq 2M(v_0). \quad (3.11)$$

It implies that $\mathbb{E}(\xi_1) \subseteq B_{v_0}(3r)$.

Since Γ is locally finite, the set $B_{v_0}(r)$ of (3.4) is finite. We also know from the SLLN for the relative frequencies $\mu_n(u)$ that $\mu_n(u) \overset{a.s.}{\to} \mu(u)$ as $n \to \infty$. These facts imply that for any $\varepsilon > 0$, the event

$$C_\varepsilon := \{\exists N = N(\varepsilon), \quad \forall n > N, \quad \forall u \in B_{v_0}(r), \quad |\mu_n(u) - \mu(u)| < \varepsilon\} \tag{3.12}$$

has probability one. In particular, this is true for

$$\varepsilon = \varepsilon^* := \frac{1}{4} \min\{\mu(u) \mid u \in B_{v_0}(r), \ \mu(u) \neq 0\},$$

and the event C_{ε^*} is a subset of

$$\left\{\exists N = N(\varepsilon^*), \quad \forall n > N, \quad \forall u \in V(\Gamma) \setminus B_{v_0}(3r), \quad M_n(u) \geq \frac{3}{2} M(v_0)\right\}. \tag{3.13}$$

Indeed, on the event C_{ε^*}, as in (3.12), we have $\mu_n(i) \geq \frac{3}{4}\mu(i)$, $i \in B_{v_0}(r)$. Using this fact together with (3.11), we can write

$$M_n(u) = \sum_{i \in V(\Gamma)} d^2(u,i)\mu_n(i) \geq 4 \sum_{i \in B_{v_0}(r)} d^2(u,i)\mu_n(i) \geq 3 \sum_{i \in B_{v_0}(r)} d^2(u,i)\mu(i) \geq \frac{3}{2} M(v_0).$$

Thus we have

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N, \ \forall u \in V(\Gamma) \setminus B_{v_0}(3r), \quad M_n(u) \geq \frac{3}{2} M(v_0)\Big) = 1. \tag{3.14}$$

By Lemma 3.13, for any $v \in V(\Gamma)$, for any $\varepsilon > 0$, we have

$$\mathbf{P}\Big(\exists N = N(\varepsilon), \quad \forall n > N, \quad |M_n(v) - M(v)| < \varepsilon\Big) = 1$$

and, since $B_{v_0}(3r)$ is a finite set, we have simultaneous convergence for all vertices in $B_{v_0}(3r)$, i.e.,

$$\mathbf{P}\Big(\exists N = N(\varepsilon), \quad \forall n > N, \quad \forall v \in B_{v_0}(3r), \quad |M_n(v) - M(v)| < \varepsilon\Big) = 1. \tag{3.15}$$

In particular, remembering that $\mathbb{E}(\xi_1) \subseteq B_{v_0}(3r)$, for $\varepsilon = M(v_0)/4$,

$$\mathbf{P}\Big(\exists N = N(\varepsilon), \quad \forall n > N, \ \forall v \in \mathbb{E}(\xi_1), \quad \frac{3}{4} M(v) < M_n(v) < \frac{5}{4} M(v)\Big) = 1. \tag{3.16}$$

Finally, we notice that on the intersection of the events in (3.14) and (3.16), we have

$$M_n(u) - M_n(v) \geq \frac{3}{2}M(v) - \frac{5}{4}M(v) = \frac{1}{4}M(v) = \frac{1}{4}M(v_0),$$

by the virtue of the fact that $M(v_0) = M(v)$ (as both $v_0, v \in \mathbb{E}(\xi_1)$), and formula (3.9) holds for any δ such that $\delta \leq \frac{1}{4}M(v_0)$.

For the second part of our proof we use statement (3.15) that holds, in particular, for

$$\varepsilon = \varepsilon' := \frac{1}{4}\min\{M(u) - M(v_0) \mid u \in B_{v_0}(3r),\ M(u) - M(v_0) > 0\}.$$

It means that, with probability 1, there exists $N = N(\varepsilon')$ such that for any $n > N$ and all $u \in B_{v_0}(3r)$, we have $|M_n(u) - M(u)| < \varepsilon'$. Moreover, since $\mathbb{E}(\xi_1) \subseteq B_{v_0}(3r)$, we can assert the same for any $v \in \mathbb{E}(\xi_1)$; namely, $|M_n(v) - M(v)| < \varepsilon'$. Together with the fact that $M(u) - M(v_0) > 0$, the obtained inequalities imply that, with probability 1, there exists number $N = N(\varepsilon')$ such that for any $n > N$ and all $u \in B_{v_0}(3r) \setminus \mathbb{E}(\xi_1)$,

$$M_n(v_0) < M(v_0) + \varepsilon' \leq M(v_0) + \frac{1}{4}(M(u) - M(v_0))$$

$$M(u) - \frac{1}{4}(M(u) - M(v_0)) \leq M(u) - \varepsilon' < M_n(u),$$

and, hence,

$$M_n(u) - M_n(v_0) \geq M(u) - \frac{1}{4}(M(u) - M(v_0)) - M(v_0) - \frac{1}{4}(M(u) - M(v_0)) =$$

$$= \frac{1}{2}(M(u) - M(v_0)) \geq 2\varepsilon',\ \text{i.e.,}$$

$$\mathbf{P}\Big(\exists N = N(\varepsilon),\quad \forall n > N,\quad \forall u \in B_{v_0}(3r) \setminus \mathbb{E}(\xi_1):\quad M_n(u) - M_n(v_0) \geq 2\varepsilon'\Big) = 1.$$

Therefore, (3.10) holds for any $\delta \leq 2\varepsilon'$. Choosing $\delta = \min(\frac{1}{4}M(v_0), 2\varepsilon')$ finishes the proof.

□

Corollary 3.15 (Inclusion Lemma). *Let Γ be a locally-finite connected graph, $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. Γ-valued random elements defined on a given probability space*

$(\Omega, \mathcal{F}, \mathbf{P})$ with values in $V(\Gamma)$, and μ the probability measure on Γ induced by ξ_1. Suppose that weight function M is totally defined. Then

$$\mathbf{P}\Big(\limsup_{n\to\infty} \mathbb{S}_n \subseteq \mathbb{E}(\xi_1)\Big) = 1.$$

Proof.

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N,\ \max_{v\in\mathbb{E}(\xi_1)} M_n(v) < \inf_{u\in V(\Gamma)\setminus\mathbb{E}(\xi_1)} M_n(u)\Big) = 1,$$

by Separation Lemma 3.14. Thus, for every $u \notin \mathbb{E}(\xi_1)$, we have

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N,\ u \notin \mathbb{S}_n\Big) = \mathbf{P}\Big(u \notin \limsup_{n\to\infty} \mathbb{S}_n\Big) = 1$$

By σ-additivity of measure, we obtain,

$$\mathbf{P}\Big(u \notin \limsup \mathbb{S}_n,\ \text{for every } u \in V(\Gamma) \setminus \mathbb{E}(\xi_1)\Big) = 1.$$

□

Now we are ready to prove the strong law of large numbers for center-sets containing only one element. This is the only case when the classical *limit* of sets works, as opposed to multi-vertex center-sets, when the law holds in the sense of *limsup*.

Theorem 3.16. (SLLN for graph-valued random elements with a singleton mean-set.) *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^\infty$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If the weight function $M_{\xi_1}(\cdot)$ is totally defined and $\mathbb{E}(\xi_1) = \{v\}$ for some $v \in V(\Gamma)$, i.e., if $\mathbb{E}(\xi_1)$ is a singleton, then the following holds almost surely:*

$$\mathbb{S}(\xi_1, \ldots, \xi_n) \longrightarrow \mathbb{E}(\xi_1) \text{ as } n \to \infty.$$

Proof. Observe that the conclusions of the theorem can be expressed by the statement

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N,\ \mathbb{S}(\xi_1, \ldots, \xi_n) = \{v\}\Big) = 1$$

or, equivalently,

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N\ \forall u \in V(\Gamma) \setminus \{v\},\ \ M_n(v) < M_n(u)\Big) = 1. \tag{3.17}$$

By Separation Lemma 3.14,

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N,\ M_n(v) < \inf_{u \in V(\Gamma)\setminus\{v\}} M_n(u)\Big) = 1,$$

and the statement $\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N,\ \{v\} = \mathbb{S}(\xi_1, \ldots, \xi_n)\Big) = 1$ is proved. □

3.3 SLLN for multi-vertex mean-sets

In this section we investigate a multi-vertex center-set case and conditions under which the strong law of large numbers holds for such set. We reduce this problem to the question of recurrence of a certain subset in $\mathbf{Z}^n$ relative to a random walk on this integer lattice. If $2 \le |\mathbb{E}(\xi)| \le 3$, no restrictive assumptions are required; we formulate and prove the strong law of large numbers for these special instances separately. The case of $|\mathbb{E}(\xi)| > 3$ requires more technical assumptions, and, therefore, more work to handle it.

3.3.1 Case of 2 vertices

Theorem 3.17. (SLLN for graph-valued random elements with two point mean-set.) *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If the weight function $M_{\xi_1}(\cdot)$ is totally defined and $|\mathbb{E}(\xi)| = 2$, then*

$$\limsup_{n\to\infty} \mathbb{S}(\xi_1, \ldots, \xi_n) = \mathbb{E}(\xi_1)$$

holds with probability 1.

Proof. By Inclusion Lemma 3.15, $\limsup_{n\to\infty} \mathbb{S}_n \subseteq \mathbb{E}(\xi_1)$. Thus, in order to prove the theorem, it suffices to show the reverse inclusion. Assume $\mathbb{E}(\xi_1) = \{v_1, v_2\}$, i.e., the weight function $M_\xi(\cdot)$ attains its minimum value at the vertices v_1 and v_2, and only those:

$$M(v_1) = M(v_2) < M(u) \quad \text{for any} \quad u \in \Gamma \setminus \{v_1, v_2\}.$$

By definition,

$$M(v_1) = \sum_{s \in \Gamma} d^2(v_1, s)\mu(s), \quad M(v_2) = \sum_{s \in \Gamma} d^2(v_2, s)\mu(s),$$

$$M_n(v_1) = \sum_{s \in \Gamma} d^2(v_1, s)\frac{|\{i \mid \xi_i = s, 1 \le i \le n\}|}{n}, \;\; M_n(v_2) = \sum_{s \in \Gamma} d^2(v_2, s)\frac{|\{i \mid \xi_i = s, 1 \le i \le n\}|}{n},$$

and by Lemma 3.13, $M_n(v_1) \to M(v_1)$ and $M_n(v_2) \to M(v_2)$ almost surely as $n \to \infty$. By Separation Lemma 3.14 it follows that, with probability one, for the sequence of random elements $\xi_1, \xi_2, \ldots$, there exists a number N such that for any $n > N$, we have

$$\max\{M_n(v_1), M_n(v_2)\} < \inf_{u \in \Gamma \setminus \{v_1, v_2\}} M_n(u).$$

Hence, for any $n > N$,

$$v_1 \in \mathbb{S}_n \;\; \text{if and only if } M_n(v_1) \le M_n(v_2)$$

$$\text{if and only if } \sum_{s \in \Gamma} \big(d^2(v_1, s) - d^2(v_2, s)\big)|\{i \mid \xi_i = s, 1 \le i \le n\}| \le 0.$$

Similarly,

$$v_2 \in \mathbb{S}_n \;\; \text{if and only if } M_n(v_2) \le M_n(v_1)$$

$$\text{if and only if } \sum_{s \in \Gamma} \big(d^2(v_2, s) - d^2(v_1, s)\big)|\{i \mid \xi_i = s, 1 \le i \le n\}| \le 0.$$

Observe that in order to show that $\{v_1, v_2\} \subseteq \limsup_{n \to \infty} \mathbb{S}_n$, it is enough to prove that the difference $M_n(v_1) - M_n(v_2)$ takes on positive and negative values infinitely often. For every $n \in \mathbb{N}$, define

$$R(n) := n(M_n(v_2) - M_n(v_1)) = \sum_{s \in \Gamma} \big(d^2(v_2, s) - d^2(v_1, s)\big) \cdot |\{i \mid \xi_i = s, 1 \le i \le n\}|$$

and notice that

$$R(n+1) - R(n) = \sum_{s \in \Gamma} \;\; [d^2(v_2, s) - d^2(v_1, s)] \;\; \mathbf{1}_{\{\xi_{n+1} = s\}}.$$

We immediately recognize $\{R(n)\}_{n \in \mathbb{N}_0}$ as a random walk on $\mathbf{Z}$ starting at 0; namely, $R(0) = 0, R(n) = \sum_{i=1}^{n} \zeta_i$, with $\zeta_i = \zeta_i(s) : V(\Gamma) \to \mathbf{Z}, i = 1, 2, \ldots$ are i.i.d. random variables such that $\mathbb{E}(\zeta_1) = 0$ (since $M(v_1) = M(v_2)$), $\zeta_i(s) = d^2(v_2, s) - d^2(v_1, s)$ with probability $\mu(s), s \in V(\Gamma)$. Now, notice that for any $n > N$

$$v_1 \in \mathbb{S}_n \quad \Leftrightarrow \quad R(n) \geq 0 \quad \text{and} \quad v_2 \in \mathbb{S}_n \quad \Leftrightarrow \quad R(n) \leq 0.$$

It is known that a general (not simple, not symmetric) one-dimensional random walk $R(n) = \sum_{i=1}^{n} \zeta_i$ on $\mathbf{Z}$ is recurrent if $\sum_{s \in \Gamma} |\zeta_1(s)|\mu(s) < \infty$ and $\sum_{s \in \Gamma} \zeta_1(s)\mu(s) = 0$ (see (36), pg. 23). We have seen that the second condition holds since

$$\mathbb{E}(\zeta_1) = \sum_{s \in \Gamma} (d^2(v_2, s) - d^2(v_1, s))\mu(s) = M(v_1) - M(v_2) = 0.$$

The first sufficient condition for recurrence is also trivial to check:

$$\sum_{s \in \Gamma} |\zeta_1(s)|\mu(s) = \sum_{s \in \Gamma} |d^2(v_2, s) - d^2(v_1, s)|\mu(s) \leq$$

$$\leq \sum_{s \in \Gamma} (d^2(v_2, s) + d^2(v_1, s))\mu(s) = M(v_1) + M(v_2) < \infty$$

since both weight functions are assumed to be defined. Thus, our random walk takes on positive and negative values infinitely often. We conclude that almost always $\limsup_{n \to \infty} \mathbb{S}_n = \{v_1, v_2\} = \mathbb{E}\xi$. □

3.3.2 Case of 3 and more vertices

Assume $\mathbb{E}(\xi_1) = \{v_1, v_2, \ldots, v_k\}$, i.e., for any $u \in \Gamma \setminus \{v_1, v_2, \ldots, v_k\}$, $M(v_1) = M(v_2) = \ldots = M(v_k) < M(u)$. Our goal is to formulate conditions that would guarantee the inclusion $\mathbb{E}(\xi_1) \subseteq \limsup_{n \to \infty} \mathbb{S}_n$ or, without loss of generality, conditions for $\{v_1\} \in \limsup_{n \to \infty} \mathbb{S}_n$.

By Separation Lemma 3.14, it follows that, with probability one, for a sequence of random elements $\xi_1, \xi_2, \ldots$, there exists a number N such that for any $n > N$ we have

$$\max\{M_n(v_1), M_n(v_2), \ldots, M_n(v_k)\} < \inf_{u \in \Gamma \setminus \{v_1, v_2, \ldots, v_k\}} M_n(u).$$

Hence, for any $n > N$, $v_1 \in \mathbb{S}_n$ if and only if $M_n(v_1) \leq M_n(v_i)$ for every $i = 2, \ldots, k$. Thus, to achieve our goal, we have to show that conditions

$$M_n(v_2) - M_n(v_1) \geq 0, \quad \ldots \quad , M_n(v_k) - M_n(v_1) \geq 0$$

simultaneously hold infinitely often.

As in the case for two-point mean-sets, for every $i = 1, \ldots, k-1$ and $n \in \mathbb{N}$, define

$$R_i(n) := n(M_n(v_{i+1}) - M_n(v_1)) = \sum_{s\in\Gamma} \left(d^2(v_{i+1}, s) - d^2(v_1, s)\right) \cdot |\{i \mid \xi_i = s, 1 \leq i \leq n\}|$$

and observe, as before, that

$$R_i(n+1) - R_i(n) = \sum_{s\in\Gamma} \ [d^2(v_{i+1}, s) - d^2(v_1, s)] \ \ \mathbf{1}_{\{\xi_{n+1}=s\}}, \tag{3.18}$$

i.e., $R_i(n)$, $i = 1, \ldots, k-1$, represent random walks, associated with v_1, on $\mathbf{Z}$ starting at 0.

Consider a random walk $\overline{R}$, associated with v_1, in $\mathbf{Z}^{k-1}$, starting at the origin $(0, \ldots, 0)$ with the position of the walk after n steps given by $\overline{R}(n) = (R_1(n), R_2(n), \ldots, R_{k-1}(n))$. An increment step in $\mathbf{Z}^{k-1}$ is given by $\overline{\zeta}(s) = (\zeta_1(s), \ldots, \zeta_{k-1}(s))$, $s \in V(\Gamma)$, with probability $\mu(s)$, where $\zeta_i(s) : V(\Gamma) \to \mathbf{Z}$, $\zeta_i(s) = d^2(v_{i+1}, s) - d^2(v_1, s)$, $i = 1, \ldots, k-1$. The following lemma shows the significance of this random walk.

Lemma 3.18. *In the notation of this section, $\{v_1\} \in \limsup_{n\to\infty} \mathbb{S}_n$ if and only if the random walk $\overline{R}$ visits the set $\mathbf{Z}_+^{k-1} = \{(a_1, \ldots, a_{k-1}) \mid a_i \geq 0\}$ infinitely often. Therefore,*

$$\mathbf{P}(v_1 \in \limsup_{n\to\infty} \mathbb{S}_n) = \mathbf{P}(\overline{R}(n) \in \mathbf{Z}_+^{k-1}, \ \textit{i.o.}).$$

Proof. Follows from the discussion preceding the lemma. □

It is worth redefining $\overline{R}$ in the terms of transition probability function, as in (36). Let $\overline{0} \in \mathbf{Z}^{k-1}$ be the zero vector and $x_i = \zeta_i(s)$, $s \in V(\Gamma)$. For every $\overline{x} = (x_1, \ldots, x_{k-1}) \in \mathbf{Z}^{k-1}$, we define a function $P(\overline{0}, \overline{x})$ by

$$P(\overline{0}, \overline{x}) = \mu\{s \mid x_i = d^2(v_{i+1}, s) - d^2(v_1, s) \text{ for every } i = 1, \ldots, k-1\}. \tag{3.19}$$

This is trivial to check that this is, indeed, the transition probability for $\overline{R}$. To continue further, we investigate some properties of our random walk $\overline{R}$.

Lemma 3.19. *Let $\overline{R}$ be a random walk defined above. Then*

$$m_1 = \sum_{\overline{x} \in \mathbf{Z}^{k-1}} \overline{x} P(\overline{0}, \overline{x}) = \overline{0} \quad \textit{and} \quad m_2 = \sum_{\overline{x} \in \mathbf{Z}^{k-1}} |\overline{x}|^2 P(\overline{0}, \overline{x}) < \infty.$$

Proof. The first equality trivially holds. Consider the left hand side of the second inequality

$$\sum_{\overline{x} \in \mathbf{Z}^{k-1}} |\overline{x}|^2 P(\overline{0}, \overline{x}) = \sum_{s \in V(\Gamma)} \sum_{i=1}^{k-1} \Big(d^2(v_{i+1}, s) - d^2(v_1, s) \Big)^2 \mu(s)$$

$$= \sum_{i=1}^{k-1} \sum_{s \in V(\Gamma)} \Big(d(v_{i+1}, s) - d(v_1, s) \Big)^2 \Big(d(v_1, s) + d(v_{i+1}, s) \Big)^2 \mu(s)$$

$$\leq \sum_{i=1}^{k-1} d^2(v_1, v_{i+1}) \sum_{s \in V(\Gamma)} \Big(d(v_1, s) + d(v_{i+1}, s) \Big)^2 \mu(s)$$

$$\leq \sum_{i=1}^{k-1} d^2(v_1, v_{i+1})(4M(v_1) + 4M(v_{i+1})) < \infty,$$

where, in the last estimate, we break the sum $\sum_{s \in V(\Gamma)} \Big(d(v_1, s) + d(v_{i+1}, s) \Big)^2 \mu(s)$ into two sums over $s \in V(\Gamma)$ with $d(v_1, s) < d(v_{i+1}, s)$ and $d(v_1, s) \geq d(v_{i+1}, s)$ and overestimate them. □

Clearly, conditions under which this random walk is recurrent would guarantee that $\{v_1\} \subseteq \limsup_{n \to \infty} \mathbb{S}_n$ (see (36, page 30, Proposition 3.3)). Sufficient conditions for the recurrence of two-dimensional random walk involve the finiteness of its second moment and can be found in (36, page 83). The result stated there indicates that genuinely 2-dimensional random walk is recurrent if its first moment is zero, and its second moment is finite. Let us recall the notion of genuinely-dimensional random walk, as well as some other relevant concepts, before we go on.

Consider an arbitrary random walk $\overline{R}$ on $\mathbf{Z}^n$ given by a transition probability P, as in (3.19). The support, *supp*(P), of the probability measure P is defined to be the set *supp*$(P) := \{\overline{v} \in \mathbf{Z}^n \mid P(\overline{v}) \neq 0\}$ of all possible one-step increments of $\overline{R}$. Further, with $\overline{R}$, one can associate an abelian subgroup $A_{\overline{R}}$ of $\mathbf{Z}^n$ generated by the vectors in *supp*(P). It is well-known in group theory that any subgroup $A_{\overline{R}}$ of $\mathbf{Z}^n$ is isomorphic to $\mathbf{Z}^k$, where $k \leq n$ (the reader can also check (36, Proposition7.1 on pg.65) for details), in which case we write $\dim(A_{\overline{R}}) = k$ and say that $\overline{R}$ is *genuinely k-dimensional.* Let us stress that we speak of an n-dimensional random walk on $\mathbf{Z}^n$ when $P(0, \overline{x})$ is defined for all $\overline{x}$ in $\mathbf{Z}^n$; this walk is genuinely n-dimensional if $\dim(A_{\overline{R}}) = n$. We say that $\overline{R}$ is *aperiodic* if $A_{\overline{R}} = \mathbf{Z}^n$. Observe that genuinely n-dimensional random walk does not have to be aperiodic. A standard simple random walk, which we denote by $S = S(n)$, is an example of an aperiodic random walk on $\mathbf{Z}^n$. It will be convenient to define a vector space $V_{\overline{R}} \subset \mathbf{R}^n$ spanned by the vectors in *supp*(P). It is easy to see that the genuine dimension of $\overline{R}$ is equal to the dimension of $V_{\overline{R}}$. We shall need another notion for our developments. Assume that D is an $k \times n$ matrix (not necessarily integer valued) which maps $A_{\overline{R}}$ onto $\mathbf{Z}^k$. Then D naturally induces a random walk $\overline{R}^D$ on $\mathbf{Z}^k$ with transition probability P^D given by $P^D(\overline{u}) = P(\overline{v} \in \mathbf{Z}^n \mid D(\overline{v}) = \overline{u})$ for every $\overline{u} \in \mathbf{Z}^k$. Now, we have our strong law of large numbers for mean-sets with three elements almost for free.

Theorem 3.20 (SLLN for graph-valued random elements with three point mean-set)**.** *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If the weight function $M_{\xi_1}(\cdot)$ is totally defined and $|\mathbb{E}(\xi)| = 3$, then*

$$\limsup_{n\to\infty} \mathbb{S}(\xi_1, \ldots, \xi_n) = \mathbb{E}(\xi_1)$$

holds with probability 1.

Proof. Let $v \in \mathbb{E}(\xi_1)$ and $\overline{R}$ a random walk in $\mathbf{Z}^2$, associated with v_1. This random walk can be genuinely m-dimensional, where $m \in \{0, 1, 2\}$. By Lemma 3.19, the first moment of $\overline{R}$ is $(0, 0)$ and the second moment is finite. Therefore, by (36, Theorem 8.1) the genuinely 2-dimensional random walk is recurrent, and, hence,

$\mathbf{Z}_+^{k-1}$ is visited infinitely often with probability 1. Finally, by Lemma 3.18, it follows that $\mathbf{P}(v \in \limsup_{n\to\infty} \mathbb{S}_n) = 1$ for every $v \in \mathbb{E}(\xi_1)$. Thus the result.

□

Recall that a subset of $\mathbf{Z}^n$ is called recurrent if it is visited by a given random walk infinitely often with probability one, and it is transient otherwise (according to the Hewitt-Savage $0-1$ law, any set is either visited infinitely often with probability one or with probability zero). The criterion for whether a given set is recurrent or transient for simple random walk was obtained by Itô and McKean (19) for $n = 3$ (it can also be found in (36, Theorem 26.1)). It turns out that the criterion does not depend on the random walk in question, if the walk is aperiodic. This is the subject of the extension of the Wiener's test, proved in (37), that we state below. This invariance principle is one of the main tools we use in our investigation of the recurrence properties of the positive octant in $\mathbf{Z}^n$ for our random walk $\overline{R}$.

