Spectra of Supersingular Isogeny Graphs

Buuren, S. van

Spectra of Supersingular Isogeny Graphs

Master Thesis of

Sam van Buuren

June 27, 2022

Thesis supervisor: dr. J. Vonk

Leiden University
Mathematical Institute

Abstract

In a 2007 paper, Charles, Lauter and Goren studied how one might use Ramanujan graphs to create cryptographic hash functions. One of the most well-known such graphs is the isogeny graph, whose vertices are indexed by the isomorphism classes of supersingular elliptic curves in a characteristic p.

In this thesis, we study the spectra of these graphs. To start, we give two algorithms to compute these graphs for small p, and present data obtained from implementing these algorithms. This data provides some statistical evidence for several properties of the spectra of these graphs. We discuss the proof of two of these properties.

Firstly, the fact that these supersingular isogeny graphs are Ramanujan, i.e. that they have large spectral gap. For this, we discuss the relation with Hecke operators and the Eichler–Shimura relation.

Secondly, the distribution of the eigenvalues as p tends to infinity. This we prove via the relation between the graph and the Brandt matrices for the quaternion algebra ramified at p and ∞. We sketch a proof of the Eichler–Selberg trace formula, and use this to conclude the proof.

Contents

1 Introduction 4

1.1 Motivation . . . 4

1.2 Overview of the thesis . . . 4

1.3 Graphs and hash functions . . . 5

2 Pizer graphs 7

2.1 General Ramanujan graphs . . . 7

2.2 Pizer graphs . . . 13

2.3 Implementation and data . . . 21

3 The Ramanujan property 27

3.1 Pizer graphs and Hecke operators . . . 27

3.2 The Riemann hypothesis for abelian varieties . . . 30

3.3 Eichler–Shimura and finishing the proof . . . 34

4 Distribution of eigenvalues 35

4.1 Measure theory of eigenvalues . . . 37

4.2 Our polynomials and measures . . . 39

4.3 Brandt matrices and their trace . . . 40

4.4 Finishing the proof . . . 45

5 Final thoughts 48

A Graph Theory 52

B SageMath Scripts 56

1 Introduction

1.1 Motivation

There has, in recent years, been much interest in isogeny-based cryptographic schemes, especially in the context of post-quantum security. A central role here is played by the supersingular isogeny graphs, whose vertices are (indexed by) the isomorphism classes of supersingular elliptic curves in characteristic p and whose edges are (indexed by) the cyclic degree-m isogenies between these curves (here p is prime and m is an integer coprime to p). For most of this thesis, we consider graphs of prime degree $m = l$.

These graphs appear also in the work of Charles–Lauter–Goren [CLG09], where they are used to construct a cryptographic hash function. We will discuss this construction below in Section 1.3. As we will see, the security of this hash function depends on the mixing properties of the used graphs, and its speed (i.e. how quickly we can compute a hash) is dependent on how efficiently steps in our graphs can be computed. We will see that the supersingular isogeny graphs have good mixing properties. Steps in these graphs are isogenies over finite fields, the computation of which is still optimized (for a quantum computing algorithm, one can for instance look at [BJS14]).

In this thesis we study we study the spectrum of these isogeny graphs, which we will call Pizer graphs (after Arnold Pizer, who was first to prove these graphs are Ramanujan). These spectra give us information on several graph-theoretic properties. For instance, in Theorem 2.8, we will see that having a large spectral gap implies that a random walk in our graph quickly converges to a uniform distribution. We will not, however, delve deeply in the cryptographic consequences of the properties of the spectra. We are interested first and foremost simply in understanding these graphs and the techniques one requires to study them.

1.2 Overview of the thesis

In Section 2, we discuss the Pizer graphs from a computational point of view. To start, in Section 2.1, we discuss some theory on the spectrum of general graphs, and define Ramanujan graphs. We prove in Theorem 2.8 that these Ramanujan graphs have excellent mixing properties. In Section 2.2, we then define our Pizer graphs and discuss two algorithms for computing these graphs. Included here is also a discussion on how one can compute a single supersingular j-invariant over the prime field $\mathbb{F}_p$, based on a paper by Bröker [Brö09]. Finally, Section 2.3 presents data on the spectra of Pizer graphs acquired via our own implementation of these algorithms. From this data, we find statistical backing for the known results that that Pizer graphs are Ramanujan graphs, and that their eigenvalues follow a distribution. The remainder of the thesis discusses the proof of these results.

The proof of the Ramanujan property is found in Section 3. We deduce this property from a theorem of Deligne on the eigenvalues of Hecke operators acting on modular forms on $\Gamma_0(p)$, from [Del74]. This result is on forms of general weight, but for our purposes, the case of weight two modular forms suffices. The proof of this specific case is a more classical consequence of the Riemann hypothesis for abelian varieties (which we discuss in Section 3.2) and the work of Eichler–Shimura, as we see in Section 3.3. We also discuss the relationship between Pizer graphs and modular forms in Section 3.1.

Section 4 discusses a proof of the distribution of the eigenvalues. This proof is directly based on Serre's proof of a distribution of the eigenvalues of Hecke operators [Ser97], and also bears a strong resemblance a classic graph-theoretic result of McKay [McK81]. Using that the distribution of eigenvalues of linear operators is determined by traces of polynomials of these operators, as discussed in Section 4.1, we compute the Eichler–Selberg trace formula for Brandt matrices in Section 4.3 and use this to deduce the distribution in Section 4.4.

Finally, in section 5, we give a more informal review of the various results observed in this thesis.

1.3 Graphs and hash functions

A *hash function* is a way to compress an arbitrary-length message to a fixed-length bit-string in such a way that one cannot deduce the original from the result. Mathematically, we can model this as a function

$$h : \mathbb{Z}_{\geq 0} \to \{0,1\}^n,$$

for some fixed $n \geq 0$, such that

- Given $s \in \{0,1\}^n$, it is hard to find $m \in \mathbb{Z}_{\geq 0}$ such that $h(m) = s$,
- Given $m_1 \in \mathbb{Z}_{\geq 0}$, it is hard to find $m_2 \in \mathbb{Z}_{\geq 0}$ such that $h(m_1) = h(m_2)$ but $m_1 \neq m_2$,
- It is hard to find any pair $m_1, m_2 \in \mathbb{Z}_{\geq 0}$ such that $h(m_1) = h(m_2)$ but $m_1 \neq m_2$.

Example 1.1. An ideal example would be the following: for $k \in \mathbb{Z}_{\geq 0}$, let X_k be a uniform random variable on $\{0,1\}^n$, such that X_{k_1} and X_{k_2} are independent for $k_1 \neq k_2$. We define

$$h : \mathbb{Z}_{\geq 0} \to \{0,1\}^n, h(m) = X_m.$$

The example above is impractical, as computing infinitely many X_k is impossible. Still, we can attempt to simulate such randomness. This is what the hash function discussed below is based on. As mentioned, this discussion is taken from [CLG09].

Let $G = (V, E)$ be a $(k+1)$-regular graph (that is, each vertex $v \in V$ has exactly $(k+1)$ edges that originate in v). For any vertex $v \in V$, let $E(v)$ be the set of edges coming from v. Define some ordering on each $E(v)$, so that we can speak of the first (second, etc) edge coming from v for all vertices $v \in V