Theorem (Extension of Wiener's test – Invariance Principle, (37))**.** *Let $n \geq 3$. Then an infinite subset A of $\mathbf{Z}^n$ is either recurrent for each aperiodic random walk $\overline{R}$ on $\mathbf{Z}^n$ with mean zero and a finite variance, or transient for each of such random walks.*

For a positive constant $\alpha \in \mathbf{R}$ and a positive integer $m \leq n$ define a subset of $\mathbf{R}^n$

$$Cone_\alpha^m = \left\{(x_1,\ldots,x_n) \in \mathbf{R}^n \mid x_1 = 0,\ldots,x_{n-m} = 0,\ \sqrt{x_{n-m+1}^2 + \ldots + x_{n-1}^2} \leq \alpha x_n\right\}$$

called an *m-dimensional cone* in $\mathbf{R}^n$. If $m = n$, then we omit the superscript in $Cone_\alpha^m$. For an $n \times n$ matrix D and a set $A \subseteq \mathbf{R}^n$, define a set $A^D = \{D \cdot \overline{v} \mid \overline{v} \in A\}$, which is a linear transformation of A. If D is an orthogonal matrix, then the set $(Cone_\alpha)^D$ is called a *rotated cone.* Following (19), for any non-decreasing function $i : \mathbb{N} \to \mathbf{R}_+$ define a set

$$Thorn_i = \{\overline{v} \in \mathbf{Z}^n \mid \sqrt{v_1^2 + \ldots + v_{n-1}^2} \leq i(v_n)\}.$$

Observe that $Cone_\alpha \cap \mathbf{Z}^n = Thorn_i$ where $i(t) = \alpha t$. In (19), Itô and McKean prove the recurrence criterion for $Thorn_i$; namely, the authors show that $Thorn_i$ is visited infinitely often by standard random walk with probability one if $\sum_{k\geq 1}\left(2^{-k}i(2^k)\right)^{n-3} = \infty$ for dimensions $n \geq 4$. Since in our case $i(t) = \alpha t$, we have $i(2^k) = \alpha 2^k$, and, hence,

$$\sum_{k\geq 1}\left(2^{-k}i(2^k)\right)^{n-3} = \sum_{k\geq 1}\left(2^{-k}\alpha 2^k\right)^{n-3} = \infty.$$

Thus, the criterion is satisfied. When dimension is $n = 3$, the same authors show that even the thinnest thorn $\bigcup_{k\geq 1}(0,0,k)$ is recurrent (see (19)). In their proofs, Itô and McKean evaluate Wiener's sum from the Wiener's test using *capacities* of the spherical shells. We are not going to get into the domain of capacities because it is not relevant to our work. We just use the results about the recurrence of sets that we need to achieve our goal. Keeping in mind that capacities of sets are invariant under orthogonal transformations (see (19) again), we arrive at the following important theorem.

Theorem 3.21. *For any $\alpha > 0$ and any orthogonal matrix D,*

$$\mathbf{P}\big(S(n) \in (Cone_\alpha)^D,\ i.o.\big) = 1,$$

i.e., the probability that the simple random walk on $\mathbf{Z}^n$ visits $(Cone_\alpha)^D$ infinitely often is 1*.*

Proof. Direct consequence of (6.1) and (4.3) in (19), where the criterion for recurrence of $Thorn_i$ is given (see also the discussion above). □

Lemma 3.22. *Assume that a set $A \subseteq \mathbf{R}^n$ contains a rotated cone. Then for any invertible $n \times n$ matrix D, the set A^D contains a rotated cone.*

Proof. Exercise in linear algebra. See Lemma 3.35 below. □

Lemma 3.23. *If $S_1 \subseteq S_2 \subseteq \mathbf{R}^n$ and S_1 is visited by the simple random walk infinitely often with probability* 1 *then S_2 is visited by the simple random walk infinitely often with probability* 1*.*

Proof. Obvious. □

Now, we return to our strong law of large numbers for multi-vertex mean-sets. Assume that $\mathbb{E}\xi = \{v_1, \ldots, v_k\}$, where $k \geq 4$. Let $\overline{R}^i$ be a random walk on $\mathbf{Z}^{k-1}$, associated with v_i, where $i = 1, \ldots, k$ (in our notation, $\overline{R} = \overline{R}^1$). This is a $(k-1)$-dimensional random walk which, in general, is not aperiodic. In fact, $\overline{R}^i$ is not even genuinely $(k-1)$-dimensional. Fortunately, it turns out that it does not matter to what vertex v_i we associate our random walk, since the choice of the vertex does not affect the dimension of the corresponding walk, as the following lemma shows.

Lemma 3.24. *Let μ be a probability measure on a locally finite graph Γ such that $\mathbb{E}\mu = \{v_1, \ldots, v_k\}$, where $k \geq 2$. Then the random walks $\overline{R}^1, \ldots, \overline{R}^k$, associated with vertices $v_1, \ldots, v_k$ respectively, all have the same genuine dimension.*

Proof. We prove that random walks $\overline{R}^1$ and $\overline{R}^2$ have the same genuine dimension. Recall that the subgroup $A_{\overline{R}^1}$ is generated by the set of vectors $\overline{v}^1 \in \mathbf{Z}^{k-1}$ such that $\overline{v}^1 = \overline{v}^1(s) = (d^2(v_2, s) - d^2(v_1, s), d^2(v_3, s) - d^2(v_1, s), \ldots, d^2(v_k, s) - d^2(v_1, s))$, where $s \in supp(\mu)$, and the subgroup $A_{\overline{R}^2}$ is generated by the set of vectors $\overline{v}^2 \in \mathbf{Z}^{k-1}$ such that $\overline{v}^2 = \overline{v}^2(s) = (d^2(v_1, s) - d^2(v_2, s), d^2(v_3, s) - d^2(v_2, s), \ldots, d^2(v_k, s) - d^2(v_1, s))$, where $s \in supp(\mu)$. Now observe that for every $s \in supp(\mu)$, we have

$$\overline{v}^2(s) = D \cdot \overline{v}^1(s),$$

where D is a $(k-1) \times (k-1)$ matrix

$$D = \begin{pmatrix} -1 & 0 & 0 & 0 & \ldots \\ -1 & 1 & 0 & 0 & \ldots \\ -1 & 0 & 1 & 0 & \ldots \\ -1 & 0 & 0 & 1 & \ldots \\ \ldots & & & & \end{pmatrix}$$

Therefore, $A_{\overline{R}^2} = (A_{\overline{R}^1})^D$. Since the matrix D is invertible it follows that $A_{\overline{R}^1}$ and $A_{\overline{R}^2}$ have the same dimension. □

Theorem 3.25 (multi-vertex SLLN-I)**.** *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^\infty$ be a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. Assume that the weight function M is totally defined and $\mathbb{E}(\xi) = \{v_1, \ldots, v_k\}$, where $k \geq 4$. If the random walk $\overline{R}^1$, associated to v_1, is genuinely $(k-1)$-dimensional, then*

$$\limsup_{n\to\infty} \mathbb{S}(\xi_1, \ldots, \xi_n) = \mathbb{E}(\xi_1)$$

holds with probability 1.

Proof. Since $\overline{R}^1$ is genuinely $(k-1)$-dimensional it follows that the subgroup $A_{\overline{R}^1}$ is isomorphic to $\mathbf{Z}^{k-1}$ and there exists an invertible matrix D that isomorphically maps $A_{\overline{R}^1} \subseteq \mathbf{Z}^{k-1}$ onto $\mathbf{Z}^{k-1}$. We are interested in $\mathbf{P}\Big(\overline{R}^1 \in \mathbf{Z}_+^{k-1},\ \text{i.o.}\Big)$, but, instead, we consider a set $\mathbf{R}_+^{k-1} = \{(x_1, \ldots, x_{k-1}) \mid x_i \geq 0\}$ and observe that

$$\mathbf{P}\Big(\overline{R}^1 \in \mathbf{Z}_+^{k-1},\ \text{i.o.}\Big) = \mathbf{P}\Big(\overline{R}^1 \in \mathbf{R}_+^{k-1},\ \text{i.o.}\Big),$$

since $\overline{R}^1$ ”lives“ in $\mathbf{Z}_+^{k-1}$ only. Let $(\overline{R}^1)^D$ be the induced random walk on $\mathbf{Z}^{k-1}$ by application of D to $\overline{R}^1$. The random walk $(\overline{R}^1)^D$ is aperiodic since D maps $A_{\overline{R}^1}$ onto $\mathbf{Z}^{k-1}$. Moreover, by construction of $(\overline{R}^1)^D$,

$$\mathbf{P}\Big(\overline{R}^1 \in \mathbf{R}_+^{k-1},\ \text{i.o.}\Big) = \mathbf{P}\Big((\overline{R}^1)^D \in (\mathbf{R}_+^{k-1})^D,\ \text{i.o.}\Big).$$

Let S be the simple random walk on $\mathbf{Z}^{k-1}$. Since $(\overline{R}^1)^D$ and S are both aperiodic random walks on $\mathbf{Z}^{k-1}$, it follows from the Invariance Principle (Extension of Wiener's test) that

$$\mathbf{P}\Big((\overline{R}^1)^D \in (\mathbf{R}_+^{k-1})^D \text{ i.o.}\Big) = \mathbf{P}\big(S \in (\mathbf{R}_+^{k-1})^D \text{ i.o.}\big).$$

Clearly, the set $\mathbf{R}_+^{k-1}$ contains a rotated cone and, hence, by Lemma 3.22, its image under an invertible linear transformation D contains a rotated cone too. Now, by Theorem 3.21 and by Lemma 3.23,

$$\mathbf{P}\big(S \in (\mathbf{R}_+^{k-1})^D,\ \text{i.o.}\big) = 1.$$

It follows that $\mathbf{P}\Big(\overline{R}^1 \in \mathbf{Z}_+^{k-1} \text{ i.o.}\Big) = 1$ and, by Lemma 3.18,

$$\mathbf{P}(v_1 \in \limsup_{n\to\infty} \mathbb{S}_n) = 1.$$

Finally, it is proved in Lemma 3.24 that for any $i = 2, \ldots, k$ the random walk $\overline{R}^i$ is genuinely $(k-1)$-dimensional. For any $i = 2, \ldots, k$ we can use the same argument as for v_1 to prove that $\mathbf{P}(v_i \in \limsup_{n\to\infty} \mathbb{S}_n) = 1$. Hence the result.

□

3.3.3 The case when random walk is not genuinely $(k-1)$-dimensional.

The case when $\overline{R}^1$ is not genuinely $(k-1)$-dimensional is more complicated. To answer the question whether v_1 belongs to $\limsup_{n\to\infty} \mathbb{S}_n$ (namely, how often v_1 visits the set $\mathbb{S}_n$), we need to analyze how the space $V_{\overline{R}^1}$ "sits" in $\mathbf{R}^{k-1}$ (in other words, we have to look at its dimension). We know that the subgroup $A_{\overline{R}^1} \subset \mathbf{Z}^{k-1}$ is isomorphic to $\mathbf{Z}^m$, where $m < k-1$ in the case under consideration. Therefore, there exists a $m \times (k-1)$ matrix D which maps the subgroup $A_{\overline{R}^1}$ onto $\mathbf{Z}^m$ and which is injective on $A_{\overline{R}^1}$. Furthermore, the mapping D maps the subspace $V_{\overline{R}^1}$ bijectively onto $\mathbf{R}^m$. The linear mapping D induces an aperiodic random walk $(\overline{R}^1)^D$ on $\mathbf{Z}^m$ in a natural way and

$$\mathbf{P}\Big(\overline{R}^1 \in (\mathbf{R}_+^{k-1}),\ \text{i.o.}\Big) = \mathbf{P}\Big(\overline{R}^1 \in (\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}),\ \text{i.o.}\Big) = \mathbf{P}\Big((\overline{R}^1)^D \in (\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1})^D,\ \text{i.o.}\Big).$$

The main problem here is to understand the structure of the set $(\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1})^D$ and, to be more precise, the structure of the set $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$. Clearly $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ is a monoid, i.e., it contains the trivial element and a sum of any two elements in $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ belongs to $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$. We can define dimension of $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ to be the maximal number of linearly independent vectors in $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$.

Theorem 3.26 (SLLN for "not genuine" dimension)**.** *Suppose $A_{\overline{R}^1} \simeq \mathbf{Z}^m$ and the set $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ has dimension m. Then $\mathbf{P}\Big(v_i \in \limsup_{n\to\infty} \mathbb{S}(\xi_1, \ldots, \xi_n)\Big) = 1$.*

Proof. Since $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ is a monoid of dimension m, we know that $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ contains an m-dimensional rotated cone (see Lemma 3.37 below). Since D is a linear isomorphism from $V_{\overline{R}^1}$ onto $\mathbf{R}^m$ it follows from Lemma 3.22 that $(\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1})^D$ contains an m-dimensional rotated cone in $\mathbf{R}^m$. If S is a simple random walk in $\mathbf{Z}^m$, then

$$\mathbf{P}\big(S \in (\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1})^D \text{ i.o.}\big) = 1.$$

Since S and $(\overline{R}^1)^D$ are both aperiodic, by the extension of Wiener's test (Invariance Principle), we see that

$$\mathbf{P}\Big((\overline{R}^1)^D \in (\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1})^D \text{ i.o.}\Big) = 1.$$

Hence, $\mathbf{P}\Big(\overline{R}^1 \in (\mathbf{R}_+^{k-1}) \text{ i.o.}\Big) = 1$ by the discussion preceding the theorem, and, finally, by Lemma 3.18, $\mathbf{P}(v_i \in \limsup_{n\to\infty} \mathbb{S}_n) = 1$.

□

Next, we investigate under what conditions the subgroup $A_{\overline{R}^1}$ and the set $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ have the same dimension m.

Lemma 3.27. *Assume that $A_{\overline{R}^1}$ contains a positive vector. Then the sets $A_{\overline{R}^1}$ and $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ have the same dimension.*

Proof. Exercise in linear algebra. See Lemma 3.36 below. □

Lemma 3.28. *Assume that $\mu(v_1) \neq 0$. Then $A_{\overline{R}^1}$ and the set $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ have the same dimension.*

Proof. Observe that if $\mu(v_1) \neq 0$ then $A_{\overline{R}^1}$ contains the vector $(d^2(v_2, v_1), \ldots, d^2(v_k, v_1))$ which has all positive coordinates. Therefore, by Lemma 3.27, the set $A_{\overline{R}^1}$ and $\mathbf{R}_+^{k-1} \cap V_{\overline{R}^1}$ have the same dimension. □

Corollary 3.29 (Multi-vertex SLLN - II)**.** *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^\infty$ be a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. Assume that the weight function $M_{\xi_1}(\cdot)$ is totally defined and $\mathbb{E}(\xi) = \{v_1, \ldots, v_k\}$, where $k \geq 4$. If $\mathbb{E}(\xi_1) \subseteq supp(\mu)$, then*

$$\limsup_{n\to\infty} \mathbb{S}(\xi_1, \ldots, \xi_n) = \mathbb{E}(\xi_1)$$

holds with probability 1*.*

Proof. Follows from Lemmas 3.28, 3.27, and Theorem 3.26. □

3.4 Cones, subspaces, and monoids.

3.4.1 Properties of cones

Lemma 3.30. *Let* $C = Cone_\alpha$*. Then for any* $u, v \in C$ *and* $a \geq 0$ *and* $b \geq 0$*,* $au + bv \in C$*, i.e.,* C *is closed under taking nonnegative linear combinations.*

Proof. This is easy to see by application of the discrete version of Minkowski inequality :

$$\Big(\sum_{k=1}^{n} |a_k + b_k|^p\Big)^{\frac{1}{p}} \leq \Big(\sum_{k=1}^{n} |a_k|^p\Big)^{\frac{1}{p}} + \Big(\sum_{k=1}^{n} |b_k|^p\Big)^{\frac{1}{p}}$$

for $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in \mathbf{R}^n$, $1 \leq p < \infty$.

Indeed, $u = (u_1, \ldots, u_n) \in C$ means $\sqrt{u_1^2 + \ldots + u_{n-1}^2} \leq \alpha u_n$ and $v = (v_1, \ldots, v_n) \in C$ means $\sqrt{v_1^2 + \ldots + v_{n-1}^2} \leq \alpha v_n$. Now,

$$\sqrt{(au_1 + bv_1)^2 + \ldots + (au_{n-1} + bv_{n-1})^2} \leq$$

$$\leq \sqrt{(au_1)^2 + \ldots + (au_{n-1})^2} +$$

$$+ \sqrt{(bv_1)^2 + \ldots + (bv_{n-1})^2} \leq$$

$$\leq a\alpha u_n + b\alpha v_n = \alpha(au_n + bv_n)$$

and, thus, $au + bv \in C$. □

Lemma 3.31. *Let* D *be a matrix and* $A \subseteq \mathbf{R}^n$*. If* A *is closed under taking nonnegative linear combinations, then so is* A^D*.*

Proof. D is a linear transformation. □

Corollary 3.32. *Let* D *be a matrix and* $C = Cone_\alpha^D$*. Then* C *is closed under taking nonnegative linear combinations.*

Lemma 3.33. *Let D be an $n \times n$ orthogonal matrix. Then for any $r > 0$ and $v \in \mathbf{R}$ we have*

$$(B_v(r))^D = B_{v^D}(r).$$

In other words any orthogonal matrix shifts a ball of radius r at v to a ball of radius r at v^D.

Proof. Orthogonal matrix preserves distances. □

Lemma 3.34. *Let $A \subseteq \mathbf{R}^n$. The set A contains a rotated cone, $Cone_\alpha^D$, if and only if there exist $B \subseteq A$ such that:*

- *B is closed under taking nonnegative linear combinations;*
- *B contains a ball $B_v(r)$ for some $r > 0$ and $v \in \mathbf{R}^n$.*

Proof. "$\Rightarrow$" Assume that for some $\alpha > 0$ and an orthogonal matrix D, $Cone_\alpha^D \subseteq A$. Clearly, $B = Cone_\alpha^D$ satisfies both of the stated properties.

"$\Leftarrow$" Assume that some $B \subseteq A$ satisfies the properties stated above for some $r > 0$ and $v \in \mathbf{R}^n$. Let D be any orthogonal matrix that maps a point v to a point $Dv = v^D = v' = (0, \ldots, 0, \sqrt{v_1^2 + \ldots + v_n^2})$. By Lemmas 3.31 and 3.33 above, the set B^D is closed under nonnegative linear combinations and contains a ball $B_{v^D}(r) = B_{v'}(r)$. It implies that B^D contains $Cone_\alpha$ where

$$\alpha = \frac{r}{\sqrt{v_1^2 + \ldots + v_n^2}}.$$

Therefore, $Cone_\alpha \subseteq B^D \subseteq A^D$. If $E = D^{-1}$, then $Cone_\alpha^E \subseteq A$. □

Lemma 3.35. *Assume that a set $A \subseteq \mathbf{R}^n$ contains a rotated cone. Then for any invertible $n \times n$ matrix D, the set A^D contains a rotated cone.*

Proof. Assume that for some $\alpha > 0$ and orthogonal D', $Cone_\alpha^{D'} \subseteq A$. Then by Lemma 3.34 there is $B \subseteq A$ which closed under taking nonnegative linear combinations and contains some ball $B_v(r)$.

Since $B \subseteq A$, we have $B^D \subseteq A^D$. As proved in Lemma 3.31, B^D is closed under taking nonnegative sums. Even though $(B_v(r))^D$ is not a ball for a general invertible matrix, it still contains a ball in itself with the center at v^D and some radius, which depends on eigenvalues of D. Thus, B^D meets the assumptions of Lemma 3.34 and, hence, it contains a rotated cone. Since $B^D \subseteq A^D$, A^D contains a rotated cone too. □

3.4.2 Subspaces and monoids

Let us accept the following notation for this section of Appendix A only:

- $S \subset \mathbf{Z}^n$,
- $A = \langle S \rangle \subseteq \mathbf{Z}^n$ is a subgroup of $\mathbf{Z}^n$ generated by S,
- $V \subseteq \mathbf{R}^n$ is a vector subspace spanned by S.

Lemma 3.36. *Assume that A contains a positive vector. Then A and $\mathbf{R}^n_+ \cap V$ have the same dimension.*

Proof. Clearly $\mathbf{R}^n_+ \cap V$ cannot have bigger dimension than A, because V is spanned by the same vectors as A.

Conversely, let $\{v_1, \ldots, v_k\}$ be a basis for A and $v = a_1v_1 + \ldots + a_kv_k \in A$ be a positive vector. Then for some sufficiently large $m \in \mathbb{N}$, the vectors $\{mv + v_1, \ldots, mv + v_k\}$ belong to $\mathbf{R}^n_+ \cap A \subseteq \mathbf{R}^n_+ \cap V$. In addition, it is not hard to see that these vectors are linearly independent. Indeed, consider an $n \times k$ matrix M, columns of which are the vectors $v_1, \ldots, v_k$. Adding the vector mv to $v_1, \ldots, v_k$ corresponds to multiplication of the matrix M by the $k \times k$ matrix

$$M_m = \begin{bmatrix} ma_1 + 1 & ma_1 & ma_1 & \ldots \\ ma_2 & ma_2 + 1 & ma_2 & \ldots \\ ma_3 & ma_3 & ma_3 + 1 & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}.$$

Now, consider a polynomial function $m \mapsto |M_m|$, where $|\cdot|$ stays for determinant. Clearly, $0 \mapsto 1$ and hence, this function is not trivial. Since non-trivial polynomial function cannot have infinitely many roots, the matrix M cannot be degenerate for infinitely many m's. Therefore, for some sufficiently large m, $|M_m| \neq 0$, and the vectors $\{mv + v_1, \ldots, mv + v_k\}$ remain independent.

Therefore, the dimension of $\mathbf{R}_+^n \cap V$ is not smaller than that of A.

□

Lemma 3.37. *Assume that a set $B \subseteq \mathbf{R}^k$ is closed under taking nonnegative linear combinations and contains k independent vectors. Then B contains a rotated cone.*

Proof. If we show that B contains a ball, then both assumptions of Lemma 3.34 will be satisfied, and the result follows.

Let $\{v_1, \ldots, v_k\}$ be independent vectors in B. Consider a mapping $\varphi : \mathbf{R}^k \to \mathbf{R}^k$ defined by

$$(\alpha_1, \ldots, \alpha_k) \overset{\varphi}{\mapsto} \alpha_1 v_1 + \ldots \alpha_k v_k.$$

Clearly, φ is a linear automorphism of $\mathbf{R}^k$ and it has an inverse automorphism ψ. Notice that both maps φ and ψ are nice continuous mappings. For $i = 1, \ldots, k$ define the usual projection functions $\pi_i : \mathbf{R}^k \to \mathbf{R}$, which are continuous.

Now, consider a point $v_0 = v_1 + \ldots + v_k$. We claim that there exists $\varepsilon > 0$ such that $B_{v_0}(\varepsilon) \subseteq B$. Indeed, note that for $i = 1, \ldots, k$, $\pi_i(\psi(v_0)) = 1$. Moreover, $\pi_i \circ \psi$ is a continuous mapping and, hence, for some $\varepsilon_i > 0$ and any $v \in B_{v_0}(\varepsilon_i)$, $\pi_i(\psi(v)) \subset (\frac{1}{2}, \frac{3}{2})$. Put $\varepsilon = \min\{\varepsilon_1, \ldots, \varepsilon_k\}$. By the choice of ε, for any $v \in B_{v_0}(\varepsilon)$, $\pi_i(\psi(v)) \in (\frac{1}{2}, \frac{3}{2})$, i.e., any $v \in B_{v_0}(\varepsilon)$ is a positive linear combination of vectors $\{v_1, \ldots, v_k\}$. It follows that $B_{v_0}(\varepsilon) \subset B$, and, by Lemma 3.34, B contains a rotated cone.

□

3.5 Example of random walk for a cyclic graph.

Let μ be a probability measure on a locally finite graph Γ. Assume that the function $M^{(c)}$ is defined for Γ. Define a $|V(\Gamma)| \times |V(\Gamma)|$ matrix $D^{(c)}$ of cth powers of distances between vertices in Γ. Clearly,

$$M^{(c)} = D^{(c)}\mu.$$

The purpose of this section is to work out a particular example of a random walk construction for the cyclic graph of length 4, see Figure 3.2. Here we assume that all four vertices are centers with $\mu_i > 0$, $1 \le i \le 4$. The corresponding distance matrix

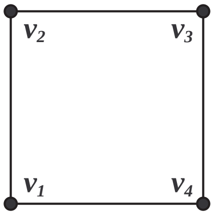

Figure 3.2: Cycle-4.

$D^{(2)}$ with $D^{(2)}(i,j) = d^2(v_i, v_j), 1 \le i, j \le 4$ is

$$D^{(2)} = \begin{pmatrix} 0 & 1 & 4 & 1 \\ 1 & 0 & 1 & 4 \\ 4 & 1 & 0 & 1 \\ 1 & 4 & 1 & 0 \end{pmatrix}$$

In order to show, for instance, that $v_1 \in \limsup \mathbb{S}_n$, we can look at the recurrence property of a corresponding random walk $\overline{R}$ in $\mathbf{Z}^3$, which we are going to construct now.

According to the above discussion, we need to write down a matrix of increments (steps) $\mathbf{I}_\zeta$ of our walk that we obtain by subtracting the first row of $D^{(2)}$ from the rest of the rows and removing the first row (since we do not consider v_1 relative to itself). We obtain 3×4 matrix of increments

$$\mathbf{I}_\zeta = \begin{pmatrix} 1 & -1 & -3 & 3 \\ 4 & 0 & -4 & 0 \\ 1 & 3 & -3 & -1 \end{pmatrix}$$

To be consistent with the notation used in the section on (two-) three-point center-sets, we denote increments in $\mathbf{Z}^3$ by $\zeta(s) = (\zeta_1(s), \zeta(s)_2, \zeta_3(s)), s \in V(\Gamma)$. For this particular example we have

$$\zeta_1(s) = d^2(v_2, s) - d^2(v_1, s)$$

$$\zeta_2(s) = d^2(v_3, s) - d^2(v_1, s)$$

$$\zeta_3(s) = d^2(v_4, s) - d^2(v_1, s)$$

$$s = v_1, v_2, v_3, v_4.$$

It is even better to look at the matrix $\mathbf{I}_\zeta$ as the following table (see Table 3.1).

$\zeta_i(s) \backslash s$	v_1	v_2	v_3	v_4
$\zeta_1(s)$	1	-1	-3	3
$\zeta_2(s)$	4	0	-4	0
$\zeta_3(s)$	1	3	-3	-1

Table 3.1: Increments of the random walk

Here, each column gives us a vector $\zeta(s) = (\zeta_1(s), \zeta(s)_2, \zeta_3(s))$ – a possible increment step in $\mathbf{Z}^3$ with $\mu(\zeta(s)) = \mu(s)$, where $s \in V(\Gamma)$. These vectors define a random walk $\overline{R}$ in $\mathbf{Z}^3$ which is genuinely 3-dimensional. Thus, SLLN holds by Theorem 3.25

Chapter 4

Chebyshev's inequality on graphs

4.1 Introduction

Every area of mathematics in general, as well as every trend of probability theory in particular, possesses some important inequalities. For instance, inequalities, as humble as they may seem, often provide necessary bounds and are at the heart of the matter of proving many theorems. One of such inequalities in classical probability theory is due to Pafnuty L. Chebyshev. It asserts that if ξ is a random variable with $\mathbb{E}(\xi^2) < \infty$, then for any $\varepsilon > 0$, we have

$$\mathbf{P}\Big(|\xi - \mathbb{E}(\xi)| \geq \varepsilon\Big) \leq \frac{\sigma^2}{\varepsilon^2}, \tag{4.1}$$

where $\sigma^2 = Var(\xi)$. This inequality can be found in any classical probability theory text, in particular, in (3).

The inequality applied to the sample mean random variable $\overline{X} = \frac{S_n}{n}$, where $S_n = \xi_1 + \ldots + \xi_n$, $\mathbb{E}(\xi_i) = m, Var(\xi_i) = \sigma^2, i = 1, \ldots, n$ results in

$$\mathbf{P}\Big(|\overline{X} - m| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n\varepsilon^2} \tag{4.2}$$

Chebyshev discovered it when he was trying to prove the law of large numbers, and the inequality is widely used ever since. We can think of Chebyshev's inequality as a result concerning the concentration of measure, giving a quantitative description of this concentration. Indeed, it provides a bound on the probability that a value of a random variable ξ with finite mean and variance will differ from the mean by more than a fixed number ε. In other words, we have a crude estimate for concentration of probabilities around the expectation, and this estimate has a big theoretical significance.

In this chapter we prove an analog of the classical Chebyshev's inequality - the concentration of measure inequality for a graph- (group-)valued random element ξ. The usual setting is as follows. We consider the image of the given probability space under the mapping $\xi(\cdot) : \Omega \to V(\Gamma)$; namely, we work with a discrete probability space $(V(\Gamma), \mathcal{S}, \mu)$, and we remember that for every fixed v, $d^2(v, \xi)$ is a real-valued random variable, i.e., $d^2(v, \cdot) : V(\Gamma) \to \mathbf{R}$.

4.2 Concentration of measure inequality on graphs (groups)

First, let us prove the following lemma, which is going to be useful in the proof of the main theorem below.

Lemma 4.1. *Let μ be a distribution on a locally finite graph Γ such that $M \equiv M^{(2)}$ is defined. If for some $r \in \mathbb{N}$ and $v_0 \in V(\Gamma)$ the inequality*

$$\sum_{s \in V(\Gamma) \setminus B_{v_0}(r/2)} d(v_0, s)\mu(s) - \frac{r}{2}\mu(v_0) < 0 \tag{4.3}$$

holds, then for any $u \in V(\Gamma) \setminus B_{v_0}(r)$, $M(u) > M(v_0)$.

Proof. Indeed, pick any $u \in V(\Gamma) \setminus B_{v_0}(r)$ and let $d = d(v_0, u)$. Then

$$\begin{aligned} M(u) - M(v_0) &= \sum_{s \in V(\Gamma)} \Big(d^2(u, s) - d^2(v_0, s)\Big)\mu(s) \\ &\geq d^2\mu(v_0) - \sum_{d(v_0,s) > d(u,s)} \Big(d^2(v_0, s) - d^2(u, s)\Big)\mu(s) \end{aligned}$$

$$\geq d^2\mu(v_0) - \sum_{d(v_0,s)>d(u,s)} (d(v_0,s)-d(u,s))(d(v_0,s)+d(u,s))\mu(s)$$

$$\geq d^2\mu(v_0) - 2d \sum_{d(v_0,s)>d(u,s)} d(v_0,s)\mu(s)$$

$$\geq d^2\mu(v_0) - 2d \sum_{s\in V(\Gamma)\setminus B_{v_0}(r/2)} d(v_0,s)\mu(s).$$

Since $d > r$, it follows from the assumption of the lemma that we get a positive quantity at the end. To make it more clear, let us note that in the last estimate, we used the observation that

$$\Big\{s \in V(\Gamma) |\ d(v_0,s) > d(u,s)\Big\} \bigcap B_{v_0}(r/2) = \emptyset,$$

which implies that $\Big\{s \in V(\Gamma) |\ d(v_0,s) > d(u,s)\Big\} \subseteq V(\Gamma) \setminus B_{v_0}(r/2)$ and, therefore,

$$\sum_{d(v_0,s)>d(u,s)} d(v_0,s)\mu(s) \leq \sum_{s\in V(\Gamma)\setminus B_{v_0}(r/2)} d(v_0,s)\mu(s).$$

We conclude that $M(u) > M(v_0)$ as required.

□

4.2.1 Singleton mean-set

Theorem 4.2. *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^\infty$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If the weight function $M_{\xi_1}(\cdot)$ is totally defined and $\mathbb{E}(\xi_1) = \{v\}$ for some $v \in V(\Gamma)$ then there exists a constant $C = C(\Gamma, \xi_1) > 0$ such that*

$$\mathbf{P}\Big(\mathbb{S}(\xi_1, \ldots, \xi_n) \neq \{v\}\Big) \leq \frac{C}{n}, \tag{4.4}$$

for a sample of random elements $\xi_1(\omega), \ldots, \xi_n(\omega)$ of size n.

Proof. Observe that, by the definition of the sample mean-set, we can rewrite the event in question in terms of sample weight functions $M_n(\cdot)$:

$$\Big\{\mathbb{S}_n \neq \{v\}\Big\} = \Big\{\exists u \in V(\Gamma) \setminus \{v\},\quad M_n(u) \leq M_n(v)\Big\}.$$

Therefore, in order to prove the theorem, it suffices to bound the probability of this equivalent representation of the event. We show that

$$\mathbf{P}\Big(\exists u \in V(\Gamma) \setminus \{v\},\quad M_n(u) \le M_n(v)\Big) \le \frac{C}{n} \tag{4.5}$$

for some constant C, in two stages; namely, we prove that for some $v_0 \in V(\Gamma)$ with $\mu(v_0) > 0$ and constants $r \in \mathbb{N}$, $C_1, C_2 \in \mathbf{R}$ such that $v \in B_{v_0}(r)$, the inequalities

$$\mathbf{P}\Big(\exists u \in B_{v_0}(r) \setminus \{v\},\quad M_n(u) \le M_n(v)\Big) \le \frac{C_1}{n} \tag{4.6}$$

and

$$\mathbf{P}\Big(\exists u \in V(\Gamma) \setminus B_{v_0}(r),\quad M_n(u) \le M_n(v_0)\Big) \le \frac{C_2}{n} \tag{4.7}$$

hold. It is not hard to see that if we find C_1 and C_2 satisfying (4.6) and (4.7) respectively, then (4.4) holds for $C = C_1 + C_2$ and the theorem is proved.

Indeed, consider the following events:

$$D_n = \Big\{\exists u \in V(\Gamma) \setminus \{v\},\quad M_n(u) \le M_n(v)\Big\}, \text{ as in (4.5)},$$

$$A_n = \Big\{\exists u \in B_{v_0}(r) \setminus \{v\},\quad M_n(u) \le M_n(v)\Big\}, \text{ as in (4.6)},$$

$$B_n = \Big\{\exists u \in V(\Gamma) \setminus B_{v_0}(r),\quad M_n(u) \le M_n(v) \text{ and } \nexists u \in B_{v_0}(r) \setminus \{v\},\quad M_n(u) \le M_n(v)\Big\},$$

and

$$E_n = \Big\{\exists u \in V(\Gamma) \setminus B_{v_0}(r),\quad M_n(u) \le M_n(v_0)\Big\}, \text{ as in (4.7)}.$$

Clearly, $D_n = A_n \bigsqcup B_n$ is a disjoint union, and $\mathbf{P}(D_n) = \mathbf{P}(A_n) + \mathbf{P}(B_n)$. Now, observe that for any $u, v_0, v \in V(\Gamma)$, if $M_n(u) \le M_n(v)$ then either $M_n(u) \le M_n(v_0)$ or $M_n(v_0) \le M_n(v)$. In particular, this is true on the event B_n. But on this event, $M_n(v_0)$ cannot be smaller than $M_n(v)$, because it would contradict the second property of B_n with u taken to be $v_0 \in B_{v_0}(r) \setminus \{v\}$. Hence, for any $\omega \in B_n$, if $M_n(u) \le M_n(v)$, then $M_n(u) \le M_n(v_0)$, and, consequently,

$$B_n \subseteq E_n.$$

Thus,

$$\mathbf{P}(D_n) = \mathbf{P}(A_n) + \mathbf{P}(B_n) \leq \mathbf{P}(A_n) + \mathbf{P}(E_n),$$

and it is sufficient to prove (4.6) and (4.7), as claimed.

First we argue (4.7). Choose any $v_0 \in V(\Gamma)$ such that $\mu(v_0) > 0$ and $r \in \mathbb{N}$ such that the inequality (4.3) holds, i.e.,

$$\sum_{s \in V(\Gamma) \setminus B_{v_0}(r/2)} d(v_0, s)\mu(s) - \frac{r}{2}\mu(v_0) < 0.$$

We can choose such r since $M^{(1)}(v_0)$ is finite. Observe that the left hand side of the inequality above is the expectation (with respect to the measure μ) of a random variable $\eta : V \to \mathbf{R}$ defined as

$$\eta(s) := d(v_0, s)\mathbf{1}_{V(\Gamma)\setminus B_{v_0}(r/2)}(s) - \frac{r}{2}\mathbf{1}_{v_0}(s),\ \ s \in V(\Gamma)$$

where

$$\mathbf{1}_A(s) = \begin{cases} 1, & \text{if } s \in A; \\ 0, & \text{if } s \notin A \end{cases}$$

is a usual indicator function with $A \subseteq \mathcal{S}$ and $s \in V(\Gamma)$.

Since by our assumption $M \equiv M^{(2)}$ is defined, it follows that $\sigma^2(\eta) < \infty$, and we can apply classical Chebyshev inequality to

$$\overline{\eta} = \frac{\sum_{i=1}^{n} \eta(s_i)}{n} = \left(\sum_{s \in V(\Gamma) \setminus B_{v_0}(r/2)} d(v_0, s)\mu_n(s) \right) - \frac{r}{2}\mu_n(v_0).$$

Now, observe that, since $\mathbb{E}(\eta) < 0$, by our choice of v_0 and r, the event

$$\Big\{ \Big| \sum_{s \in V(\Gamma) \setminus B_{v_0}(r/2)} d(v_0, s)\mu_n(s) - \frac{r}{2}\mu_n(v_0) - \mathbb{E}\eta \Big| < |\mathbb{E}\eta|/2 \Big\}$$

implies that $\displaystyle\sum_{s \in V(\Gamma) \setminus B_{v_0}(r/2)} d(v_0, s)\mu_n(s) - \frac{r}{2}\mu_n(v_0) < 0$, and, by Lemma (4.1), it follows that for any $u \in V(\Gamma) \setminus B_{v_0}(r)$, we have $M_n(u) > M_n(v_0)$. Thus

$$\Big\{ \Big| \sum_{s \in V(\Gamma) \setminus B_{v_0}(r/2)} d(v_0, s)\mu_n(s) - \frac{r}{2}\mu_n(v_0) - \mathbb{E}\eta \Big| < |\mathbb{E}\eta|/2 \Big\}$$

$$\subseteq \Big\{ \forall u \in V(\Gamma) \setminus B_{v_0}(r), \quad M_n(u) > M_n(v_0) \Big\}.$$

Taking complements of the above events as well as employing Chebyshev inequality (4.8) applied to $\overline{\eta}$ with $\varepsilon = |\mathbb{E}\eta|/2$, we can obtain the following estimate:

$$\mathbf{P}\Big(\exists u \in V(\Gamma) \setminus B_{v_0}(r), \quad M_n(u) \le M_n(v_0) \Big)$$

$$\le \mathbf{P}\Big(\Big| \sum_{s \in V(\Gamma) \setminus B_{v_0}(r/2)} d(v_0, s)\mu_n(s) - \frac{r}{2}\mu_n(v_0) - \mathbb{E}\eta \Big| \ge |\mathbb{E}\eta|/2 \Big) \le \frac{4\sigma^2(\eta)}{n|\mathbb{E}\eta|^2}.$$

Hence, inequality (4.7) holds for $C_2 = C_2(r, v_0, \mu) = \frac{4\sigma^2(\eta)}{|\mathbb{E}\eta|^2}$.

To prove (4.6) we notice that for any $u \in V(\Gamma) \setminus \{v\}$,

$$M(u) - M(v) = \sum_{s \in V(\Gamma)} (d(u,s) - d(v,s))(d(u,s) + d(v,s))\mu(s),$$

i.e., $M(u) - M(v)$ is the expectation of a random variable $\tau : V \to \mathbf{R}$ defined as

$$\tau_{u,v}(s) := (d(u,s) - d(v,s))(d(u,s) + d(v,s)), \ s \in V(\Gamma).$$

Furthermore, since $M_{\xi_1}(\cdot)$ is defined by assumption and because for every $s \in V(\Gamma)$, we have $d(u,s) - d(v,s) \le d(v,u)$, it is easy to see that $\sigma^2(\tau_{u,v}(s)) < \infty$ and, therefore, classical Chebyshev inequality applies to

$$\overline{\tau} = \frac{\sum_{i=1}^n \tau(s_i)}{n} = M_n(u) - M_n(v).$$

Thus,

$$\mathbf{P}\Big(|M_n(u) - M_n(v) - (M(u) - M(v))| \ge \varepsilon \Big) \le \frac{\sigma^2(\tau_{u,v}(s))}{n\varepsilon^2}.$$

holds. Now, if $0 < \varepsilon < M(u) - M(v)$ then

$$\mathbf{P}\Big(M_n(u) < M_n(v) \Big) \le \mathbf{P}\Big(|M_n(u) - M_n(v) - (M(u) - M(v))| \ge \varepsilon \Big),$$

by the reasoning that we used above with taking complements of sets.

Finally, we choose ε to be $\frac{1}{2}\inf\{M(u) - M(v) \mid u \in B_{v_0}(r) \setminus \{v\}\}$ and, using σ additivity (or, rather, just additivity in our case) of measure, we can write

$$\mathbf{P}\Big(\exists u \in B_{v_0}(r) \setminus \{v\}, \quad M_n(u) \le M_n(v) \Big)$$

$$\le \sum_{u \in B_{v_0}(r) \setminus \{v\}} \mathbf{P}\Big(M_n(u) \le M_n(v) \Big)$$

$$\le \frac{\sum_{u \in B_{v_0}(r)} \sigma^2(\tau_{u,v}(s))}{n\varepsilon^2}$$

we conclude that inequality (4.6) holds for the constant

$$C_1 = \varepsilon^{-2} \sum_{u \in B_{v_0}(r)} \sigma^2(\tau_{u,v}(s)).$$

□

4.2.2 Multi-vertex mean-set

In fact, one can easily generalize the previous theorem to the following statement for the multi-vertex mean-sets.

Theorem 4.3. *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^{\infty}$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$. If the weight function $M_{\xi_1}(\cdot)$ is totally defined then there exists a constant $C = C(\Gamma, \xi_1) > 0$ such that*

$$\mathbf{P}\Big(\mathbb{S}(\xi_1, \ldots, \xi_n) \not\subseteq \mathbb{E}(\xi) \Big) \le \frac{C}{n}. \tag{4.8}$$

Proof. Suppose $\mathbb{E}(\xi) = \{v_1, \ldots, v_k\}$. As before, by definition of the sample mean-set,

$$\Big\{ \mathbb{S}_n \not\subseteq \mathbb{E}(\xi) \Big\} = \Big\{ \exists u \in V(\Gamma) \setminus \mathbb{E}(\xi), \quad M_n(u) \le M_n(v_i) \ \forall i = 1, \ldots k \Big\}.$$

Clearly,

$$\Big\{ \exists u \in V(\Gamma) \setminus \mathbb{E}(\xi), \quad M_n(u) \le M_n(v_i) \ \forall i = 1, \ldots k \Big\} \subseteq \Big\{ \exists u \in V(\Gamma) \setminus \mathbb{E}(\xi), \quad M_n(u) \le M_n(v_1) \Big\},$$

and we have the result by the Theorem 4.2.

□

Chapter 5

Computations, applications, and experiments

5.1 Configurations of mean-sets

In this section we demonstrate several configurations of mean-sets on graphs and how they lead to results that allow us to make some implications about trees and free groups. In addition, considerations of this section help us in dealing with computational problems. First, we make a simple observation stated in the lemma below.

Lemma 5.1. *Let Γ be a connected graph. Then for any $v \in V(\Gamma)$ there exists a measure μ such that $\mathbb{E}\mu = \{v\}$.*

Proof. Indeed, the statement of the lemma holds for the distribution defined by

$$\mu(u) = \begin{cases} 1, & \text{if } u = v; \\ 0, & \text{otherwise.} \end{cases}$$

□

On the other hand, it is easy to see that not any subset of $V(\Gamma)$ can be realized as $\mathbb{E}\mu$. For instance, consider a graph as in Figure 5.1. Let $\mu_0 = \mu(v_0)$, $\mu_1 = \mu(v_1)$,

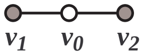

Figure 5.1: Impossible configuration of centers (gray vertices).

$\mu_2 = \mu(v_2)$, $M_0 = M(v_0)$, $M_1 = M(v_1)$, $M_2 = M(v_2)$ Then

$$M_1 = \mu_0 + 4\mu_2,$$

$$M_0 = \mu_1 + \mu_2,$$

$$M_2 = 4\mu_1 + \mu_0.$$

Clearly, for no values of μ_0, μ_1, and μ_2 both inequalities $M_0 > M_1$ and $M_0 > M_2$ can hold simultaneously (since we can not have $2M_0 > M_1 + M_2$). Thus, v_1 and v_2 can not comprise $\mathbb{E}(\mu)$. In fact a tree can have only a limited configuration of centers as proved in Proposition 5.8 below.

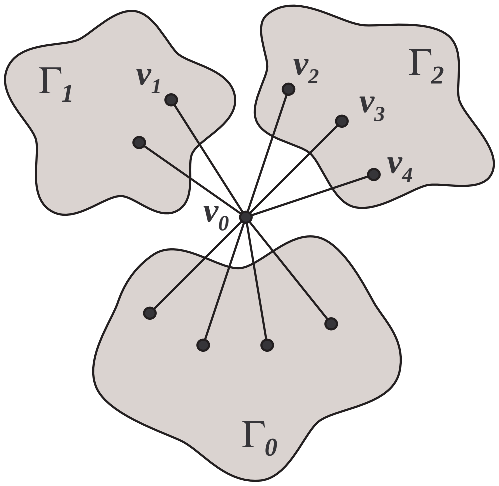

Figure 5.2: A graph with a cut point v_0.

Let Γ be a graph and $v_0 \in V(\Gamma)$. We say that v_0 is a *cut-point* if removing v_0 from Γ results into a disconnected graph (see Figure 5.2). The same definition applies to any metric space (Γ, d). It turns out that existence of a *cut-point* in Γ affects possible configurations of center-sets dramatically. The following Lemma provides a useful inequality that holds for any metric space with a cut-point.

Lemma 5.2 (Cut-point inequality)**.** *Let (Γ, d) be a metric space and v_0 a cut point in Γ. If v_1, v_2 belong to distinct connected components of $\Gamma - \{v_0\}$ then for any $s \in V(\Gamma)$ the inequality*

$$d(v_0, v_2)(d^2(v_1, s) - d^2(v_0, s)) + d(v_0, v_1)(d^2(v_2, s) - d^2(v_0, s)) \geq C > 0 \qquad (5.1)$$

holds, where $C = C(v_0, v_1, v_2) = d(v_0, v_2)d(v_0, v_1)(d(v_0, v_1) + d(v_0, v_2))$.

Proof. Denote the left hand side of (5.1) by $g(s)$. There are 3 cases to consider.

Case 1. Assume that s does not belong to the components of v_1 and v_2. Then $d^2(v_1, s) = [d(v_1, v_0) + d(v_0, s)]^2 = d^2(v_1, v_0) + 2d(v_1, v_0)d(v_0, s) + d^2(v_0, s)$. With this in mind, we get

$$d(v_0, v_2)\big(d^2(v_1, s) - d^2(v_0, s)\big) + d(v_0, v_1)\big(d^2(v_2, s) - d^2(v_0, s)\big)$$

$$= d(v_0, v_2)d(v_0, v_1)(2d(v_0, s) + d(v_0, v_1)) + d(v_0, v_1)d(v_0, v_2)(2d(v_0, s) + d(v_0, v_2))$$

$$= d(v_0, v_2)d(v_0, v_1)(4d(v_0, s) + d(v_0, v_1) + d(v_0, v_2))$$

$$\geq d(v_0, v_2)d(v_0, v_1)(d(v_0, v_1) + d(v_0, v_2))$$

and hence (5.1) holds.

Case 2. Assume that s belongs to the component of v_1. Define

$$x := x(s) = d(v_1, s) \text{ and } y := y(s) = d(v_0, s).$$

In this notation we get

$$d^2(v_2, s) = [y + d(v_0, v_2)]^2 = y^2 + 2yd(v_0, v_2) + d^2(v_0, v_2),$$

and

$$g(s) = g(x, y) = d(v_0, v_2)\big(x^2 - y^2\big) + d(v_0, v_1)\big(d^2(v_2, s) - y^2\big) =$$

$$= d(v_0, v_2)\big(x^2 - y^2\big) + d(v_0, v_1)\big(2yd(v_0, v_2) + d^2(v_0, v_2)\big).$$

Dividing by a positive value $d(v_0, v_2)$, we get

$$g(s) > 0 \quad \text{if and only if} \quad \frac{g(x, y)}{d(v_0, v_2)} = x^2 - y^2 + d(v_0, v_1)(2y + d(v_0, v_2)) > 0.$$

Now, observe that the numbers x, y, and $d(v_0, v_1)$ satisfy triangle inequalities

$$\begin{cases} x + y \geq d(v_0, v_1); \\ x + d(v_0, v_1) \geq y; \\ y + d(v_0, v_1) \geq x; \end{cases}$$

which bound the area visualized in Figure 5.3. The function of two variables $\frac{g(x,y)}{d(v_0,v_2)}$ attains the minimal value $d^2(v_0, v_1)+d(v_0, v_1)d(v_0, v_2)$ on the boundary of the specified area. Hence the inequality

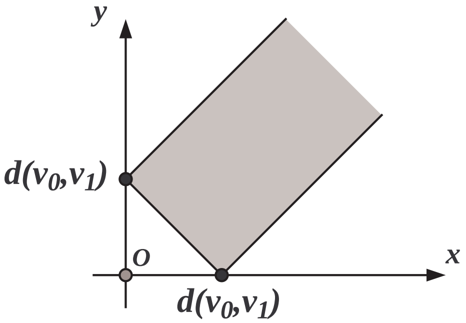

Figure 5.3: Area of possible triangle side lengths.

$$g(s) \geq d(v_0, v_2)d(v_0, v_1)(d(v_0, v_1) + d(v_0, v_2))$$

holds for any s in the component of v_1.

Case 3. If s belongs to the component of v_2 then using same arguments as for the previous case one shows that (5.1) holds.

□

Corollary 5.3. *Let Γ be a connected graph, v_0 a cut-point in Γ, and v_1, v_2 belong to distinct components of $\Gamma - \{v_0\}$. Then the inequality*

$$d(v_0, v_2)(M(v_1) - M(v_0)) + d(v_0, v_1)(M(v_2) - M(v_0)) \geq C > 0$$

holds, where $C = C(v_0, v_1, v_2) = d(v_0, v_2)d(v_0, v_1)(d(v_0, v_1) + d(v_0, v_2))$.

Proof. Indeed,

$$\begin{aligned} & d(v_0, v_2)(M(v_1) - M(v_0)) + d(v_0, v_1)(M(v_2) - M(v_0)) \\ = & \sum_{s \in V(\Gamma)} \Big(d(v_0, v_2)\big(d^2(v_1, s) - d^2(v_0, s)\big) + d(v_0, v_1)\big(d^2(v_2, s) - d^2(v_0, s)\big)\Big)\mu(s) \\ \geq & \sum_{s \in V(\Gamma)} C\mu(s) = C = d(v_0, v_2)d(v_0, v_1)(d(v_0, v_1) + d(v_0, v_2)) \end{aligned}$$

by Lemma 5.2 □

The following result is very important for future developments concerning local properties of our weight function M.

Corollary 5.4. (Cut Point Lemma) *Let Γ be a connected graph, v_0 a cut-point in Γ. If v_1 and v_2 belong to distinct connected components of $\Gamma - \{v_0\}$, then the inequalities $M(v_0) \geq M(v_1)$ and $M(v_0) \geq M(v_2)$ cannot hold simultaneously.*

Proof. Assume to the contrary that $M(v_0) \geq M(v_1)$ and $M(v_0) \geq M(v_2)$ hold simultaneously which is equivalent to

$$M(v_1) - M(v_0) \leq 0 \text{ and } M(v_2) - M(v_0) \leq 0.$$

Then, multiplying by positive constants and adding the inequalities above, we get

$$d(v_0, v_2)(M(v_1) - M(v_0)) + d(v_0, v_1)(M(v_2) - M(v_0)) \leq 0$$

which is impossible by Corollary 5.3. This contradiction finishes the proof.

□

The following series of corollaries together with Proposition 5.8 are featuring some theoretical applications of our theory to trees, free groups, and a free product of finitely generated groups.

Corollary 5.5. (Mean-set in a graph with a cut-point) *Let v_0 be a cut-point in a graph Γ and $\Gamma - \{v_0\}$ a disjoint union of connected components $\Gamma_1, \ldots, \Gamma_k$. Then for any distribution μ on Γ there exists a unique $i = 1, \ldots, k$ such that $\mathbb{E}(\mu) \subseteq V(\Gamma_i) \cup \{v_0\}$.*

Corollary 5.6. (Mean-set in a graph with several cut-points) *Let $v_1, \ldots, v_n$ be cut-points in a graph Γ and $\Gamma - \{v_1, \ldots, v_n\}$ a disjoint union of connected components $\Gamma_1, \ldots, \Gamma_k$. Then for any distribution μ on Γ there exists a unique $i = 1, \ldots, k$ such that $\mathbb{E}(\mu) \subseteq V(\Gamma_i) \cup \{v_1, \ldots, v_n\}$.*

Corollary 5.7. *Let G_1 and G_2 be finitely generated groups and $G = G_1 * G_2$ a free product of G_1 and G_2. Then for any distribution μ on G the set $\mathbb{E}(\mu)$ is a subset of elements of the forms gG_1 or gG_2 for some element $g \in G$.*

Note: every point in a Cayley graph of $G = G_1 * G_2$ is a cut point.

Proposition 5.8. *Let Γ be a tree and μ a probability measure on $V(\Gamma)$. Then $|\mathbb{E}\mu| \leq 2$. Moreover, if $\mathbb{E}(\mu) = \{u, v\}$ then u and v are adjacent in Γ.*

Proof. Observe that any points v_1, v_0, v_2 such that v_0 is connected to v_1 and v_2 satisfy the assumptions of Cut Point Lemma (Corollary 5.4). Assume that $v_0 \in \mathbb{E}(\mu)$. At most one of the the neighbors of v_0 can belong to $\mathbb{E}(\mu)$, otherwise we would have 3 connected vertices with equal M values which contradicts Cut Point Lemma.

□

Corollary 5.9. *Let μ be a probability distribution on a free group F. Then $|\mathbb{E}\mu| \leq 2$.*

Remark 5.10. (On interpretation and benefits of Cut Point Lemma.) Let us make a brief and informal remark on how the Cut Point Lemma (Corollary 5.4), proved above, is useful. When we are dealing with any graph Γ satisfying conditions of Cut Point Lemma in general, and with any tree, in particular, we can see that existence of proper local minima of the *weight function* M in such Γ is impossible since M cannot have a jump of values around cut points. As we shall see in the Section 5.2, it makes the computation of mean-sets for trees manageable – we can easily find the global minimum without taking risks of being "lost" in the vicinity of just a local one.

In general, the number of central points is unlimited. To see this, consider the complete graph K_n on n vertices and let μ be a uniform probability distribution on $V(K_n)$. Clearly $\mathbb{E}(\mu) = V(K_n)$. Another example of the same type is a cyclic graph C_n on n vertices with a uniform probability distribution μ on $V(C_n)$. Clearly $\mathbb{E}(\mu) = V(C_n)$.

In all previous examples, the centers in a graph formed a connected subgraph. This is not always the case. See for instance the graph in Figure 5.4. In this figure, each vertex marked by black has probability 0.1, others have probability 0. Gray vertices are centers. One can construct similar graphs with as many centers as required and property that distances between centers are very large (as large as one wishes).

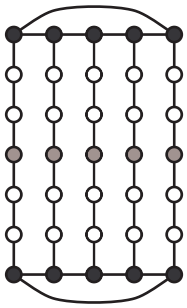

Figure 5.4: Another example of configuration of centers (gray vertices)

5.2 Computation of mean-sets

In this section we discuss computational issues that we face in practice. Let G be a group and $\{\xi\}_{i=1}^{n}$ a sequence of random i.i.d. elements taking values in G such that the weight function is totally defined. One of the technical difficulties is that, unlike the average value S_n/n for real-valued random variables, the sample mean-set $\mathbb{S}_n \equiv \mathbb{S}(\xi_1, \ldots, \xi_n)$ is hard to compute. In other words, our sample mean-set might not be efficiently computable in general. Several problems arise when trying to compute $\mathbb{S}_n$:

- Straightforward computation of the set $\{M(g) \mid g \in G\}$ requires $O(|G|^2)$ steps. This is computationally infeasible for large groups G, and simply impossible for infinite groups. Hence we might want to reduce the search of a minimum to some small part of G.

- There exist infinite groups in which the distance function $|\cdot|$ is very difficult to compute. The braid group B_∞ is one of such groups. The computation of the distance function for B_∞ is an NP-hard problem, see (31). Such groups require special treatment.

  Moreover, there exist infinite groups for which the distance function $|\cdot|$ is not computable. We omit consideration of such groups.

To deal with the first problem, we can devise a heuristic procedure. As we show below, if the function M satisfies certain local monotonicity properties, then our procedure achieves good results. The following algorithm is a simple direct descent heuristic which can be used to compute the minimum of a function f.

Algorithm 5.11. (Direct Descend Heuristic)

Input: A graph Γ and a function $f : V(\Gamma) \to \mathbf{R}$.

Output: A vertex v that locally minimizes f on Γ.

Computations:

A. Choose a random $v \in V(\Gamma)$.

B. If v has no adjacent vertex with smaller value of f, then output current v.

C. Otherwise put $v \leftarrow u$ where u is any adjacent vertex such that $f(u) < f(v)$. Go to step B.

Note: As any other direct descend heuristic method, Algorithm 5.11 might not work if the function f has local minima.

In the Lemma 5.12 below, we prove that if a function f satisfies certain local properties, then we achieve good results; namely, the proposed algorithm finds the vertex that minimizes f on Γ exactly. Furthermore, we demonstrate that our weight function $M(\cdot)$ meets the required properties and prove that the Direct Descend algorithm finds a central point for trees, and, hence, for free groups. These tasks are carried out in the rest of this section, in Lemma 5.13 and Theorem 5.14 below.

We say that a function $f : V(\Gamma) \to \mathbf{R}$ is *locally decreasing* if at any vertex $v \in V(\Gamma)$, such that f does not have minimum at v, there exists an adjacent vertex u such that $f(u) < f(v)$. We say that a function f is *locally finite* if for any $a, b \in \mathbf{R}$ the set $f(V(\Gamma)) \cap [a, b]$ is finite.

Lemma 5.12. *Let Γ be a graph and $f : V(\Gamma) \to \mathbf{R}$ a real-valued function that attains its minimum on Γ. If f is locally decreasing and locally finite, then Algorithm 5.11 for Γ and f finds the vertex that minimizes f on Γ.*

Proof. Let $v \in V(\Gamma)$ be a random vertex chosen by Algorithm 5.11 at Step A. If v is a minimum of f, then the algorithm stops with the correct answer v. Otherwise, the algorithm, at Step C, chooses any vertex u adjacent to v such that $f(u) < f(v)$. Such a vertex u exists, since the function f is locally decreasing by assumption. Next, Algorithm 5.11 performs the same steps for u. Essentially, it produces a succession of vertices $v_0, v_1, v_2, \ldots$ such that $v_0 = v$ and, for every $i = 0, 1, 2, \ldots$, the vertices v_i, v_{i+1} are adjacent in Γ with the property $f(v_i) > f(v_{i+1})$.

We claim that the constructed succession cannot be infinite. Assume, to the contrary, that the chain $v_0, v_1, v_2, \ldots$ is infinite. Let m be the minimal value of f on Γ. Then $f(V(\Gamma)) \cap [m, f(v)]$ is infinite, and, therefore, f cannot be locally finite. Contradiction. Hence the sequence is finite, and the last vertex minimizes f on $V(\Gamma)$.

□

Lemma 5.13. *Let μ be a distribution on a locally finite graph Γ such that a weight function $M(\cdot)$ is defined. Then the function M is locally finite on Γ.*

Proof. Since the function $M(\cdot)$ is non-negative, it suffices to prove that for any $b \in \mathbf{R}_+$ the set $M(V(\Gamma)) \cap [0, b]$ is finite. Let $v \in \mathbb{E}(\xi)$, i.e., v minimizes the value of $M(\cdot)$, and choose $r \in \mathbb{N}$ such that

$$0 < \frac{1}{2}M(v) \leq \sum_{i \in B_v(r)} d^2(v,i)\mu(i),$$

as in the proof of Lemma 3.5. Choose an arbitrary value $b \in \mathbf{R}_+$ and put $\alpha = \max\{2, b/M(v)\}$. Let $u \in \Gamma \backslash B_v((\alpha+2)r)$. If $i \in B_v(r)$, we have $d(u,i) \geq (\alpha+2)r - r = (\alpha+1)r$. Then

$$M(u) = \sum_{i \in V(\Gamma)} d^2(u,i)\mu(i) = \sum_{i \in B_v(r)} d^2(u,i)\mu(i) + \sum_{i \notin B_v(r)} d^2(u,i)\mu(i) \geq$$

$$\geq \sum_{i \in B_v(r)} [(\alpha+1)r]^2 \mu(i) \geq (\alpha+1)^2 \sum_{i \in B_v(r)} d^2(v,i)\mu(i) \geq (\alpha+1)^2 \frac{1}{2} M(v) >$$

$$> \frac{(\alpha+1)^2}{\alpha+1} M(v) = (\alpha+1)M(v).$$

In the last inequality we used the fact that $\frac{1}{2}M(v) \geq \frac{1}{\alpha}M(v) > \frac{M(v)}{\alpha+1}$, by the choice of α. Thus,

$$M(u) > (\alpha+1)M(v) > \alpha M(v) > b,$$

by the choice of α again. It means that for vertices u outside of the ball $B_v((\alpha+2)r)$, we have $M(u) > b$. Therefore, $M(V(\Gamma)) \cap [0, b] \subset M(B_v((\alpha+2)r))$, and the set $B_v((\alpha+2)r)$ is finite.

□

Theorem 5.14. *Let μ be a distribution on a locally finite tree T such that a function M is totally defined. Then Algorithm 5.11 for T and M finds a central point (mean-set) of μ on T.*

Proof. Follows from Lemmata 5.12, 5.4, 5.13, and 3.5. More precisely, to prove the theorem, we need to show that the assumptions of Lemma 5.12 are met for the weight function $M(\cdot)$. Indeed, $M(\cdot)$ attains its minimum by Lemma 3.5, and it is locally finite by Lemma 5.13. Finally, it is locally decreasing by the Cut Point Lemma 5.4, because the Cut Point Lemma implies the non-existence of local minimum of $M(\cdot)$ for trees, as discussed in the Remark 5.10. □

Unfortunately, the function M is not always locally decreasing, as shown in Figure 5.5, and a local minimum, computed by Algorithm 5.11, is not always a global minimum. In Figure 5.5, each vertex marked by black has probability 0.5, others have probability 0. The gray vertex v_1 is the center of (Γ, μ) and $M(v_3) = M(v_5) = 5$, $M(v_4) = 4$. Hence the vertex v_4 is a local minimum of this graph, but is not a center (global minimum).

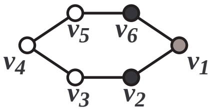

Figure 5.5: Graph with local minimum of M, which is not a center (global minimum)

5.3 Experiments

In this section we demonstrate how the technique of computing mean-sets, employing the Direct Descend Algorithm 5.11 described in section 5.2, works in practice and produces results supporting our SLLN for graphs and groups. More precisely, we arrange series of experiments in which we compute the sample center-sets of randomly generated samples of n random elements and observe a universal phenomenon: the greater the sample size n, the closer the sample mean gets to the actual mean of a given distribution. In particular, we experiment with two classes of groups, free and free abelian, in which the length function is computable. All experiments were done using the CRAG software package, see (9).

5.3.1 Free group results

One of the most frequently used distributions on the free groups is a uniform distribution μ_L on a *sphere* of radius L defined as

$$S_L := \{w \in F(X) \mid |w| = L\}.$$

For example, Figure 2.1 in Section 2.3 illustrates a sphere of radius 2, i.e., S_2 in F_2. Clearly, S_L is finite. Therefore, we can easily define a uniform distribution μ_L on it as follows

$$\mu_L(w) = \begin{cases} \frac{1}{|S_L|} & \text{if } |w| = L; \\ 0 & \text{otherwise.} \end{cases}$$

The reader interested in the question of defining probabilities on groups can find several approaches to this issue in (7). One of the properties of μ_L is that its center-set is just the trivial element of $F(X)$, which is usually denoted by ε. Observe also that the distance of any element of $F(X)$ to the center-set is just the length of this element (or length of the corresponding word, basically).

Tables 5.1, 5.2, and 5.3 below contain the results of experiments for the distributions $\mu_5, \mu_{10}, \mu_{20}, \mu_{50}$ on the groups F_2, F_4, and F_6.

The main parameters in our experiments are

$$\text{the rank } \; r \; \text{ of the free group, the length } \; L, \; \text{ and the sample size } \; n.$$

For every particular triple of parameter values (r, L, n), we perform series of 1000 experiments to which we refer (in what follows), somewhat loosely, as series (r, L, n).

Each cell in the tables below corresponds to a certain series of experiments with parameters (r, L, n). In each experiment from the series (r, L, n), we randomly generate n words $w_1, \ldots, w_n$, according to distribution μ_L, compute the sample mean-set $\mathbb{S}_n = \mathbb{S}_n(w_1, \ldots, w_n)$ for this sample, and compute the displacement of the actual center ε of μ_L from $\mathbb{S}_n$. The set $\mathbb{S}_n$ is computed using Algorithm 5.11 which, according to Theorem 5.14, always produces correct answers for free groups. Every cell in the tables below contains a pair of numbers (d, N); it means that in N experiments out of 1000 the displacement from the real mean was d.

One can clearly see from the tables that the bigger the sample size n is, the closer we get to the actual mean value, i.e., we get an obvious convergence of the empirical (sample) mean to the theoretical (population) mean-set. Another interesting observation that one can extract from the tables is that as the rank of the free group grows, we get better and faster convergence. As we can see, for F_2, sample size of $n = 20$ gives us a pretty good result. For F_4, the convergence is much better though, since starting with a sample size of $n = 18$, we have a perfect agreement of the sample mean-set with the actual one. For F_6, only $n = 14$ is enough. Intuitively, one may think about this outcome as follows: the greater the rank is, the more branching in the corresponding Cayley graph we have, which means that more elements are concentrated in a ball, and the bigger growth (in that sense) causes the better and faster convergence.

At the end, the important conclusion is that these experimental results support the strong law of large numbers for graphs and groups proved in Chapter 3, Section 3.1, and we can say that the law actually works on practice.

5.3.2 Free abelian group results

In this section we describe our experiments with free abelian groups A_n. As we mentioned in Chapter 2, Section 2.4, any free abelian group of rank n is isomorphic to a direct power of the infinite cyclic group $\mathbf{Z}$. Let L be a positive integer. We

L\n	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**	**20**
$\mu_{\mathbf{5}}$	(0,792) (1,183) (2,25)	(0,854) (1,138) (2,8)	(0,891) (1,109)	(0,927) (1,73)	(0,950) (1,50)	(0,962) (1,38)	(0,970) (1,30)	(0,968) (1,32)	(0,992) (1,8)
$\mu_{\mathbf{10}}$	(0,771) (1,197) (2,28) (3,4)	(0,834) (1,158) (2,8)	(0,902) (1,97) (2,1)	(0,932) (1,68)	(0,941) (1,59)	(0,956) (1,44)	(0,966) (1,34)	(0,987) (1,13)	(0,987) (1,13)
$\mu_{\mathbf{20}}$	(0,789) (1,185) (2,21) (3,5)	(0,859) (1,132) (2,9)	(0,871) (1,127) (2,2)	(0,917) (1,82) (2,1)	(0,936) (1,64)	(0,961) (1,39)	(0,963) (1,37)	(0,977) (1,23)	(0,984) (1,16)
$\mu_{\mathbf{50}}$	(0,790) (1,180) (2,27) (3,3)	(0,854) (1,140) (2,6)	(0,906) (1,93) (2,1)	(0,921) (1,79)	(0,951) (1,49)	(0,963) (1,37)	(0,965) (1,35)	(0,981) (1,19)	(0,979) (1,21)

Table 5.1: The results of experiment for F_2.

define a system of probability measures on A_n as follows. Let μ_L to be the uniform distribution on a finite set $[-L, L]^n$. An important property of μ_L is that its center is a singleton set containing the trivial element only.

Tables 5.4, 5.5, and 5.6 below contain the results of experiments for the distributions $\mu_5, \mu_{10}, \mu_{20}$ on the group A_2 and A_4.

The main parameters in our experiments are

the rank r of the free group, the length L, and the sample size n.

For every particular triple of parameter values (r, L, n), we perform series of 1000 experiments to which we refer as series (r, L, n), as in the case of free groups. Again, each cell in the tables below corresponds to a certain series of experiments with parameters (r, L, n). In each experiment from the series (r, L, n), we randomly generate

L\n	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**	**20**
$\mu_{\mathbf{5}}$	(0,943) (1,55) (2,2)	(0,978) (1,22)	(0,988) (1,12)	(0,999) (1,1)	(0,998) (1,2)	(0,1000)	(0,999) (1,1)	(0,1000)	(0,1000)
$\mu_{\mathbf{10}}$	(0,930) (1,69) (2,1)	(0,976) (1,24)	(0,993) (1,7)	(0,994) (1,6)	(0,999) (1,1)	(0,1000)	(0,1000)	(0,1000)	(0,1000)
$\mu_{\mathbf{20}}$	(0,940) (1,58) (2,2)	(0,975) (1,25)	(0,985) (1,15)	(0,991) (1,9)	(0,1000)	(0,999) (1,1)	(0,999) (1,1)	(0,1000)	(0,1000)
$\mu_{\mathbf{50}}$	(0,928) (1,71) (2,1)	(0,984) (1,16)	(0,991) (1,9)	(0,998) (1,2)	(0,997) (1,3)	(0,998) (1,2)	(0,999) (1,1)	(0,1000)	(0,1000)

Table 5.2: The results of experiment for F_4.

n words $w_1, \ldots, w_n$, according to distribution μ_L, compute the sample mean-set $\mathbb{S}_n$ for this sample, and compute the displacement of the actual center ε of μ_L from $\mathbb{S}_n$.

Every cell in the tables below contains a pair of numbers (d, N), meaning that in N experiments out of 1000 the displacement from the real mean was d. The set $\mathbb{S}_n$ is computed using Algorithm 5.11. We observe, from the results of the experiments, that this algorithm does not guarantee us the optimal solution for the case of abelian groups. Nevertheless, we can still observe the convergence, though at a much slower rate. The reason is that in the abelian case the elements are more connected; we have more geodesics and less growth (number of elements in a ball) and, as a result, slower convergence.

$\mathbf{L\backslash n}$	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**
$\mu_{\mathbf{5}}$	(0,932) (1,63) (2,5)	(0,978) (1,22)	(0,991) (1,9)	(0,998) (1,2)	(0,999) (1,1)	(0,1000)	(0,1000)	(0,1000)	(0,1000)
$\mu_{\mathbf{10}}$	(0,903) (1,87) (2,9) (3,1)	(0,971) (1,29)	(0,996) (1,4)	(0,999) (1,1)	(0,999) (1,1)	(0,1000)	(0,1000)	(0,1000)	(0,1000)
$\mu_{\mathbf{20}}$	(0,915) (1,76) (2,8) (3,1)	(0,972) (1,27) (2,1)	(0,991) (1,9)	(0,1000)	(0,1000)	(0,1000)	(0,1000)	(0,1000)	(0,1000)
$\mu_{\mathbf{50}}$	(0,894) (1,95) (2,9) (3,2)	(0,980) (1,20)	(0,990) (1,10)	(0,997) (1,3)	(0,1000)	(0,1000)	(0,1000)	(0,1000)	(0,1000)

Table 5.3: The results of experiment for F_6.

L\n	**10**	**30**	**50**	**100**	**200**
$\mu_{\mathbf{5}}$	(0,104) (1,280) (2,306) (3,212) (4,76) (5,20) (6,1) (7,1)	(0,273) (1,438) (2,266) (3,20) (4,3)	(0,390) (1,470) (2,138) (3,2)	(0,640) (1,308) (2,52)	(0,848) (1,151) (2,1)
$\mu_{\mathbf{10}}$	(0,42) (1,105) (2,155) (3,174) (4,180) (5,138) (6,96) (7,58) (8,26) (9,18) (10,4) (11,3) (12,1)	(0,77) (1,240) (2,325) (3,227) (4,82) (5,39) (6,9) (8,1)	(0,127) (1,317) (2,350) (3,164) (4,36) (5,6)	(0,250) (1,471) (2,244) (3,33) (4,2)	(0,433) (1,445) (2,121) (3,1)
$\mu_{\mathbf{20}}$	(0,7) (1,25) (2,47) (3,74) (4,100) (5,94) (6,106) (7,91) (8,83) (9,100) (10,67) (11,61)	(0,23) (1,95) (2,140) (3,173) (4,166) (5,147) (6,95) (7,64) (8,55) (9,20) (10,12) (11,5)	(0,35) (1,122) (2,195) (3,233) (4,170) (5,126) (6,56) (7,45) (8,12) (9,5) (11,1)	(0,84) (1,231) (2,313) (3,207) (4,109) (5,39) (6,16) (7,1)	(0,144) (1,351) (2,321) (3,145) (4,36) (5,1) (6,2)

Table 5.4: The results of experiment for A_2 - cut to fit.

L\n	**10**	**30**	**50**	**100**	**200**
$\mu_{\mathbf{5}}$	(0,4)	(0,43)	(0,107)	(0,280)	(0,609)
	(1,34)	(1,166)	(1,268)	(1,409)	(1,320)
	(2,107)	(2,309)	(2,373)	(2,230)	(2,68)
	(3,190)	(3,285)	(3,198)	(3,73)	(3,3)
	(4,209)	(4,138)	(4,49)	(4,8)	
	(5,201)	(5,55)	(5,4)		
	(6,128)	(6,2)	(6,1)		
	(7,66)	(7,2)			
	(8,39)				
	(9,15)				
	(10,4)				
	(11,2)				
	(12,1)				
$\mu_{\mathbf{10}}$	(0,1)	(0,2)	(0,10)	(0,31)	(0,112)
	(1,3)	(1,25)	(1,48)	(1,154)	(1,317)
	(2,10)	(2,71)	(2,155)	(2,299)	(2,361)
	(3,26)	(3,147)	(3,212)	(3,269)	(3,171)
	(4,38)	(4,188)	(4,261)	(4,179)	(4,37)
	(5,92)	(5,209)	(5,163)	(5,53)	(5,2)
	(6,106)	(6,165)	(6,98)	(6,14)	
	(7,106)	(7,107)	(7,39)	(7,1)	
	(8,124)	(8,43)	(8,11)		
	(9,121)	(9,23)	(9,3)		
	(10,97)	(10,15)			
	(11,83)	(11,3)			
	(12,69)	(12,1)			
	(13,31)	(13,1)			
	(14,38)				
	(15,20)				
	(16,15)				
	(17,9)				
	(18,10)				

Table 5.5: The results of experiment for A_4.

L\n	**10**	**30**	**50**	**100**	**200**
μ_{20}	(2,1)	(1,3)	(0,1)	(0,1)	(0,14)
	(3,1)	(2,6)	(1,6)	(1,20)	(1,70)
	(4,4)	(3,14)	(2,20)	(2,59)	(2,154)
	(5,5)	(4,39)	(3,43)	(3,126)	(3,223)
	(6,17)	(5,55)	(4,77)	(4,168)	(4,245)
	(7,19)	(6,67)	(5,110)	(5,173)	(5,151)
	(8,26)	(7,115)	(6,131)	(6,160)	(6,83)
	(9,28)	(8,107)	(7,148)	(7,124)	(7,47)
	(10,45)	(9,100)	(8,137)	(8,73)	(8,9)
	(11,22)	(10,113)	(9,97)	(9,59)	(9,4)
	(12,54)	(11,77)	(10,63)	(10,18)	
	(13,58)	(12,89)	(11,67)	(11,11)	
	(14,84)	(13,64)	(12,50)	(12,5)	
	(15,65)	(14,51)	(13,20)	(13,1)	
	(16,62)	(15,26)	(14,16)	(14,1)	
	(17,67)	(16,27)	(15,6)	(15,1)	
	(18,52)	(17,23)	(16,5)		
	(19,61)	(18,12)	(18,1)		
	(20,55)	(19,3)	(19,2)		
	(21,36)	(20,4)			
	(22,37)	(22,3)			
	(23,38)	(24,2)			
	(24,33)				
	(25,30)				
	(26,24)				
	(27,15)				
	(28,23)				
	(29,13)				
	(30,5)				
	(31,7)				

Table 5.6: The results of experiment for A_4 - continue. Cut to fit.

Chapter 6

Refinements. Central order on $V(\Gamma)$.

6.1 Medians as mean-sets of class one

As indicated in Chapter 3, Section 3.1.2, it is possible to consider center-sets of class c for graph-valued random elements by defining

$$\mathbb{E}^{(c)}(\xi) := \{v \in V(\Gamma) \mid M^{(c)}(v) \leq M^{(c)}(u), \ \ \forall u \in V(\Gamma)\}.$$

Even though the weight function $M^{(c)}$ and the mean-set $\mathbb{E}^{(c)}$ of class one do not suit our purposes in working with centers of groups and graphs (see Section 3.1.2 for more details), it turns out that $M^{(1)}$ and $\mathbb{E}^{(1)}$ find their own interpretation related to another useful measure of central tendency in classical probability theory and statistics, namely, the median, as we shall see below.

The goal of this section is to define a notion of median set for graphs and to bring forth its connection with $M^{(1)}$ and $\mathbb{E}^{(1)}$. Recall that, according to classical theory, a *median* of a probability distribution μ on the real line $\mathbf{R}$ is a point $\mathfrak{M}$ satisfying

$$\mu((-\infty, \mathfrak{M}]) \geq 1/2 \ \text{ and } \ \mu([\mathfrak{M}, \infty)) \geq 1/2,$$

i.e., $\mathfrak{M}$ can be viewed as a midpoint of μ. Note that according to this definition, μ may not have a unique median. We denote the set of medians by $\mathbb{M}$. Observe also that the set $\mathbb{M}$ is connected, i.e., it is always an interval.

Proposition 6.1 (Connection of $\mathfrak{M}$ and $\mathbb{E}^{(1)}(\cdot)$)). *Let $\xi : \Omega \to \mathbf{Z}$ be an integer-valued random variable with the classical median set $\mathbb{M}$. Assume that $M^{(1)}(\cdot)$ is defined on $\mathbf{Z}$. Then we have $\mathbb{E}^{(1)}(\xi) = \mathbb{M} \cap \mathbf{Z}$.*

Proof. We can naturally continue the function $M^{(1)}(\cdot)$ from $\mathbf{Z}$ to $\mathbf{R}$ by putting $M^{(1)}(v) = \sum_{n \in \mathbf{Z}} |v - n| \mu(n)$ for every $v \in \mathbf{R}$. The resulting function $M^{(1)} : \mathbf{R} \to [0, \infty)$ is piecewise linear with corners at $\{(n, M^{(1)}(n)) \mid n \in \mathbf{Z}\}$. Furthermore, for every $v \in \mathbf{Z}$,

$$M^{(1)}(v+1) - M^{(1)}(v) = \sum_{n \le v} \mu(n) - \sum_{n \ge v+1} \mu(n).$$

Therefore, $M^{(1)}(v+1) - M^{(1)}(v)$ changes sign at the same point where $\sum_{n \le v} \mu(n) - \sum_{n \ge v+1} \mu(n)$ changes sign and

$$v \in \mathbb{E}^{(1)}(\xi) \Leftrightarrow \mu((-\infty, v-1]) - \mu([v, \infty)) \le 0 \text{ and } \mu((-\infty, v]) - \mu([v+1, \infty)) \ge 0.$$

The later is clearly equivalent to conditions that $\mu((-\infty, v]) \ge 1/2$ and $\mu([v, \infty)) \ge 1/2$. Therefore, it follows from definitions of $\mathbb{E}^{(1)}(\xi)$ and median-set $\mathbb{M}$ that $\mathbb{E}^{(1)}(\xi) = \mathbb{M} \cap \mathbf{Z}$, which is not empty.

□

Observe that classical median $\mathfrak{M}$ is defined for any distribution μ on $\mathbf{R}$, but the function $M^{(1)}(\cdot)$ is not always defined. We need to fix this inconsistency in order to continue conveniently work with medians further. The following lemma helps us to tackle this problem.

Lemma 6.2. *For any $u, v \in V(\Gamma)$ the limit*

$$\lim_{n \to \infty} \sum_{i=1}^{n} (d(u, v_i) - d(v, v_i)) \mu(v_i)$$

exists. Moreover, such limit does not depend on the order of vertices v_i.

Proof. Let $d = d(u, v)$. Then by the triangle inequality, for any $v_i \in \Gamma$, both $d(u, v_i) - d(v, v_i) \leq d$ and $d(v, v_i) - d(u, v_i) \leq d$ hold, and, therefore, $|d(u, v_i) - d(v, v_i)| \leq d$. Thus, $\sum_{i=1}^{\infty} |d(u, v_i) - d(v, v_i)|\mu(v_i) \leq d < \infty$, and the sum $\sum_{i=1}^{\infty}(d(u, v_i) - d(v, v_i))\mu(v_i)$ converges absolutely. Hence the result. □

Definition 6.3. Define a function $\rho^{(1)} : \Gamma \times \Gamma \to \mathbb{N}$ to be equal to $\sum_{s\in\Gamma}(d(u, s) - d(v, s))\mu(s)$, for $u, v \in \Gamma$.

Lemma 6.4. *Let $\xi : \Omega \to \mathbf{Z}$ be an integer-valued random variable such that $M^{(1)}(\cdot)$ is defined on $\mathbf{Z}$. Then for every $u, v \in \mathbf{Z}$, $\rho^{(1)}(u, v) = M^{(1)}(u) - M^{(1)}(v)$.*

Proof. Obvious. □

It follows from Lemma 6.2 that for any $u, v \in \Gamma$ the value

$$\rho^{(1)}(u, v) = \sum_{s\in\Gamma} \Big(d(u, s) - d(v, s)\Big)\mu(s)$$

is correctly defined. Moreover, we can rewrite it in the following, more insightful in some sense, way

$$\rho^{(1)}(u, v) = \sum_{\delta=-d(u,v)}^{d(u,v)} \delta \cdot \mu\Big(s \in V(\Gamma) \mid d(u, s) - d(v, s) = \delta\Big).$$

The function $\rho^{(1)}$ allows us to introduce a notion of order on the vertices of the graph Γ that will eliminate the problem of dependence of medians on finiteness of $M^{(1)}$.

Definition 6.5. We define a binary relation $<^{(1)}$ on the vertices of Γ by

$$u <^{(1)} v \quad \Leftrightarrow \quad \rho^{(1)}(u, v) < 0.$$

It is easy to check that the above relation defines a certain partial order on Γ; this is not a partial order in classical sense since it is not anti-symmetric (for more on binary relations see (14)).

Proposition 6.6. Let μ be a distribution on a locally finite graph Γ. The binary relation $<^{(1)}$ is

- (anti-reflexive), i.e., for no $v \in V(\Gamma)$, $v <^{(1)} v$;
- (neither anti-symmetric nor symmetric), i.e., for no $u, v \in V(\Gamma)$, $v <^{(1)} u$ and $u <^{(1)} v$;
- (transitive), i.e., for every $u, v, w \in V(\Gamma)$, if $u <^{(1)} v$ and $v <^{(1)} w$ then $u <^{(1)} w$.

Proof. Indeed, for every $v \in \Gamma$, $v \not<^{(1)} v$ since $\rho^{(1)}(v, v) = 0$. For every u, v, inequalities $\rho^{(1)}(u, v) < 0$ and $\rho^{(1)}(u, v) > 0$ cannot hold simultaneously and hence at most one of $v <^{(1)} u$ and $u <^{(1)} v$ are true.

Finally, assume that for some $u, v, w \in \Gamma$, $u <^{(1)} v$ and $v <^{(1)} w$. This means that $\rho^{(1)}(u, v) < 0$ and $\rho^{(1)}(v, w) < 0$. Notice that for every $s \in \Gamma$, $d(u, v_i) - d(w, v_i) = [d(u, v_i) - d(v, v_i)] + [d(v, v_i) - d(w, v_i)]$ and hence $\rho^{(1)}(u, w) = \rho^{(1)}(u, v) + \rho^{(1)}(v, w) < 0$, therefore $u <^{(1)} w$.

□

Now, we can see the mean-set $\mathbb{E}^{(1)}(\mu)$ of class one in a totally new light employing the binary relation above and define a median-set in graph Γ.

Definition 6.7. Let $\mathbb{E}^{(1)}(\mu) = \{v \in \Gamma \mid u \not<^{(1)} v, \forall u \in \Gamma\}$. The set $\mathbb{E}^{(1)}(\mu)$ is called the *median-set* in Γ relative to μ.

It turns out that the set $\mathbb{E}^{(1)}(\mu)$ is always defined, i.e., it is finite and non-empty (see Proposition 6.8 below). It is important to note that the median-set $\mathbb{E}^{(1)}\mu$ is defined even if the function $M^{(1)}(\cdot)$ takes infinite values, in contrast with our mean-sets on graphs which depend totally on the finiteness of $M^{(2)}(\cdot)$.

Proposition 6.8. *Let μ be a probability measure on a locally finite graph Γ. Then the set $\mathbb{E}^{(1)}(\mu)$ is finite and non-empty.*

Proof. Fix any vertex $v \in \Gamma$ and choose r such that $\mu(B_v(r)) \geq 0.99$. Consider a vertex $u \in \Gamma \setminus B_v(3r)$ and put

$$\alpha := d(v, u)/r.$$

Then

$$\rho^{(1)}(v,u) = \sum_{s\in\Gamma} \Big(d(v,s) - d(u,s)\Big)\mu(s)$$

$$= \sum_{s\in B_v(r)} \Big(d(v,s) - d(u,s)\Big)\mu(s) + \sum_{s\in\Gamma\setminus B_v(r)} \Big(d(v,s) - d(u,s)\Big)\mu(s) \leq$$

which, using the observation that, for $s \in B_v(r)$, $d(v,s) - d(u,s) \leq r - (\alpha r - r) = -r(\alpha - 2)$ for the first sum, and employing the triangle inequality for the second sum, can be bounded by

$$\leq -(\alpha-2)r\mu(B_v(r)) + \alpha r\mu(\Gamma \setminus B_v(r)) \leq -(\alpha-2)r\cdot 0.99 + \alpha r \cdot 0.01$$

$$= (1.98 - 0.98\alpha)r,$$

and the last bound is negative since $\alpha \geq 3$. Hence, for every $u \in \Gamma \setminus B_v(3r)$, we have $v <^{(1)} u$. Thus, $\mathbb{E}^{(1)}(\mu) \subseteq B_v(3r)$. Since the set $B_v(3r)$ is finite, it contains finitely many least elements.

□

Remark 6.9 (An Optimality Property)**.** Recall that in classical probability theory, the median possesses an optimality property asserting that the median is a central point minimizing the average of the absolute deviations; namely, we call median of the distribution of the real-valued random variable X a value c that minimizes the classical expectation $\mathbb{E}(|X - c|)$. Clearly, this is in agreement with our definition, since our generalized median set minimizes $M^{(1)}(v) = \sum_{u\in\Gamma} d(v,u)\mu(u)$. In the sense of L^p-spaces, our median sets correspond to L^1 settings, as well as the classical ones.

6.2 Expectation when $M^{(1)}$ is defined

Recall that our Definition 3.1 of the mean-set $\mathbb{E}$ for locally finite graphs, given in Chapter 3, Section 3.1, is closely related to the classical definition of expectation on the real line $\mathbf{R}$ (see Proposition 3.9 for details). It turns out, though, that Definition 3.1 of a mean-set generalized to graphs and groups is weaker when applied to $\mathbf{R}$, in the sense that the classical expectation $\mathbb{E}$ may be finite, while the generalized one may not be defined, see Lemma 6.10 below. The goals of this section is to fix this flaw and consider a possibility of even more general definition for mean-sets on graphs and groups.

6.2.1 Central order on Γ and general $\mathbb{E}(\xi)$

We know that the classical mean of a distribution on $\mathbf{R}$ coincides with its first moment. On the other hand, if we look at $\mathbf{R}$ as a graph and apply Definition 3.1 of a mean-set for graphs, we see that the existence of the second moment is necessary for the mean-set to be defined, as the following lemma shows.

Lemma 6.10. *Let Γ be a graph with vertices corresponding to integer points on the real line $\mathbf{R}$ and μ a probability measure on Γ. Then $\mathbb{E}(\mu) \equiv \mathbb{E}^{(2)}(\mu)$ is defined if and only if the second classical moment for μ is finite.*

Proof. Recall that $\mathbb{E} \equiv \mathbb{E}^{(2)}$ is defined if and only if the function $M \equiv M^{(2)}$ is finite for every vertex in the graph Γ, and $M(\cdot)$, as defined in (3.2), is exactly the second classical moment of μ.

□

This means, in particular, that when applied to $\mathbf{R}$, the classical mean $\mathbb{E}$ may be finite, while the generalized one may not be defined. We eradicate this problem by introducing a binary relation that gives us a way to compare elements (vertices) of Γ. This method may be employed in a more general definition of $\mathbb{E}^{(c)}$ on graphs, independently of whether a weight function $M^{(c)}$ is finite or not.

Let μ be a probability measure on a locally finite connected graph Γ, and u, v two elements of $V(\Gamma)$. The values $M^{(c)}(u)$ and $M^{(c)}(v)$ can be infinite, but it is possible, sometimes, to compare these values in the following sense.

Definition 6.11. Let μ be a probability measure on a locally finite connected graph Γ, and $u, v \in V(\Gamma)$. Let

$$\rho^{(c)}(v,u) = \sum_{s \in V(\Gamma)} \Big(d^c(v,s) - d^c(u,s) \Big) \mu(s), \tag{6.1}$$

which can be a finite or an infinite value, or it can be undefined. We say that v is *more c-central* than u if the above sum equals to a finite negative value, i.e., $\rho^{(c)}(v,u) < 0$, or properly diverges to negative infinity, i.e., $\rho^{(c)}(v,u) = -\infty$.
In this case, we write $v <^{(c)} u$ and call this binary relation the *central order*.

Proposition 6.12. The relation $<^{(c)}$ is:

- anti-reflexive, i.e., there is no $v \in V(\Gamma)$ such that $v <^{(c)} v$;
- transitive, i.e., for any $t, u, v \in V(\Gamma)$,

$$t <^{(c)} u, \quad u <^{(c)} v \Rightarrow t <^{(c)} v;$$

- neither anti-symmetric nor symmetric, i.e., there is no pair $u, v \in V(\Gamma)$ such that $u <^{(c)} v$ and $v <^{(c)} u$;

Moreover, $\rho^{(c)}(\cdot,\cdot)$ has the following property

$$\rho^{(c)}(u,v) = \rho^{(c)}(u,w) + \rho^{(c)}(w,v), \ \forall\, u, w, v \in V(\Gamma).$$

Proof. Straightforward verification. □

For more insights about binary relations and their properties see (14). We would like to emphasize that our *central order* represents a partial order relation on $V(\Gamma)$ (not classical one, because it is not anti-symmetric). The advantage of this new development is that it allows us to weaken the assumptions in the major theorems in the sequel; for instance, instead of having $M \equiv M^{(2)} < \infty$, we can assume just $M^{(1)} < \infty$ in the Strong Law of Large Numbers for graphs (groups). Next proposition is the first step in this direction.

Proposition 6.13. *Let μ be a probability measure on a locally finite connected graph Γ. Then for every $c \in \mathbb{N}$ the following are equivalent:*

- *$M^{(c-1)}(\cdot)$ is defined;*
- *$\rho^c : V(\Gamma) \times V(\Gamma) \to \mathbf{R}$ is well-defined, i.e., $|\rho^c(u,v)| < \infty$ for any $u, v \in \Gamma$.*

Proof. Assume that $M^{(c-1)}$ is defined on $V(\Gamma)$. Let $v_1, v_2 \in V(\Gamma)$ and $d = d(v_1, v_2)$. For every $i = -d, \ldots, d$, define a set

$$S_i = \{s \in V(\Gamma) \mid d(v_1, s) - d(v_2, s) = i\}$$

and note that $V(\Gamma) = S_{-d} \sqcup \ldots \sqcup S_d$, i.e., we have a partition of $V(\Gamma)$. Using this partition and an elementary formula

$$a^c - b^c = (a-b)\sum_{j=0}^{c-1} a^{c-j-1}b^j, \tag{6.2}$$

we can write

$$\begin{aligned}\rho^{(c)}(v_1, v_2) &= \sum_{s \in V(\Gamma)} \Big(d^c(v_1, s) - d^c(v_2, s) \Big) \mu(s) \\ &= \sum_{i=-d}^{d} \left(i \sum_{s \in S_i} \left(\sum_{j=0}^{c-1} d^j(v_1, s) d^{c-j-1}(v_2, s) \mu(s) \right) \right)\end{aligned}$$

$$= \sum_{i=-d}^{-1} \left(i \sum_{s \in S_i} \left(\sum_{j=0}^{c-1} d^j(v_1, s) d^{c-j-1}(v_2, s) \mu(s) \right) \right) + \sum_{i=1}^{d} \left(i \sum_{s \in S_i} \left(\sum_{j=0}^{c-1} d^j(v_1, s) d^{c-j-1}(v_2, s) \mu(s) \right) \right),$$

and observe that, depending on the sign of i,

$$i \sum_{s \in S_i} \left(\sum_{j=0}^{c-1} d^j(v_1, s) d^{c-j-1}(v_2, s) \mu(s) \right) \le ic \sum_{s \in S_i} d^{c-1}(v_2, s) \mu(s), \quad \text{if} \quad i < 0$$

$$i \sum_{s \in S_i} \left(\sum_{j=0}^{c-1} d^j(v_1, s) d^{c-j-1}(v_2, s) \mu(s) \right) \le ic \sum_{s \in S_i} d^{c-1}(v_1, s) \mu(s), \quad \text{if} \quad i > 0.$$

When doing the above estimate, we just observed that, when $i < 0$, we have $d(v_1, s) < d(v_2, s)$, and, when $i > 0$, we have $d(v_1, s) > d(v_2, s)$, which is clear from the definition of the sets S_i, $i = -d, \ldots, d$. Now, since $M^{(c-1)}$ is defined for v_1 and v_2, it follows that the sum $\sum_{s\in S_i} d^{c-1}(v_2,s)\mu(s)$, as well as the sum $\sum_{s\in S_i} d^{c-1}(v_1,s)\mu(s)$, above is finite. Thus, the original infinite sum was broken into the finite sum of convergent sums. Therefore, the whole series converges to some real value.

Conversely, assume that ρ^c is a well defined function from $V(\Gamma)\times V(\Gamma)$ to $\mathbf{R}$. Let v be an arbitrary, but fixed, vertex in Γ and $\{v_1,\ldots,v_n\}$ the set of vertices adjacent to v in Γ. For every vertex v_i, $i=1,\ldots,n$, define a set $T_i=\{s\in V(\Gamma)\mid d(v_i,s)=d(v,s)-1\}$. Since for every i, the value $\rho^{(c)}(v,v_i)$ is finite, it follows that

$$\infty > \sum_{s\in T_i}(d^c(v,s)-d^c(v_i,s))\mu(s) = \sum_{s\in T_i}\left(\sum_{j=0}^{c-1} d^j(v_i,s)d^{c-j-1}(v,s)\mu(s)\right)$$
$$\geq \sum_{s\in T_i} d^{c-1}(v,s)\mu(s),$$

where we leave only the term corresponding to $j=0$ and ignore the rest in the estimate above. Thus, $\sum_{s\in T_i} d^{c-1}(v,s)\mu(s)$ converges. Now, we notice that $V(\Gamma)\setminus\{v\}=T_1\cup\ldots\cup T_n$ and

$$M^{(c-1)}(v)=\sum_{s\in V(\Gamma)} d^{c-1}(v,s)\mu(s)=\sum_{s\in V(\Gamma)\setminus\{v\}} d^{c-1}(v,s)\mu(s)$$
$$\leq \sum_{i=1}^{n}\left(\sum_{s\in T_i} d^{c-1}(v,s)\mu(s)\right)<\infty,$$

since the last sum is a finite sum of absolutely converging series. Hence $M^{(c-1)}$ is defined at v, and, therefore, on the whole $V(\Gamma)$ (this can be seen easily, similar to the Lemma 3.3, which was proved for the case $c=2$).

□

Definition 6.14. Let $(\Omega,\mathcal{F},\mathbf{P})$ be a probability space and $\xi:\Omega\to V(\Gamma)$ a random element. We define $\mathbb{E}^{(c)}(\xi)$ to be the set of vertices in Γ which are minimal relative to $<^{(c)}$, i.e.,

$$\mathbb{E}^{(c)}(\xi)=\{v\in\Gamma\mid \nexists u \text{ s.t. } u<^{(c)}v\}=\{v\in\Gamma\mid \nexists u \text{ s.t. } \rho^{(c)}(u,v)<0\}.$$

We call this set the *general mean-set of ξ of class c relative to the central order* (or simply *mean-set relative to $<^{(c)}$*, or just *ρ-mean-set*, for short).

Note: If $c = 2$, then we do not specify the class of the mean-set: our expectation for a graph- (group-)valued random element is denoted by $\mathbb{E} \equiv \mathbb{E}^{(2)}$ then.

Observe that this definition makes sense even if the weight function $M^{(c)}$ is infinite. If $M^{(c)}$ is finite, then this definition agrees with what we had before, namely, the Definition 3.8 (as we shall see below in Corollary 6.18). If $M^{(c)}$ is infinite, we have to show that our general mean-set relevant to the central order is finite and non-empty under some conditions. This is the subject of the next theorem.

Theorem 6.15. *Let μ be a distribution on a locally finite graph Γ. If $M^{(c-1)}$ is defined on Γ then $1 \le |\mathbb{E}^{(c)}(\mu)| < \infty$.*

Proof. Assume that $M^{(c-1)}(\cdot)$ is defined on Γ. By Proposition 6.13, this implies that the function $\rho^{(c)}(\cdot,\cdot)$ is well defined on $V(\Gamma) \times V(\Gamma)$. Fix an arbitrary $v_0 \in V(\Gamma)$ such that $\mu(v_0) > 0$. Define a region $S^+_{v_0,v}$ to be a set of vertices that are closer to v than to v_0, i.e.,

$$S^+_{v_0,v} = \{s \in V(\Gamma) \mid d(v_0,s) > d(v,s)\}.$$

Then, for every $v \in V(\Gamma)$, if $d = d(v_0, v)$, we have

$$\rho^{(c)}(v_0,v) = \sum_{s \in V(\Gamma)} (d^c(v_0,s) - d^c(v,s))\mu(s)$$

$$\le -d^c\mu(v_0) + \sum_{s \in S^+_{v_0,v}} (d^c(v_0,s) - d^c(v,s))\mu(s)$$

where we just disregarded the negative part of the sum, corresponding to the complement of $S^+_{v_0,v}$, and, continuing, with the help of formula (6.2) and triangle inequality,

$$\le -d^c\mu(v_0) + d \sum_{s \in S^+_{v_0,v}} \left(\sum_{j=0}^{c-1} d^j(v_0,s) d^{c-j-1}(v,s) \right) \mu(s)$$

$$\le -d^c\mu(v_0) + dc \sum_{s \in S^+_{v_0,u}} d^{c-1}(v_0,s)\mu(s),$$

employing the definition of the set $S^+_{v_0,v}$ in the last inequality. Observe that since $M^{(c-1)}(v_0) < \infty$ and $S^+_{v_0,v} \subseteq V(\Gamma) \setminus B_{v_0}(d/2)$, it follows that

$$\sum_{s \in S^+_{v_0,v}} d^{c-1}(v_0,s)\mu(s) \le \sum_{s \in V(\Gamma)\setminus B_{v_0}(d/2)} d^{c-1}(v_0,s)\mu(s) \to 0 \quad \text{as} \quad d \to \infty.$$

In particular, we can choose $d \in \mathbb{N}$ such that

$$\sum_{s \in V(\Gamma)\setminus B_{v_0}(d/2)} d^{c-1}(v_0,s)\mu(s) < d^{c-1}c^{-1}\mu(v_0)$$

or

$$dc \sum_{s \in V(\Gamma)\setminus B_{v_0}(d/2)} d^{c-1}(v_0,s)\mu(s) < d^{c}\mu(v_0)$$

Hence, for any v such that $v \in V(\Gamma)\setminus B_{v_0}(d)$, with d chosen above, we have $\rho^{(c)}(v_0,v) < 0$ and, therefore, $v_0 <^{(c)} v$. This means that $\mathbb{E}^{(c)}(\mu) \subseteq B_{v_0}(d)$ and $|\mathbb{E}^{(c)}\mu| < \infty$. Moreover, any order has a minimal element in a finite set and $1 \le |\mathbb{E}^{(c)}\mu|$.

□

Now we can state the following theorem without assumption of finiteness of $M(\cdot)$.

Proposition 6.16. (Equivalence of the general mean-set to the classical mean) *Let μ be a probability distribution on $\mathbf{Z}$ such that the classical expectation*

$$\mathfrak{m} = \sum_{i \in \mathbf{Z}} i\mu(i)$$

is finite. Then the mean-set $\mathbb{E}(\mu) \equiv \mathbb{E}^{(2)}(\mu)$, relative to the central order, is finite, and for any $v \in \mathbb{E}(\mu)$ we have $|\mathfrak{m} - v| \le 1/2$.

Proof. Note that the function $\rho^{(2)}$ can be naturally extended from $\mathbf{Z} \times \mathbf{Z}$ to $\mathbf{R} \times \mathbf{R}$. For any $v \in \mathbf{R}$ we have

$$\rho^{(2)}(v,\mathfrak{m}) = \sum_{s \in \mathbf{Z}} (d^2(v,s) - d^2(\mathfrak{m},s))\mu(s) = \sum_{s \in \mathbf{Z}} ((v-s)^2 - (\mathfrak{m}-s)^2)\mu(s)$$

$$= v^2 - \mathfrak{m}^2 + 2(\mathfrak{m} - v)\sum_{s \in \mathbf{Z}} s\mu(s) = v^2 - \mathfrak{m}^2 + 2\mathfrak{m}^2 - 2v\mathfrak{m} = (v - \mathfrak{m})^2.$$

Hence $\mathbb{E}(\mu)$ contains integer values v that minimize the function $(v - \mathfrak{m})^2$. There are 2 cases possible:

- If $\mathfrak{m} = k + 1/2$, where $k \in \mathbf{Z}$, then $\mathbb{E}(\mu) = \{k, k+1\}$ and $|\mathfrak{m} - k| = 1/2$.
- If $\mathfrak{m}$ is not of the form $k+1/2$ for any $k \in \mathbf{Z}$, then $|\mathfrak{m} - k| < 1/2$ for some $k \in \mathbf{Z}$ and $\mathbb{E}(\mu) = \{k\}$.

In both cases we see that the statement of the proposition holds.

□

Having proved the proposition above, we see that classical mean on the real line and the *generale mean-set relative to central order* on Γ (applied to $\mathbf{R}$), are either both finite or both infinite.

Proposition 6.17. If $M^{(c)}(\cdot)$ is totally defined on (Γ, μ), then for any $v, u \in \Gamma$,

$$\rho^{(c)}(u,v) = M^{(c)}(u) - M^{(c)}(v)$$

and

$$u <^{(c)} v \quad \Leftrightarrow \quad M^{(c)}(u) < M^{(c)}(v).$$

Proof. If $M^{(c)}(\cdot)$ is defined on Γ for a given μ , then the series

$$\sum_{s \in V(\Gamma)} d^c(v,s)\mu(v_i) \quad \text{and} \quad \sum_{s \in V(\Gamma)} d^c(u,s)\mu(v_i)$$

converge to $M^{(c)}(v)$ and $M^{(c)}(u)$ respectively. Therefore

$$\rho^{(c)}(u,v) = \sum_{s \in V(\Gamma)} (d^c(u,s) - d^c(v,s))\mu(v_i) = M^{(c)}(u) - M^{(c)}(v)$$

and $u <^{(c)} v$ if and only if $\rho^{(c)}(u,v) < 0$ if and only if $M^{(c)}(u) < M^{(c)}(v)$.

□

Corollary 6.18. *If the weight function $M^{(c)}$ of class c is totally defined, then the definitions 3.8 and 6.14 coincide.*

6.2.2 SLLN and other results for central order

Now we prove some of the main theorems for ρ-centers $\mathbb{E}(\xi)$ under the weaker assumption of $M^{(1)}$ being finite instead of $M \equiv M^{(2)} < \infty$. Our first result corresponds to Lemma 3.15.

Lemma 6.19. (Inclusion Lemma for the central order) *Let Γ be a locally-finite connected graph, $\{\xi_i\}_{i=1}^{\infty}$, $\xi_i : \Omega \to V(\Gamma)$, a sequence of i.i.d. Γ-valued random elements defined on a given probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and μ the probability measure on Γ induced by ξ_1. Suppose that the weight function of class one $M^{(1)}$ is totally defined. Then*

$$\mathbf{P}\left(\limsup_{n\to\infty} \mathbb{S}(\xi_1, \ldots, \xi_n) \subseteq \mathbb{E}(\xi_1)\right) = 1.$$

Proof. The proof is basically the same as the proof of Lemma 3.15. We just note that, corresponding to the sampling distribution μ_n of (3.1) and the sample weight function $M_n(\cdot)$, we can define a *sample central order function* $\rho_n^{(2)}(\cdot,\cdot)$ with *sample central order* at $v, u \in V(\Gamma)$ as

$$\rho_n^{(2)}(v,u) = \frac{1}{n}\sum_{i=1}^{n}\big(d^2(v,\xi_i) - d^2(u,\xi_i)\big),$$

or, equivalently,

$$\rho_n^{(2)}(v,u) = \sum_{s\in V(\Gamma)}\big(d^2(v,s) - d^2(u,s)\big)\mu_n(s) = M_n(v) - M_n(u).$$

In addition, using the central order notation, we can rewrite the definition of the *sample mean-set*; namely,

$$\mathbb{S}_n = \mathbb{S}(\xi_1, \ldots, \xi_n) = \Big\{v \in \Gamma \mid u \not<^{(c,\mu_n)} v, \forall\ u \in \Gamma\Big\},$$

where $<^{(c,\mu_n)}$ indicates that we deal with *sample central order* in the binary relation of central order.

Now we can supply the outline of the proof:

- The function $\rho^{(2)}(u,v)$ is the expectation of the function $d^2(u,\xi) - d^2(v,\xi)$.

- Hence, for every $u, v \in V(\Gamma)$, $\mathbf{P}\Big(\rho_n^{(2)}(u,v) \to \rho^{(2)}(u,v)\Big) = 1$.

- Hence, $\mathbf{P}\Big(\rho_n^{(2)}(u,v) \to \rho^{(2)}(u,v), \quad \forall u, v \in V(\Gamma)\Big) = 1$.

- Hence, if $v \notin \mathbb{E}(\xi_1)$, i.e., $u <^{(2)} v$ ($\rho^{(2)}(u,v) < 0$) for some u, then, with probability 1, $u <^{(2,\mu_n)} v$ in samples, i.e., $\rho_n^{(2)}(u,v) < 0$ eventually, and, therefore, $v \notin \limsup_{n\to\infty} \mathbb{S}_n$.

□

Next is the analogue of Theorem 3.16 (SLLN), but with the assumption that $M^{(1)} < \infty$, as opposed to the original requirement of the finiteness of $M \equiv M^{(2)}$.

Theorem 6.20 (SLLN for graph-valued random elements with a singleton mean-set)**.** *Let Γ be a locally-finite connected graph and $\{\xi_i\}_{i=1}^\infty$ a sequence of i.i.d. Γ-valued random elements $\xi_i : \Omega \to V(\Gamma)$ such that $M^{(1)}$ is totally defined. If the general mean-set relative to central order $\mathbb{E}(\xi_1) = \{v\}$ for some $v \in V(\Gamma)$, i.e., if $\mathbb{E}(\xi_1)$ is a singleton, then the following holds almost surely:*

$$\mathbb{S}(\xi_1, \ldots, \xi_n) \longrightarrow \mathbb{E}(\xi_1) \text{ as } n \to \infty.$$

Proof. We prove that for some vertex $v_0 \in V(\Gamma)$ there exists a sufficiently large number $m > 0$ such that $v \in B_{v_0}(m)$ and the following inequalities hold:

$$\mathbf{P}(\exists N \text{ s.t. } \forall n > N \ \forall u \in B_{v_0}(m) \setminus \{v\}, \quad \rho_n^{(2)}(v,u) < 0) = 1. \tag{6.3}$$

$$\mathbf{P}(\exists N \text{ s.t. } \forall n > N \ \forall u \in V(\Gamma) \setminus B_{v_0}(m), \quad \rho_n^{(2)}(v,u) < 0) = 1, \tag{6.4}$$

The first equality obviously holds for any $v_0 \in V(\Gamma)$ and m because $B_{v_0}(m)$ is a finite set of points and using the strong law of large numbers applied to the sequence of i.i.d. random variables $d^2(u,\xi_i) - d^2(v,\xi_i)$, $i = 1, 2, \ldots$, with $\rho^{(2)}(u,v) = \mathbb{E}\Big(d^2(u,\xi_1) - d^2(v,\xi_1)\Big)$, finitely many times we get the result.

To prove the second inequality, we fix any $v_0 \in V(\Gamma)$ such that $\mu(v_0) \neq 0$ and define

$$S_{v_0,u}^+ = \{s \in V(\Gamma) \mid d(v_0,s) > d(u,s)\}.$$

With this in mind, for every $u \in V(\Gamma)$, we have

$$\rho^{(2)}(v_0,u) = \sum_{s\in V(\Gamma)} \big(d^2(v_0,s) - d^2(u,s)\big)\mu(s)$$

$$= \sum_{s\in V(\Gamma)} \Big(d(v_0,s) - d(u,s)\Big)\Big(d(v_0,s) + d(u,s)\Big)\mu(s)$$

$$\leq -d^2(v_0,u)\mu(v_0) + \sum_{s\in S^+_{v_0,u}} d(v_0,u)\Big(d(v_0,s) + d(u,s)\Big)\mu(s),$$

$$\leq -d^2(v_0,u)\mu(v_0) + 2d(v_0,u) \sum_{s\in S^+_{v_0,u}} d(v_0,s)\mu(s),$$

where we used the triangle inequality and the definition of $S^+_{v_0,u}$ to make our estimates. Hence, for any $u \in V(\Gamma) \setminus \{v_0\}$, we can make the following implication

$$\sum_{s\in S^+_{v_0,u}} d(v_0,s)\mu(s) < \frac{1}{2} d(v_0,u)\mu(v_0) \quad \Rightarrow \quad \rho^{(2)}(v_0,u) < 0. \tag{6.5}$$

Note that $S^+_{v_0,u} \cap B_{v_0}(d(v_0,u)/2) = \emptyset$ and, hence, $S^+_{v_0,u} \subseteq V(\Gamma) \setminus B_{v_0}(d(v_0,u)/2)$. It implies that

$$\sum_{s\in S^+_{v_0,u}} d(v_0,s)\mu(s) \leq \sum_{s\in V(\Gamma)\setminus B_{v_0}(d(v_0,u)/2)} d(v_0,s)\mu(s).$$

The sum $\sum_{s\in V(\Gamma)\setminus B_{v_0}(d(v_0,u)/2)} d(v_0,s)\mu_n(s)$ is a part of the sum for $M^{(1)}(v_0)$ that converges by our assumption. Since $M^{(1)}(v_0)$ is finite, it follows that for every $\delta > 0$ there exists a number m such that for every u with $d(u,v_0) > m$, the inequality

$$\sum_{s\in V(\Gamma)\setminus B_{v_0}(m/2)} d(v_0,s)\mu(s) < \delta \tag{6.6}$$

holds (summation over the complements of the balls converges to zero, as in the proof of Theorem 6.15). Choose m for $\delta := \frac{1}{8}\mu(v_0)$. It follows from (6.5) that for every $u \in V(\Gamma) \setminus B_{v_0}(m)$, $\rho^{(2)}(v_0,u) < 0$.

Next, if we prove that

$$\mathbf{P}\left(\exists N \text{ s.t. } \forall n > N\ \forall u \in V(\Gamma) \setminus B_{v_0}(m),\quad \sum_{s\in S^+_{v_0,u}} d(v_0,s)\mu_n(s) < \frac{1}{2} m\mu_n(v_0)\right) = 1, \tag{6.7}$$

then, from (6.5), with μ_n instead of μ, we immediately get

$$\mathbf{P}\big(\exists N \text{ s.t. } \forall n > N \ \forall u \in V(\Gamma) \setminus B_{v_0}(m), \quad \rho_n^{(2)}(v_0, u) < 0\big) = 1,$$

which implies (6.4) (using the fact that v is the singleton center-set and, hence, $\rho^{(2)}(v, v_0) \leq 0$ together with Proposition 6.12 and remembering that $\rho_n \to \rho$ almost surely as $n \to \infty$).

Fix $\varepsilon := \frac{1}{8}\mu(v_0)$. Since $M^{(1)}(v_0)$ is an expectation of a real valued random variable $d(v_0, \xi_1)$, we know from the strong law of large numbers that for any ε

$$\mathbf{P}(\exists N \ \ \forall n > N, \ \ |M_n^{(1)}(v_0) - M^{(1)}(v_0)| < \varepsilon) = 1. \tag{6.8}$$

Moreover, it follows from the strong law of large numbers for relative frequencies that

$$\mathbf{P}\left(\exists N \ \ \forall n > N \ \ \forall s \in B_{v_0}(m), \ \ |\mu_n(s) - \mu(s)| < \frac{\varepsilon}{m|B_{v_0}(m)|}\right) = 1. \tag{6.9}$$

and, hence,

$$\mathbf{P}\left(\exists N \ \ \forall n > N, \ \ \left|\sum_{s \in B_{v_0}(m)} d(v_0, s)\mu_n(s) - \sum_{s \in B_{v_0}(m)} d(v_0, s)\mu(s)\right| < \varepsilon\right) = 1. \tag{6.10}$$

Now, observe that

$$\sum_{s \in S^+_{v_0,u}} d(v_0, s)\mu_n(s) \leq M_n^{(1)}(v_0) - \sum_{s \in B_{v_0}(m/2)} d(v_0, s)\mu_n(s)$$

$$= \big(M_n^{(1)}(v_0) - M^{(1)}(v_0)\big) + \left(M^{(1)}(v_0) - \sum_{s \in B_{v_0}(m/2)} d(v_0, s)\mu(s)\right) +$$

$$+ \left(\sum_{s \in B_{v_0}(m/2)} d(v_0, s)\mu(s) - \sum_{s \in B_{v_0}(m/2)} d(v_0, s)\mu_n(s)\right) \leq \varepsilon + \varepsilon + \varepsilon = 3\varepsilon,$$

where we used (6.8), (6.10), and observed that

$$M^{(1)}(v_0) - \sum_{s \in B_{v_0}(m/2)} d(v_0, s)\mu(s) = \sum_{s \in V(\Gamma) \setminus B_{v_0}(m/2)} d(v_0, s)\mu(s) < \delta = \frac{1}{8}\mu(v_0) = \varepsilon.$$

It follows from (6.8), (6.6) and (6.10) that

$$\mathbf{P}\left(\exists N \text{ s.t. } \forall n > N \ \forall u \in V(\Gamma) \setminus B_{v_0}(m), \quad \sum_{s \in S^+_{v_0,u}} d(v_0,s)\mu_n(s) < 3\varepsilon\right) = 1.$$

Now, as in (6.9) we have

$$\mathbf{P}\Big(\exists N \text{ s.t. } \forall n > N, \quad |\mu_n(v_0) - \mu(v_0)| < \varepsilon\Big) = 1.$$

Combining two previous equalities we get

$$\mathbf{P}\left(\exists N \text{ s.t. } \forall n > N \ \forall u \in V(\Gamma) \setminus B_{v_0}(m), \quad \sum_{s \in S^+_{v_0,u}} d(v_0,s)\mu_n(s) < \frac{3}{8}\mu(v_0) < \frac{1}{2}\mu_n(v_0)\right) = 1.$$

Keeping in mind that $m \geq 1$, we conclude that (6.7) and, hence, (6.4) hold.

□

Finally, we observe that all the statements from Chapter 5 concerning cut points and configurations of mean-set hold for the central order framework. We restate them below, using $\rho^{(2)}(\cdot,\cdot)$ for the convenience of the reader.

Theorem 6.21 (Cut point inequality)**.** *Let μ be a probability distribution on a connected graph Γ such that $M^{(1)}$ is defined. Let v_0 be a cut-point in Γ, and v_1, v_2 belong to distinct components of $\Gamma - \{v_0\}$. Then the inequality*

$$d(v_0,v_2)\rho^{(2)}(v_1,v_0) + d(v_0,v_1)\rho^{(2)}(v_2,v_0) \geq C > 0$$

holds, where $C = C(v_0,v_1,v_2) = d(v_0,v_2)d(v_0,v_1)(d(v_0,v_1) + d(v_0,v_2))$.

Corollary 6.22 (Cut Point Lemma)**.** *Let Γ be a connected graph, v_0 a cut-point in Γ. If v_1 and v_2 belong to distinct connected components of $\Gamma - \{v_0\}$, then the inequalities $\rho^{(2)}(v_1,v_0) \leq 0$ and $\rho^{(2)}(v_2,v_0) \leq 0$ cannot hold simultaneously.*

Corollary 6.23 (Mean-set in a graph with a cut-point)**.** *Let v_0 be a cut-point in a graph Γ and $\Gamma - \{v_0\}$ a disjoint union of connected components $\Gamma_1, \ldots, \Gamma_k$. Then for any distribution μ on Γ there exists a unique $i = 1, \ldots, k$ such that $\mathbb{E}\mu \subseteq V(\Gamma_i) \cup \{v_0\}$.*

Corollary 6.24 (Mean-set in a graph with several cut-points)**.** *Let $v_1, \ldots, v_n$ be cut-points in a graph Γ and $\Gamma - \{v_1, \ldots, v_n\}$ a disjoint union of connected components $\Gamma_1, \ldots, \Gamma_k$. Then for any distribution μ on Γ there exists a unique $i = 1, \ldots, k$ such that $\mathbb{E}\mu \subseteq V(\Gamma_i) \cup \{v_1, \ldots, v_n\}$.*

Corollary 6.25. *Let G_1 and G_2 be finitely generated groups and $G = G_1 * G_2$ a free product of G_1 and G_2. Then for any distribution μ on G the set $\mathbb{E}\mu$ is a subset of elements of the forms gG_1 or gG_2 for some element $g \in G$.*

Proposition 6.26. *Let Γ be a tree and μ a probability measure on $V(\Gamma)$. Then $|\mathbb{E}\mu| \leq 2$. Moreover, if $\mathbb{E}\mu = \{u, v\}$ then u and v are connected in Γ.*

Corollary 6.27. *Let μ be a probability distribution on a free group F. Then $|\mathbb{E}\mu| \leq 2$.*

Also, observe that all results of Chapter 5, Section 5.2, hold when used with the order defined by $\rho^{(2)}$.

6.2.3 Mean sets and L^c-spaces

Fix an order of the vertices $(v_1, v_2 \ldots)$ in Γ. Relative to the fixed order on $V(\Gamma)$ one can associate to a vertex $v \in V(\Gamma)$ the vector

$$\overline{d}_v := (d(v, v_1)\sqrt[c]{\mu(v_1)},\ d(v, v_2)\sqrt[c]{\mu(v_2)}, \ldots)$$

in the infinite-dimensional space $\mathbf{R}^\infty$. The subspace of vectors $\overline{d} = (d_1, d_2, \ldots)$ in $\mathbf{R}^\infty$ with

$$\|\overline{d}\|_c := \sqrt[c]{\sum_i d_i^c} < \infty$$

is called the L^c space and the function $\|\cdot\|_c$ satisfies all the properties of a norm in that space. Hence, if $M^{(c)}$ is defined on $V(\Gamma)$ then the set $\{\overline{d}_v \mid v \in V(\Gamma)\}$ is a subset of an L^c space. Furthermore, if $M^{(c)}$ is totally defined, then by Proposition 6.17 the inequality $u <^{(c)} v$ holds if and only if $\|\overline{d}_u\|_c < \|\overline{d}_v\|_c$. In other words, in that case the set $\mathbb{E}^{(c)}$ is the set of vertices in $V(\Gamma)$ with the shortest corresponding vectors $\overline{d}_v$. The situation is more complicated when $\|\overline{d}_v\|_c$ is not finite. The next proposition shows that we cannot use L^c-norm point of view in Definition 6.11 of central order in that case. The reason for this impossibility is that L^c-norms cannot distinguish vertices if $M^{(c)} = \infty$, i.e., square roots of weight functions do not separate the vertices of Γ, but, to the contrary, collapse them all to just one point.

Proposition 6.28. *For any $c \in \mathbb{N}$, locally finite graph Γ, distribution μ on Γ, and vertices $u, v \in \Gamma$,*

$$\gamma = \lim_{n\to\infty} \left(\sqrt[c]{\sum_{i=1}^{n} d^c(v, v_i)\mu(v_i)} - \sqrt[c]{\sum_{i=1}^{n} d^c(u, v_i)\mu(v_i)} \right)$$

exists for any ordering of the vertices. In particular, $\gamma \equiv 0$ if $M^{(c)} = \infty$.

Proof. If $M^{(c)}$ is finite, then $\gamma = \sqrt[c]{M^{(c)}(v)} - \sqrt[c]{M^{(c)}(u)}$ and the proposition is proved. Also, the case when $c = 1$ is proved in Section 6.1. Assume that $c \geq 2$ and $M^{(c)} = \infty$, i.e., for any $v \in V(\Gamma)$,

$$\sum_{i=1}^{n} d^c(v, v_i)\mu(v_i) \to \infty \quad \text{as} \quad n \to \infty. \tag{6.11}$$

We claim that in the case under consideration γ is identically 0. Our main tool in the proof of this claim is the following inequality

$$\left| \sqrt[c]{(a^c + b^c)} - b \right| < \frac{a^c}{cb^{c-1}}$$

which holds for any $a, b \in \mathbf{R}_+$. The inequality can be deduced by applying the Binomial theorem to $\left(b + \frac{a^c}{cb^{c-1}}\right)^c$.

Fix a sequence $\{v_1, v_2, \ldots,\}$ and a number $\varepsilon > 0$. Our goal is to show that there exists a constant N^* such that for any $n > N^*$

$$\left| \sqrt[c]{\sum_{i=1}^{n} d^c(v, v_i)\mu(v_i)} - \sqrt[c]{\sum_{i=1}^{n} d^c(u, v_i)\mu(v_i)} \right| < \varepsilon.$$

For every i define

$$\delta(i) = d(v, v_i) - d(u, v_i).$$

It follows from the triangle inequality that the function δ is bounded by $d = d(v,u)$. Since μ is a probability measure, we can choose a number N such that

$$\mu(v_N, v_{N+1}, \ldots) < \varepsilon/3d.$$

It is convenient to define the numbers

$$\alpha_1 := \sqrt[c]{\sum_{i=1}^{N-1} d^c(v,v_i)\mu(v_i)} \quad \text{and} \quad \alpha_2 := \sqrt[c]{\sum_{i=1}^{N-1} d^c(u,v_i)\mu(v_i)}$$

and for every $M > N$ define

$$\beta_{1,M} := \sqrt[c]{\sum_{i=N}^{M} d^c(v,v_i)\mu(v_i)} \quad \text{and} \quad \beta_{2,M} = \sqrt[c]{\sum_{i=N}^{M} d^c(u,v_i)\mu(v_i)}.$$

The inequalities below follow from triangle inequality for the norm $\|\cdot\|_c$

$$|\alpha_1 - \alpha_2| \le d \quad \text{and} \quad |\beta_{1,M} - \beta_{2,M}| < \varepsilon/3$$

for every $M > N$. It follows from (6.11) that we can choose a number $N^* > N$ such that

$$\frac{\alpha_1^c}{c\beta_{1,N^*}^{c-1}} < \varepsilon/3 \quad \text{and} \quad \frac{\alpha_2^c}{c\beta_{2,N^*}^{c-1}} < \varepsilon/3.$$

Moreover, since $\beta_{i,n}$ is non-decreasing in n, it follows that for every $n > N^*$ the inequalities $\frac{\alpha_1^c}{c\beta_{1,n}^{c-1}} < \varepsilon/3$ and $\frac{\alpha_2^c}{c\beta_{2,n}^{c-1}} < \varepsilon/3$ hold. Therefore, we get for every $n > N^*$

$$\begin{aligned}
&\left| \sqrt[c]{\sum_{i=1}^{n} d^c(v,v_i)\mu(v_i)} - \sqrt[c]{\sum_{i=1}^{n} d^c(u,v_i)\mu(v_i)} \right| \\
&= \left| \sqrt[c]{\alpha_1^c + \beta_{1,n}^c} - \sqrt[c]{\alpha_2^c + \beta_{2,n}^c} \right| \\
&\le \left| \sqrt[c]{\alpha_1^c + \beta_{1,n}^c} - \beta_{1,n} \right| + |\beta_{1,n} - \beta_{2,n}| + \left| \beta_{2,n} - \sqrt[c]{\alpha_2^c + \beta_{2,n}^c} \right| \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon,
\end{aligned}$$

and the proposition is proved.

□

6.2.4 Examples of central order

Let us conclude the theme of ordering of the vertices of Γ by giving several examples of central order, illuminating the idea of comparability of vertices, in general, and, hopefully, facilitating a better perception of the reader, in particular. The example below demonstrates the case when $\mathbb{E} \equiv \mathbb{E}^{(2)}$ contains all the vertices of an infinite graph, i.e., when $|\mathbb{E}(\cdot)| = \infty$.

Example 6.29. Consider $\mathbf{Z}$ with the probability measure μ given by

$$\mu(n) = \begin{cases} c/n^2 & n > 0; \\ 0 & n = 0; \\ c/n^2 & n < 0; \end{cases}$$

where c is a constant normalizing the measure on $\mathbf{Z}$. The classical expectation is not defined for such (Γ, μ) because the series $\sum_i i \cdot \dfrac{c}{i^2}$ diverges. For such μ the order $<^{(2)}$ is defined for no pair of elements, meaning that all the vertices are not comparable under the central order.

Indeed, let $u < v$ be elements of $\mathbf{Z}$. Then $\rho^{(2)}(u, v) = \sum_{n \in \mathbf{Z} \setminus \{0\}} (d^2(u, n) - d^2(v, n))\mu(n) = \sum_{n \in \mathbf{Z} \setminus \{0\}} ((u^2 - v^2) + 2n(v - u)) \frac{c}{n^2}$ which diverges whenever $v \neq u$. Similarly, $\rho^{(2)}(v, u)$ is undefined as well. Hence, the relation $<^{(2)}$ for the distribution μ under consideration is empty. Therefore, $\mathbb{E}^{(2)} = \mathbf{Z}$ in this case.

□

Next example shows that order relation can be non-trivial (not empty), and, at the same time we can have infinitely many points in the center-set.

Example 6.30. Consider a probability measure on $\mathbf{Z}^2$ defined as follows

$$\mu((n, i)) = \begin{cases} c/n^2 & n > 0 \text{ and } i = 0; \\ 0 & n = 0 \text{ or } i \neq 0; \\ c/n^2 & n < 0 \text{ and } i = 0. \end{cases}$$

It is easy to check that $(x_1, y_1) <^{(2)} (x_2, y_2)$ if and only if the conditions $|y_1| < |y_2|$ and $|x_2 - x_1| \le |y_2| - |y_1|$. This condition is visualized for a particular point $v = (3, 2)$ in Figure 6.1. In that Figure, vertices marked with black are less than $v = (3, 2)$ relative to $<^{(2)}$.

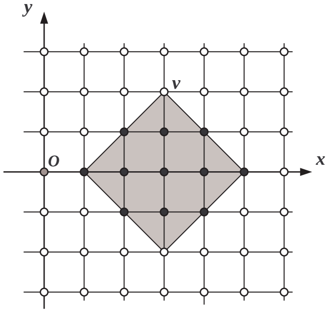

Figure 6.1: Example of central order relation $<^{(2)}$ in $\mathbf{Z}^2$.

It is not hard to see that $\mathbb{E}(\mu) = \mathbf{Z} \times \{0\}$ in this example. Indeed, let $v_1 = (x_1, y_1) \in \mathbf{Z}^2$, $v_2 = (x_2, y_2) \in \mathbf{Z}^2$. We have

$$\rho^{(2)}(v_1, v_2) = \sum_{i=-\infty}^{\infty} \Big((|y_1| + |x_1 - i|)^2 - (|y_2| + |x_2 - i|)^2 \Big) \frac{c}{i^2}$$

It is easy to see that $\rho^{(2)}((x_1, y_1), (x_2, y_2)) = \rho^{(2)}((x_1, y_1), (x_2, -y_2)) = \rho^{(2)}((x_1, -y_1), (x_2, -y_2))$. Hence we assume that $y_1 \ge 0$ and $y_2 \ge 0$. Now we have $\rho^{(2)}(v_1, v_2) =$

$$= \sum_{i=-\infty}^{\min\{x_1,x_2\}} \Big((y_1 + |x_1 - i|)^2 - (y_2 + |x_2 - i|)^2 \Big) \frac{c}{i^2}$$

$$+ \sum_{i=\min\{x_1,x_2\}+1}^{\max\{x_1,x_2\}-1} \Big((y_1 + |x_1 - i|)^2 - (y_2 + |x_2 - i|)^2 \Big) \frac{c}{i^2}$$

$$+ \sum_{i=\max\{x_1,x_2\}}^{\infty} \Big((y_1 + |x_1 - i|)^2 - (y_2 + |x_2 - i|)^2 \Big) \frac{c}{i^2}.$$

The summand in the middle is a constant. The first summand equals to

$$\sum_{i=-\infty}^{\min\{x_1,x_2\}} \Big(x_1^2 + y_1^2 - x_2^2 - y_2^2 + 2x_1y_1 - 2x_2y_2 + 2i(x_2 + y_2 - x_1 - y_1) \Big) \frac{c}{i^2}$$

which converges if and only if $x_2 + y_2 - x_1 - y_1 = 0$, diverges to $-\infty$ if and only if $x_2 + y_2 - x_1 - y_1 < 0$, and diverges to ∞ if and only if $x_2 + y_2 - x_1 - y_1 > 0$. The third summand equals to

$$\sum_{i=\max\{x_1,x_2\}}^{\infty} \Big(x_1^2 + y_1^2 - x_2^2 - y_2^2 - 2x_1y_1 + 2x_2y_2 + 2i(x_2 - y_2 - x_1 + y_1)\Big)\frac{c}{i^2}.$$

Summing everything up we see that for $v_1 = (x_1, y_1)$ and $v_2 = (x_2, y_2)$, $\rho^{(2)}(v_1, v_2) < 0$ if and only if

$$x_2 + y_2 \leq x_1 + y_1 \text{ and } x_2 - y_2 < x_1 - y_1$$

or

$$x_2 + y_2 < x_1 + y_1 \text{ and } x_2 - y_2 \leq x_1 - y_1.$$

Taking into account negative values for y_1 and y_2 we encode these conditions into $|y_1| < |y_2|$ and $|x_2 - x_1| \leq |y_2| - |y_1|$. □

The next example demonstrates the case when $\mathbb{E}^{(2)} = \emptyset$ for some infinite graph.

Example 6.31. Consider $\mathbf{Z}$ with the probability measure μ given by

$$\mu(n) = \begin{cases} c/n^2, & n > 0; \\ 0, & n \leq 0. \end{cases}$$

where c is a constant normalizing the measure on $\mathbf{Z}$. The classical expectation is not defined for such (Γ, μ) because the series $\sum_{i=1}^{\infty} i \cdot \frac{c}{i^2}$ diverges. We claim that for so defined μ and for any pair of integers u, v the inequality $u <^{(2)} v$ holds if and only if $u > v$ (intuitively: the farther the element is to infinity, the smaller the element is; it means that the center is empty).

Indeed, $\rho^{(2)}(u, v) = \sum_{n\in\mathbb{N}}((u^2 - v^2) + 2n(v - u))\frac{c}{n^2}$ which properly diverges to $-\infty$ if $v < u$ and properly diverges to ∞ if $v > u$. In other words, we have $\ldots 4 <^{(2)} 3 <^{(2)} 2 <^{(2)} 1$. Therefore the center set $\mathbb{E}^2$ is empty.

We may interpret this as follows: "ray determines an element on the boundary", i.e., the farther to ∞, the smaller the element is, with respect to the comparison relation $<^{(c)}$. It means that the center $\mathbb{E}^2$ is empty. It is natural to say that, in this case, the center belongs to the "boundary" of $\mathbf{Z}$, but this notion requires further research.

$\square$

Chapter 7

Practical applications to cryptanalysis

7.1 Introduction

The group-based cryptography attracted a lot attention after the invention of the Anshel-Anshel-Goldfeld (1) and Ko-Lee et al. (8) key-exchange protocols. Since than many new cryptographic protocols based on infinite groups were invented and analyzed. In particular, public-key *authentication protocols* such as the Sibert et al. protocol in (34) (see also (10)) already mentioned in Chapter 1, Section 1.1.4, were proposed. As the name itself suggests, authentication schemes (protocols), also called identification protocols, are basically systems designed for the purpose of authentication (identification).

In this chapter we use the technique of mean-sets to design a heuristic attack on this particular cryptographic primitive. We describe and analyze the scheme, proving that it does not meet the security compliances claimed in (34). In addition, we test the protocol, conducting some actual experiments, which show in practice that the secret information can be revealed.

It is claimed in (34), as well as in the later survey (10), that the security of this scheme relies on the difficulty of the famous algebraic *conjugacy search problem* problem for the platform group G. For an element $w \in G$, we define its *conjugate* by $s \in G$ to be an element $t = s^{-1}ws$. This is a slight notational modification of conjugation used in Sibert's article (34), namely, sws^{-1}. Nevertheless, since this is only about the notational convention, the change does not affect the protocol at all. Recall that the *conjugacy search problem* is stated as follows:

> If $w, t \in G$ and t is a conjugate of w, find $s \in G$ (sometimes called a *witness*), such that $t = s^{-1}ws$, provided that such s exists.

The use of algebraic methods that rely on certain algebraic problems being algorithmically difficult, is traditional in group-based cryptography. In particular, cryptographic schemes usually have security features that depend on the complexity of such problems. Naturally, different attacks on these systems are designed, trying to break them, again using algebraic approaches.

In this chapter we explain the practical use of our probability theory on graphs and groups in detail. We look at the algebraic problem from the probabilistic angle and attack the security of the protocol from a totally unexpected side – it has nothing to do with the originally claimed security assumptions and at no point uses the conjugacy operation. We obtain the secret information using our strong law of large numbers, the "shift“ property, and an algorithm solving an underlying computational problem of $\mathbb{S}_n$. In other words, if the platform group G has efficiently computable length function, we can show that the scheme is insecure using our probabilistic approach. It turns out that even “reasonable” approximation of the length function does the job. We provide experimental evidence that our algorithm works even for groups with no efficiently computable length function such as braid groups.

7.2 Zero-Knowledge (security-preservation) property. Some ideas.

Zero-knowledge property is considered to be a very desirable property of interactive (randomized) proofs of identity used in modern cryptography. Recall that modern cryptography is concerned with the construction of efficient schemes for which it is infeasible to violate the security feature (schemes that are computationally easy to operate, but hard to foil).

7.2.1 Intuition

As we mentioned in Chapter 1, one of the various cryptographic problems is the *authentication (identification) problem*: the Prover wishes to prove his/her identity to the Verifier via some private key without enabling an intruder watching the communication to deduce anything about the secret key. Many existing authentication schemes (protocols) use the so-called zero-knowledge (”security-preserving“) proofs ((12), (17)) as a major tool in verifying that the secret-based actions of the Prover are correct, without revealing these secrets. The ”baby-example“ of such ”security-preserving“ proof (protocol) was given in Chapter 1, Section 1.1.4, so that the reader could easily gain some insight into what it is about.

Let us just add some more ideas to this intuition. Recall that *static (non-interactive) proof)* is a fixed sequence of statements that are either axioms or are derived from previous statements (see (16)). As opposed to that, *interactive (dynamic, randomized) proof* is a (multi-round) randomized protocol for two parties, called the Verifier and the Prover, in which the Prover wishes to convince the Verifier of the validity of a given assertion. The reader can find more information on that topic in (10) and (17). Loosely speaking, we can think of interactive proof as of a game between the Verifier and the Prover consisting of questions asked by the Verifier, to which the Prover must reply ”convincingly“. As the reader can see in many sources, the interactive proof of knowledge should satisfy *completeness* and *soundness* properties ((12), (17)). It means that if the statement is true, it should be accepted by the Verifier with high probability. In addition, if the assertion is false then the Verifier must reject with high probability, i.e., so-called *soundness error* should be small. We can say that interactive proofs are *probabilistic proofs* by their nature, because there is some small probability, the soundness error, that a cheating Prover will be able to convince the Verifier of a false statement. However, the soundness error can usually be decreased to negligibly small values by repetitions of steps.

It happens that completeness and soundness properties are not enough to make an interactive proof secure. The Verifier can be a potential adversary (cheating verifier, intruder, eavesdropper) that tries to gain knowledge (secret) from the Prover. This means that we need another property of the randomized proof (identification protocol, in particular) to ensure that security is preserved. The property needed is exactly the *zero-knowledge property*; it assures the user (the Prover) that he/she does not compromise the privacy of his/her secrets in the process of providing the proof (of identity, for instance). In other words, one can think of zero-knowledge as of preservation of security on the intuitive level.

7.2.2 More precise realization

The interested reader can find in (10) that the precise realization of zero-knowledge property is as follows: using public key data only, one can construct a probabilistic Turing machine able to simulate the instances of the communication between the Prover and the Verifier in a way that cannot be distinguished from the real communication – if this can be done, no information about the secret data can be extracted from the exchanges performed in the scheme.

Let us not dwell on this definition for too long though; this is not of a paramount importance for the present exposition. Precise realization is much less illuminating for an uninitiated reader than the intuitive notion given above. For readers who wish to go deeper in these matters, we can suggest the surveys (10) and (17) which provide a lot of additional information and sources for this material.

A useful comment can be made in connection with the precise notion of zero-knowledge property. Observe that the definition above involves the notion of indistinguishability (or, more loosely, ”similarity“), which is open to interpretations. In fact, there are three interpretation, yielding three different notions of zero-knowledge that have been commonly used in the literature:

(PZ) *Perfect Zero-Knowledge* requires two communications to be indistinguishable.

(SZ) *Statistical Zero-Knowledge* requires two communications be statistically close in a sense of the variation distance.

(CZ) *Computational Zero-Knowledge* requires two communications be computationally indistinguishable.

We all live in the real world, in which ”perfect“ is almost the same as impossible. Indeed, (PZ) is the most strict definition of zero-knowledge that is very rarely used in practice. In practice, we want to be able to compute, and we can analyze only what we can, in fact, actually compute. Thus, the last notion (CZ) of the zero-knowledge is the most practical and liberal notion that is used more frequently than the others, and this is exactly the property we can analyze for authentication protocols in practice.

7.3 The protocol

The Sibert et al. protocol is an iterated two-party three-pass identification protocol, in which one of the parties (called Prover) wants to prove his identity to the other party (called Verifier) via some secret private key. A general version of the protocol uses the conjugation operation over some (infinite, non-commutative) group G for which the conjugacy search problem is presumably hard (Braid group, in particular). The authors of (34) claim that the scheme described below is ”zero-knowledge interactive proof of knowledge“, at least computationally, in practice.

7.3.1 Set up and description

The set up for the protocol is as follows. Let the Prover's *private key* be an element $s \in G$ and his *public key* a pair (w, t), where w is an arbitrary element of the group G, called *the base element*, and $t = s^{-1}ws$. The authentication protocol goes as follows:

1. The Prover chooses a random element $r \in G$ and sends the element $x = r^{-1}tr$, called the *commitment*, to the Verifier. Now the Verifier knows (w, t, x).

2. The Verifier chooses a random bit c, called the *challenge*, and sends it to the Prover.

    - If $c = 0$, then the Prover sends $y = r$ to the Verifier and the Verifier checks if the equality $x = y^{-1}ty$ is satisfied.

    - If $c = 1$, then the Prover sends $y = sr$ to the Verifier and the Verifier checks if the equality $x = y^{-1}wy$ is satisfied.

The authors of the protocol say the the scheme should be repeated k times to achieve k bits of security, i.e., to guarantee the probability of 2^{-k} of foiling the scheme, which is negligible.

Note that if an intruder (called the Eavesdropper) can efficiently compute an element $s' \in G$ such that $t = s'^{-1}ws'$, i.e., if the Eavesdropper can solve *the conjugacy search problem* for G, then he/she can authenticate as the Prover (s' can be used in place of s, and the identification will go through). Therefore, as indicated in (34), the security of this protocol is based on the complexity of the conjugacy search problem, and only this problem, thus, avoiding the attacks and making the scheme seemingly unbreakable.

Originally it was proposed to use braid groups B_n as platform groups (in this and many other protocols) because there was no feasible solution of the conjugacy search problem for B_n known. This motivated a lot of research about braid groups. As a result of very recent developments (in 2007, 2008), there is an opinion that the conjugacy search problem for B_n can be solved in polynomial time. The reader can find more information on this issue in (4), (6), (5), where J. Birman and her co-authors express this opinion and provide experimental evidence to support it. If that is true in fact, then the authentication protocol under consideration is insecure for B_n because the underlying algebraic problem can be efficiently solved.

We show in the present chapter that it is not necessary to solve the conjugacy search problem for G to break the scheme, i.e., that the algebraic approach to the problem is not the only one possible. Instead, we analyze zero-knowledge property of the protocol by employing ideas from probability theory and show that the protocol is often insecure under a mild assumption of existence of an efficiently computable length function for the platform group G.

7.3.2 Analysis

Now, let us do the analysis. It is trivial to check that a correct answer of the Prover at the second step leads to acceptance by the Verifier. Indeed, if $c = 0$, then $y^{-1}ty = r^{-1}tr = x$, and if $c = 1$, then $y^{-1}wy = r^{-1}s^{-1}wsr = r^{-1}tr = x$. Hence the Verifier accepts a correct answer at each repetition, so he accepts the Prover's proof of identity with probability one.

It is somewhat less obvious why such an arrangement (with a random bit) is needed; it may seem that the Prover could just reveal $y = sr$: this does not reveal the secret s, and yet allows the Verifier to verify the equality $x = y^{-1}wy$. The point is that in this case, the adversary (eavesdropper), called Eve for the moment, who wants to impersonate the Prover can just take an arbitrary element u and send $x = u^{-1}wu$ to the Verifier as a commitment. Then this u would play the same role as $y = sr$ for verification purposes. Similarly, if Eve knew for sure that the Prover would send $y = r$ for verification purposes, she would use an arbitrary element u in place of r at the commitment step. These considerations also show that the above authentication protocol should be run several times for better reliability because with a single run, Eve can successfully impersonate Prover with probability $\frac{1}{2}$. After k runs, this probability goes down to $\frac{1}{2^k}$, which is considered to be negligible. The scheme is reliable (complete) with a negligible soundness error (see (34) for details). In addition, the authors claim that the protocol is secure in practice and ensures the confidentiality of the private key; namely, it is computationally zero-knowledge (see (34) again for explanations why the scheme does not possess the perfect zero-knowledge property (PZ)).

Now, let us see how we may develop this analysis further, having our new theoretical tools at our disposal. Observe that the Prover sends to the Verifier a sequence of random elements of 2 types: r and sr, where r is a randomly generated element and s is a secretly chosen fixed element. We can make a table where each row corresponds to a single round of the protocol:

Round	Challenge	Response # 1	Response # 2
1	$c = 1$	–	sr_1
2	$c = 0$	r_2	–
3	$c = 0$	r_3	–
4	$c = 1$	–	sr_4
5	$c = 0$	r_5	–
...	...	...	...
n	$c = 0$	r_n	–

After the Prover's authentication, the intruder, named Eve, possesses 2 sets of elements corresponding to $c = 0$ and $c = 1$ respectively:

$$R_0 = \{r_{i_1}, \dots, r_{i_k}\}$$

and

$$R_1 = \{sr_{j_1}, \ldots, sr_{j_{n-k}}\}.$$

Eve's goal is to recover the element s based on the intercepted sets above. We observe that the Prover cannot guess the Verifier's request c, he chooses r before he gets c from the Verifier and hence we make the following important assumption:

> The random elements $r_{i_1}, \ldots, r_{i_k}$ and $r_{j_1}, \ldots, r_{j_{n-k}}$ have the same distribution, i.e., all these elements are generated by the same random generator.

The theory developed in this work allows us to show that eventually, after sufficiently many iterations, Eve is able to recover the secret element s. The "shift" property (1.5) of the expectation (mean-set) of the group-valued random element, together with the strong law of large numbers for groups, allow us to define the following simple attack. Eve, for the collected sets of elements $R_0 = \{r_{i_1}, \ldots, r_{i_k}\}$ and $R_1 = \{sr_{j_1}, \ldots, sr_{j_{n-k}}\}$, computes the set

$$\mathbb{S}(sr_{j_1}, \ldots, sr_{j_{n-k}}) \cdot [\mathbb{S}(r_{i_1}, \ldots, r_{i_k})]^{-1}.$$

When n is sufficiently large, this set contains the private key s, or rather, a very good guess of what s is. We conclude that the proposed zero-knowledge authentication (security-preserving) protocol is not secure in practice, namely, it does not ensure the confidentiality of the private key.

This is a completely different approach to cryptanalysis of such protocols - it does not rely on the solvability of hard algebraic problems, but, instead, attacks the protocol from the probabilistic angle of view.

7.4 Effective computation of a mean-set

Clearly, in order to apply our results in practice, we have to be able to compute $\mathbb{S}_n$ efficiently (see the discussion in Chapter 5). In Section 5.2, we outlined several problems arising when trying to compute $\mathbb{S}_n$. One of them is concerned with minimizing the sample weight function over G, and the other problem has to do with computation of the distance function $d(\cdot,\cdot)$, which is very difficult for braid groups.

To avoid the first problem we devise a heuristic procedure for this task, namely, we propose a Direct Descent Heuristic Algorithm 5.11 that finds a local minimum for a given function f. As proved in Section 5.2, if the function f satisfies certain local monotonicity properties, then our procedure achieves the desired result. In the same section we prove that our weight function, in fact, satisfies the desired local monotonicity and local finiteness properties for free groups.

Let us repeat the algorithm here, but this time instead of using a generic function f on a graph Γ, we shall use the sample weight function $M_n(\cdot)$ that comes from a sample of group random elements $\{g_1,\ldots,g_n\}$ on a finitely-generated group G.

Algorithm 7.1 ((Direct Descent Heuristic for M_n))**.**
INPUT: A group G with a finite set of generators $X \subseteq G$ and a sequence of elements $\{g_1,\ldots,g_n\}$ in G.
OUTPUT: An element $g \in G$ that locally minimizes $M_n(\cdot)$.
COMPUTATIONS:

```
A. Choose a random $g \in G$ according to some probability measure $\nu$ on $G$.
B. If for every $x \in X^{\pm 1}$, $M_n(g) \le M_n(gx)$, then output $g$.
C. Otherwise put $g \leftarrow gx$, where $x \in X^{\pm 1}$ is any generator such that $M_n(g) > M_n(gx)$ and goto step B.
```

7.4.1 Length function for braids

The second problem of computing $\mathbb{S}_n$ concerns practical computations of length function in G. It turns out that we need a relatively mild assumption to deal with it – the existence of an efficiently computable distance function $d_X(\cdot,\cdot)$; even a “reasonable” approximation of the length function does the job.

There are two length functions for braids that were used in cryptanalysis of braid based systems (see (11)):

- the length defined as the geodesic length $|\cdot|$ relative to the set of generators $\{\sigma_1, \ldots, \sigma_{n-1}\}$, i.e. the length of the shortest path in the corresponding Cayley graph of a group B_n;
- the canonical length of the Garside normal form $|\cdot|_\Delta$.

Unfortunately both of them are not practically useful because of the following reasons.

The geodesic length of a braid denoted by $|\cdot|$ seems to be the best candidate. However, there is no known efficient algorithm for computing $|\cdot|$. Moreover, it is proved in (31) that the set of geodesic braids in B_∞ is co-NP complete.

The canonical length of the Garside normal form $|\cdot|_\Delta$ is efficiently computable but very crude, in a sense that many braids consisting of many crossings have very small lengths (see (11)).

In this work we use the method to approximate geodesic length proposed in (27). Even though it does not guarantee the optimal result, it proved to be practically useful in a series of attacks, see (18; 15; 29; 30).

7.5 The Mean-Set Attack

In this section we explain how we can actually attack the Sibert's protocol in practice. Let $\xi_1, \xi_2, \ldots$ be a sequence of *session keys* (this is what we saw above as r_{i_k} and sr_{j_k}, but written as one sequence in order of their generation). Let R_0 and R_1 be sequences of two types of responses of the Prover in the protocol corresponding to the challenge bits $c_i = 0$ and $c_i = 1$ respectively, $i = 1, 2, \ldots$. In other words,

$$R_0 = \Big(\xi_i \mid c_i = 0, i = 1, \ldots, n\Big)$$

$$R_1 = \Big(\xi_i \mid c_i = 1, i = 1, \ldots, n\Big)$$

We may find this notation useful when we want to be very specific about these sets. By Theorem 4.2, there exists a constant C such that

$$\mathbf{P}\Big(\mathbb{S}(R_0) \not\subset \mathbb{E}\xi\Big) \leq C/|R_0| \text{ and } \mathbf{P}\Big(\mathbb{S}(R_1) \not\subset s\mathbb{E}\xi\Big) \leq C/|R_1|.$$

In other words, the probability that the sample mean-sets contain an undesired (not central) element decreases linearly with the number of rounds. In particular, if $\mathbb{E}(\xi) = \{g\}$ for some $g \in G$ then the probability that $\mathbb{S}(R_0) \neq \{g\}$ and $\mathbb{S}(R_1) \neq \{sg\}$ decrease as $C/|R_0|$ and $C/|R_1|$ respectively. Therefore, in a case of a singleton mean-set $\mathbb{E}(\xi)$ a general strategy for Eve is clearly based on the above analysis. She computes the sets $A = \mathbb{S}(R_0)$, $B = \mathbb{S}(R_1)$, and then the set $B \cdot A^{-1}$. As the number of rounds n increases the probability that $B \cdot A^{-1} \neq \{s\}$ decreases as $C/|R_0| + C/|R_1|$. This is the idea of the *Mean-set attack principe* that we formulate below as a theorem.

Theorem 7.2 (Mean-set attack principle)**.** *Let G be a group, X a finite generating set for G, $s \in G$ a secret fixed element, and $\xi_1, \xi_2, \ldots$ a sequence of randomly generated i.i.d. group elements, $\xi_1 : \Omega \to G$, such that $\mathbb{E}\xi_1 = \{g_0\}$. If $\xi_1, \ldots, \xi_n$ is a sample of random elements of G generated by the Prover, $c_1, \ldots, c_n$ a succession of random bits (challenges) generated by the Verifier, and*

$$y_i = \begin{cases} r_i & \text{if } c_i = 0; \\ sr_i & \text{if } c_i = 1 \end{cases}$$

random elements representing responses of the Prover, then there exists a constant D such that

$$\mathbf{P}\left(s \notin \mathbb{S}\Big(\{y_i \mid c_i = 1, i = 1, \ldots, n\}\Big) \cdot \mathbb{S}\Big(\{y_i \mid c_i = 0, i = 1, \ldots, n\}\Big)^{-1}\right) \leq \frac{D}{n}.$$

Proof. It follows from Theorem 4.2 that there exists a constant C such that

$$\mathbf{P}(\mathbb{S}(\{\xi_i \mid c_i = 0, i = 1, \ldots, n\}) \neq \{g\}) \leq \frac{C}{\Big|\{i \mid c_i = 0, i = 1, \ldots, n\}\Big|}. \tag{7.1}$$

Now, our challenges c_i are just Bernoulli random variables with parameter $\frac{1}{2}$, i.e., for each $i = 1, \ldots, n$, we have

$$c_i = \begin{cases} 1, & \text{with probability } p = 1/2; \\ 0, & \text{with probability } p = 1/2 \end{cases}$$

with $\mathbb{E}(c_i) = \frac{1}{2}$ and $\sigma^2_{c_i} = \frac{1}{4}$. Then, using Chebyshev's inequality (4.2), we obtain

$$\mathbf{P}\Big(\Big|\{i \mid c_i = 0, i = 1, \ldots, n\}\Big| < \frac{n}{4}\Big) < \frac{4}{n}. \tag{7.2}$$

Indeed, if number of zero's in our sample of challenges is less than $\frac{n}{4}$, then the number of one's is greater or equal to $\frac{3n}{4}$, and we have

$$\mathbf{P}\Big(\Big|\{i \mid c_i = 0, i = 1, \ldots, n\}\Big| < \frac{n}{4}\Big) < \mathbf{P}\left(\left|\sum_{i=1}^{n} c_i - \frac{n}{2}\right| \geq \frac{n}{4}\right)$$

from the inclusion of the corresponding events. Note that

$$\left|\sum_{i=1}^{n} c_i - \frac{n}{2}\right| \geq \frac{n}{4} \Leftrightarrow \left|\frac{\sum_{i=1}^{n} c_i}{n} - \frac{1}{2}\right| \geq \frac{1}{4}$$

and

$$\mathbf{P}\left(\left|\frac{\sum_{i=1}^{n} c_i}{n} - \frac{1}{2}\right| \geq \frac{1}{4}\right) \leq \frac{4}{n}$$

from (4.2) for $\varepsilon = 1/4$. Thus, we have (7.2), as claimed.

Denoting $A = \Big\{\Big|\{i \mid c_i = 0, i = 1, \ldots, n\}\Big| < \frac{n}{4}\Big\}$, we get

$$\begin{aligned} \mathbf{P}\Big(\mathbb{S}(\{\xi_i \mid c_i = 0\}) \neq \{g\}\Big) = \mathbf{P}\Big(\Big\{\mathbb{S}(\{\xi_i \mid c_i = 0\}) \neq \{g\}\Big\} \bigcap A\Big) + \\ +\mathbf{P}\Big(\Big\{\mathbb{S}(\{\xi_i \mid c_i = 0\}) \neq \{g\}\Big\} \bigcap A^c\Big) \leq \\ \leq \mathbf{P}(A) + \mathbf{P}\Big(\Big\{\mathbb{S}(\{\xi_i \mid c_i = 0\}) \neq \{g\}\Big\} \mid A^c\Big)\mathbf{P}(A^c). \\ \leq \frac{4}{n} + \frac{4C}{n} \leq \frac{4+4C}{n}, \end{aligned}$$

where, on the event A^c, we estimate the right-hand side of (7.1) by $\frac{4C}{n}$.

Similarly we prove that $\mathbf{P}\Big(\mathbb{S}(\{s\xi_i \mid c_i = 1, i = 1, \ldots, n\}) \neq \{sg\}\Big) \leq \frac{4+4C}{n}$. Hence

$$\mathbf{P}\Big(s \notin \mathbb{S}(\{s\xi_i \mid c_i = 1, i = 1, \ldots, n\}) \cdot \mathbb{S}(\{\xi_i \mid c_i = 0, i = 1, \ldots, n\})^{-1}\Big) \leq \frac{8+8C}{n}.$$

□

We compute $\mathbb{S}_n$ using the Algorithm 7.1 in the following heuristic attack.

Algorithm 7.3. (The attack)

INPUT: The Prover's public element (t,w). Sequences R_0 and R_1 as in the protocol.

OUTPUT: The secret element $s \in B_n$ used in generation of R_1, or $Failure$.

COMPUTATIONS:

A. Apply Algorithm 7.1 to R_0 and obtain g_0.

B. Apply Algorithm 7.1 to R_1 and obtain g_1.

C. Check if $z = g_1 g_0^{-1}$ satisfies $t = z^{-1}wz$ and if so output z. Otherwise output $Failure$.

Observe that in the last step of the algorithm we perform the check for the secret element s (see the description of the protocol in Section 7.3.1).

7.6 Experimental results

7.6.1 Platform group and parameters

The scheme was suggested to be used with the braid group B_n which has the following (Artin's) presentation

$$B_n = \left\langle \sigma_1, \ldots, \sigma_{n-1} \;\middle|\; \begin{array}{ll} \sigma_i\sigma_j\sigma_i = \sigma_j\sigma_i\sigma_j & \text{if } |i-j| = 1 \\ \sigma_i\sigma_j = \sigma_j\sigma_i & \text{if } |i-j| > 1 \end{array} \right\rangle,$$

with some additional requirements.

The length function relative to the Artin generators $\{\sigma_1, \ldots, \sigma_{n-1}\}$ is NP-hard (see (11)). In this work we use the method to approximate geodesic length proposed in (27). Even though it does not guarantee the optimal result, it proved to be practically useful in a series of successful attacks, see (18; 15; 29; 30). We want to point out that we compute the sample weight values in Algorithm 7.1 (which a subroutine in Algorithm 7.3) using the approximated distance function values in B_n.

7.6.2 Experiments

To demonstrate the practical use of our attack we performed a system of experiments. In each experiment we randomly generated an instance of the authentication protocol and tried to break it, i.e., find the private key, using the technique developed in this paper. Recall that each authentication is a series of k 3-pass commitment-challenge-response rounds. Therefore, an instance of authentication is k triples (x_i, c_i, y_i) obtained as described in Section 7.3.1.

A random bit c_i is chosen randomly and uniformly from the set $\{0, 1\}$. In our experiments we make an assumption that exactly half of c_i's are 0 and half are 1. This allows us to have a pair of equinumerous sets $R_0 = \{r_1, \ldots, r_{k/2}\} \subset B_n$ and $R_1 = \{sr'_1, \ldots, sr'_{k/2}\} \subset B_n$ corresponding to an instance of the protocol.

The main parameters for the system are the rank n of the braid group, the number of rounds in the protocol k, and the length of secret keys L. We generate a single instance of the problem with parameters (n, k, L) as follows:

- A braid s is chosen randomly and uniformly as a word of length L over a group alphabet $\{\sigma_1, \ldots, \sigma_{n-1}\}$. This braid is a secret element which is used only to generate further data and to compare the final element to.

- A sequence $R_0 = \{r_1, \ldots, r_{k/2}\}$ of braid words chosen randomly and uniformly as words of length L over a group alphabet $\{\sigma_1, \ldots, \sigma_{n-1}\}$.

- A sequence $R_1 = \{sr'_1, \ldots, sr'_{k/2}\}$ of braid words, where r'_i are chosen randomly and uniformly as words of length L over a group alphabet $\{\sigma_1, \ldots, \sigma_{n-1}\}$.

For every parameter set (n, k, L) we generate 1000 random instances of the protocol and run Algorithm 7.3 which attempts to find the secret key s used in the generation of R_1.

Below we present the results of actual experiments done for groups B_5, B_{10}, and B_{20}. Horizontally we have increasing sample sizes k (number of rounds) from 10 to

320 and vertically we have increasing length L from 10 to 100. Each cell in the tables below contains a success rate of finding secret s.

L\k	**10**	**20**	**40**	**80**	**160**	**320**
10	13%	54%	93%	99%	100%	100%
50	0%	0%	2%	21%	65%	83%
100	0%	0%	0%	0%	0%	4%

Table 7.1: Experiments in B_5.

L\k	**10**	**20**	**40**	**80**	**160**	**320**
10	12%	71%	97%	100%	100%	100%
50	0%	8%	53%	92%	99%	99%
100	0%	0%	0%	0%	17%	48%

Table 7.2: Experiments in B_{10}.

L\k	**10**	**20**	**40**	**80**	**160**	**320**
10	17%	85%	100%	100%	100%	100%
50	0%	16%	78%	99%	99%	100%
100	0%	1%	38%	87%	98%	99%

Table 7.3: Experiments in B_{20}.

We immediately observe from the data above that:

- the success rate increases as sample size increases;
- the success rate decreases as the length of the key increases;
- the success rate increases as the rank of the group increases.

The first observation is the most interesting since the number of rounds is one of the main parameters in the protocol (see Section 7.3.1). According to Sibert, one has to increase the number of rounds for higher reliability of the protocol. But, in fact, we observe previously unforeseen (by the authors of the protocol) behavior – security of the scheme decreases as k increases. The second observation can be interpreted as follows – the longer the braids the more difficult it is to compute the approximation of their lengths. So, here we face the problem of computing the length function. The third observation is easy to explain. The bigger the index of B_n the more braid generators commute and the simpler random braids are.

7.7 Defending against the Mean-Set Attack

In this section we describe some principles on how to defend against the mean-set attack presented above or make it computationally infeasible. Defending can be done through a special choice of the platform group G or a special choice of a distribution μ on G.

7.7.1 Large mean-set

To foil the attack one can use a distribution μ on G such that the set $\mathbb{E}(\mu)$ is huge. Recall that Algorithm 7.3 can find up to one element of G minimizing the weight function. For that it uses Algorithm 7.1 which randomly (according to some measure ν) chooses an element $g \in G$ and then gradually changes it so that to minimize its M value. This way the distribution ν on the initial choices $g \in G$ defines a distribution ν^*_μ on the set of local minima of M on G. More precisely, for $g' \in G$, $\nu^*_\mu(g') = \mu\{g \in G \mid \text{Algorithm 7.1 stops with the answer } g' \text{ on input } g\}$.

Now, denote by μ_s the *shifted probability measure* on G by an element s defined by $\mu_s(g) = \mu(s^{-1}g)$. If $S \subseteq G$ is the set of local minima of the weight function M relative to μ then the set sS is the set of local minima relative to μ_s. But the distribution $\nu^*_{\mu_s}$ does not have to be induced from ν^*_μ by the shift s, i.e., the equality $\nu^*_{\mu_s}(g) = \nu^*_\mu(s^{-1}g)$ does not have to hold. In fact, the distributions ν^*_μ and $\nu^*_{\mu_s}$ can "favor" unrelated subsets of S and sS respectively. That would definitely foil the attack presented in this paper. On the other hand if ν_μ^* and $\nu_{\mu_s}^*$ are related then our attack can work.

Finally we would like to point out again that probability measures on groups were not extensively studied and there are no good probability measures on general groups and no general methods to construct measures satisfying the desired properties ((7)). In addition, the problem of making distributions with large mean sets is very complicated because not every subset of a group G can be realized as a mean set (see Section 5.1).

7.7.2 Undefined mean-set

Another way to foil the attack is to define a distribution μ on G so that $\mathbb{E}(\mu)$ is not defined, i.e., the weight function $M(\cdot) \equiv M^{(2)}(\cdot)$ is not finite. In that case we do not have theoretical means for the attack. The sample weights tend to ∞ with probability 1. Nevertheless, we still can compare the sample weight values using the theory of central order presented in Chapter 6, where we show that the assumption $M^{(2)} < \infty$ can be relaxed to $M^{(1)} < \infty$. If $M^{(1)}$ is not defined then the lengths of commitments are too large and are impractical.

7.7.3 Groups with no efficiently computable length functions

One of the main tools in our technique is an efficiently computable function d_X on G. To prevent the intruder from employing our attack, one can use a platform group G with a hardly computable (or approximated) length function d_X relative to any “reasonable” finite generating set X. By ”reasonable“ generating set we mean a set small relative to the main security parameters. Examples of such groups exist. On the other hand, it is hard to work with such groups in practice.

Chapter 8

Further ideas. Problems and questions.

In this chapter we collect some ideas for further developments of our theory. In addition, we state several problems that appear to be interesting from our point of view.

8.1 Centers and group geometry

8.1.1 Expectation and boundary of the graph

It would be very interesting, especially for applications in geometric group theory, to intertwine the geometry of a graph and our probability theory on it. It is already proved in in Chapter 5 that the size of the center set on a free group F_n is a set containing up to 2 elements. The next goal is to look into further connections.

Recall that we defined the generalized mean-set $\mathbb{E}$ relative to central order in Chapter 6. Next step is to extend the definition of the generalized mean-set $\mathbb{E}$ to the boundary elements of the graph. When we are on the real line $\mathbf{R}$, the boundary consists of only two points, namely, $\pm\infty$, and, perhaps, is not of any particular interest. Nevertheless, for instance, we may say that $\mathbb{E}\mu = \pm\infty$ when $\sum_{x\in\mathbf{Z}} x\mu(x) = \pm\infty$, thinking of extended image (phase) space $\mathbf{R} = [-\infty, +\infty]$. Moreover, the generalized law of large numbers holds for certain sequences of random variables without expectation (see (13, pages 235–236)). These considerations, together with understanding that the geometry of the graph is more complicated than the geometry of the real line, would serve as a good motivation to embark upon considering mean-sets $\mathbb{E}$ from the geometric point of view and extend them to the boundary of the graph in future developments of our theory.

Finite graphs are not interesting for us because for such graphs, for any c, the set $\mathbb{E}^{(c)}$ is always finite. Consider an infinite graph. There are three principal cases that we distinguish:

- The set $\mathbb{E}^{(c)}$ is finite. This corresponds to the case when classical expectation is defined.
- The set $\mathbb{E}^{(c)}$ is empty. This corresponds to the case when classical expectation is not defined.
- The set $\mathbb{E}^{(c)}$ is infinite. This corresponds to the case when classical expectation is not defined.

In example 6.31, we have already considered a special case for a generalized $\mathbb{E} \equiv \mathbb{E}^{(2)}$ – the case when the set $\mathbb{E}^{(2)}$ is empty. In that example, intuitively, it is natural to say that center points belong to a *boundary* of the graph Γ.

There are several different ways to define a boundary of a graph. One of them uses the so-called end-compactification. The reader can consult (41) for the definition of this type of boundary.

Meanwhile, we can only say that the question of extension of the notion of the expectation of random graph/group element to the boundary requires further research.

8.1.2 Some small problems about mean-sets and group geometry

Problem 8.1. *Let $\varphi : G \to H$ be a group homomorphism. Is it true that if the measure on H is induced by the measure on G then $\mathbb{E}(\varphi(\xi)) = \varphi(\mathbb{E}(\xi))$.*

Proof. This property is not true in general. Here is a counterexample.

Consider free abelian groups $\mathbf{Z}$ and $\mathbf{Z}^2$ of rank 1 and 2. Let $\varphi : \mathbf{Z}^2 \to \mathbf{Z}$ be a group epimorphism which maps a pair (a, b) to the first coordinate a. Assume that μ is a distribution on $\mathbf{Z}^2$ such that $\mu(0, 100) = \mu(0, -100) = \mu(10, 0) = 1/3$. The center for such distribution is $(0, 0)$. Now, the epimorphism φ induces a probability distribution μ' on $\mathbf{Z}$ such that $\mu'(0) = 2/3$ and $\mu'(10) = 1/3$. The corresponding center is 3. Hence, the property $\mathbb{E}(\varphi(\xi)) = \varphi(\mathbb{E}(\xi))$ does not hold and the reason is that group homomorphisms can dramatically change distances between the points. □

Problem 8.2. *Let ξ_1 be a random element in a graph Γ_1 and ξ_2 a random element in a graph Γ_2. Suppose that $\mathbb{E}\xi_1$ and $\mathbb{E}\xi_2$ are defined. Is it true that*

$$\mathbb{E}(\xi_1 \times \xi_2) = \mathbb{E}\xi_1 \times \mathbb{E}\xi_2?$$

Proof. The answer is "no“. Here is the example. Consider the probability measure μ on $\mathbf{Z}$ such that $\mu(0) = 2/3$ and $\mu(15) = 1/3$. Then $\mathbb{E}\mu = 5$ and $\mathbb{E}\mu \times \mathbb{E}\mu = (5, 5)$.

On the other hand $\mu \times \mu$ is a probability measure on $\mathbf{Z} \times \mathbf{Z}$ defined by $\mu(0, 0) = 4/9$, $\mu(15, 0) = 2/9$, $\mu(0, 15) = 2/9$, $\mu(15, 15) = 1/9$. To compute a central point $v = (x, y)$ we solve the following optimization problem:

$$M(v) = (x + y)^2 \mu(0, 0) + ((15 - x) + y)^2 \mu(15, 0) +$$

$$+ (x + (15 - y))^2 \mu(0, 15) + ((15 - x) + (15 - y))^2 \mu(15, 15) \to \min .$$

which equals to

$$x^2 + \frac{2}{9}xy + y^2 - \frac{20}{3}x - \frac{20}{3}y + 200 \to \min .$$

The solution is $(3,3)$. Hence

$$\mathbb{E}(\mu \times \mu) = (3,3) \neq (5,5) = \mathbb{E}\mu \times \mathbb{E}\mu.$$

□

Below, we state several problems without solutions.

Problem 8.3. *Changing the generating set for a group changes the distance function. How does a change of the generating set affect centers for a given distribution?*

Problem 8.4. *Let $\varphi : G \to H$ be a quasi-isometry between 2 groups. Let μ be a distribution on G and μ' be a distribution on H induced by φ. Is there any relation between $\varphi(\mathbb{E}\mu)$ and $\mathbb{E}\mu'$?*

(Probably not much, but it would be interesting to find actual examples when something strange happens).

Problem 8.5. *Assume that $G \leq H$ and μ is a probability measure on G. What is the relation between $\mathbb{E}\mu$ in H and in G.*

Remark 8.6. It would be interesting to continue this line of research, trying to find connections of our centers and group geometry, and see how the geometry of the Cayley graph of a group affects the structure of a mean-set.

8.2 Other goals.

One of the future goals of this research is formulating and proving an analogue of the central limit theorem (another fundamental result of probability theory) for distribution on graphs and groups, as well as generalization of other classical results.

Another endeavor would be to attempt to generalize the theory of mean-sets to general metric spaces.

8.3 Conclusion

We stop at this point. We proved a generalization of the Strong Law of Large Numbers for graphs and groups and discussed its possible applications. We enhanced our novel theory with other theoretical tools, such as an analogue of Chebyshev inequality for graphs and the notion of central order on graphs. We discussed technical difficulties of computations and some practical ways of dealing with this issue. Moreover, we provided results of actual experiments supporting many of our conclusions. At the end, we indicated possible directions of the future research and developments and stated some problems.

What really important is that this work helps us to see that generalization of probabilistic results to combinatorial objects can lead to unanticipated applications in practice and to interconnections with other areas of mathematics. Indeed, Joseph Fourier was right when he said that mathematics compares the most diverse phenomena and discovers the secret analogies that unite them.

Bibliography

[1] I. Anshel, M. Anshel, and D. Goldfeld, *An algebraic method for public-key cryptography*, Math. Res. Lett. 6 (1999), pp. 287–291.

[2] R. Bellman, *Limit theorems for non-commutative operations*, Duke Math. J. 21 (1954), pp. 491–500.

[3] P. Billingsley, *Probability and Measure*. Wiley-Interscience, 1995.

[4] J. S. Birman, V. Gebhardt, and J. Gonzalez-Meneses, *Conjugacy in Garside groups I: Cyclings, powers, and rigidity*, Groups, Geometry, and Dynamics 1 (2007), pp. 221–279.

[5] ———, *Conjugacy in Garside Groups III: Periodic braids*, J. Algebra 316 (2007), pp. 746–776.

[6] ———, *Conjugacy in Garside groups II: Structure of the ultra summit set*, Groups, Geometry, and Dynamics 2 (2008), pp. 13–61.

[7] A. Borovik, A. Myasnikov, and V. Shpilrain, *Measuring sets in infinite groups*. Computational and Statistical Group Theory, Contemporary Mathematics 298, pp. 21–42. American Mathematical Society, 2002.

[8] J. C. Cha, K. H. Ko, S. J. Lee, J. W. Han, and J. H. Cheon, *An Efficient Implementation of Braid Groups*. Advances in Cryptology – ASIACRYPT 2001, Lecture Notes in Computer Science 2248, pp. 144–156. Springer, Berlin, 2001.

[9] CRyptography And Groups (CRAG) C++ Library, Available at `http://www.acc.stevens.edu/downloads.php`.

[10] P. Dehornoy, *Braid-based cryptography*. Group theory, statistics, and cryptography, Contemporary Mathematics 360, pp. 5–33. American Mathematical Society, 2004.

[11] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, and W. P. Thurston, *Word processing in groups*. Jones and Bartlett Publishers, 1992.

[12] U. Feige, A. Fiat, and A. Shamir, *Zero knowledge proofs of identity*, STOC '87: Proceedings of the nineteenth annual ACM Conference on Theory of Computing (1987), pp. 210–217.

[13] W. Feller, *An Introduction to Probability Theory and Its Applications: Volume 2*. John Wiley & Sons, New York, 1971.

[14] F. Folland, *Real Analysis: Modern Techniques and Their Applications*. Wiley-Interscience, 1999.

[15] D. Garber, S. Kaplan, M. Teicher, B. Tsaban, and U. Vishne, *Length-based conjugacy search in the Braid group*. Algebraic Methods in Cryptography, Contemporary Mathematics 418, pp. 75–88. American Mathematical Society, 2006.

[16] O. Goldreich, *Foundations of Cryptography*. Cambridge University Press, 2001.

[17] ______, *Zero-Knowledge twenty years after its invention*, preprint, available at `http://citeseer.ist.psu.edu/556429.html`, 2002.

[18] J. Hughes and A. Tannenbaum, *Length-based attacks for certain group based encryption rewriting systems*, preprint. Available at `http://front.math.ucdavis.edu/0306.6032`.

[19] K. Itô and H. P. McKean, Jr., *Potentials and the random walk*, Illinois J. Math. 4 (1960), pp. 119–132.

[20] L. Kantorovich and G. Akilov, *Functional Analysis*. Elsevier, 1982.

[21] L. Kantorovich and G. Rubinshtein, *On a certain function space and some extremal problems*, Dokl. Akad. Nauk SSSR 115 (1957), pp. 1058–1061.

[22] ______, *On some space of countably additive functions*, Univ. Mat. Mekh. Astronom. 7 (1958), pp. 52–59.

[23] A. Karlsson and F. Ledrappier, *On Laws of Large Numbers for Random Walks*, Ann. Probab. 34 (2006), pp. 1693–1706.

[24] H. Kesten, *Symmetric random walks on groups*, T. Am. Math. Soc. 92 (1959), pp. 336–354.

[25] K. H. Ko, S. J. Lee, J. H. Cheon, J. W. Han, J. Kang, and C. Park, *New public-key cryptosystem using braid groups*. Advances in Cryptology – CRYPTO 2000, Lecture Notes in Computer Science 1880, pp. 166–183. Springer, Berlin, 2000.

[26] A. Kurosh, *Theory of Groups*. Chelsea Publishing Corp., 1979.

[27] A. G. Miasnikov, V. Shpilrain, and A. Ushakov, *Random Subgroups of Braid Groups: An Approach to Cryptanalysis of a Braid Group Based Cryptographic Protocol*. Advances in Cryptology – PKC 2006, Lecture Notes in Computer Science 3958, pp. 302•–314. Springer, Berlin, 2006.

[28] ———, *Group-based Cryptography*, Advanced Courses in Mathematics - CRM Barcelona. Birkhäuser Basel, 2008.

[29] A. D. Myasnikov and A. Ushakov, *Cryptanalysis of Anshel-Anshel-Goldfeld-Lemieux key agreement protocol*, preprint. Available at `http://arxiv.org/abs/0801.4786v1`.

[30] ———, *Length Based Attack and Braid Groups: Cryptanalysis of Anshel-Anshel-Goldfeld Key Exchange Protocol*. Advances in Cryptology – PKC 2007, Lecture Notes in Computer Science 4450, pp. 76–88. Springer, Berlin, 2007.

[31] M. Paterson and A. Razborov, *The set of minimal braids is co-NP-complete*, J. Algorithms 12 (1991), pp. 393•–408.

[32] J. J. Quisquater, L. C. Guillow, and T. A. Berson, *How to explain zero-knowledge protocols to your children*. Advances in Cryptology – CRYPTO 1989, Lecture Notes in Computer Science 435, pp. 628–631. Springer, Berlin, 1990.

[33] G. Rubinshtein, *On multiple-point centers of normalized measures on locally compact metric spaces*, Siberian Math. J. 36 (1995), pp. 143–146.

[34] H. Sibert, P. Dehornoy, and M. Girault, *Entity authentication schemes using braid word reduction*, Discrete Appl. Math. 154 (2006), pp. 420–436.

[35] A. V. Skorohod, *Basic Principles and Applications of Probability Theory*. Springer, 2004.

[36] F. Spitzer, *Principles of Random Walk*. Springer, 2001.

[37] K. Uchiyama, *Wiener's test for random walks with mean zero and finite variance*, Ann. Prob. 26 (1998), pp. 368–376.

[38] A. M. Vershik, *L. V. Kantorovich and linear programming*. Leonid Vital'evich Kantorovich: a man and a scientist, pp. 130–152. SO RAN, Novosibirsk, 2002.

[39] ______, *Kantorovich metric: initial history and little-known applications*, 133, pp. 1410–1417, 2006.

[40] D. West, *Introduction to Graph Theory (2nd edition)*. Prentice Hall, 2000.

[41] W. Woess, *Random Walks on Infinite Graphs and Groups*, Cambridge Tracts in Mathematics. Cambridge University Press, 2000.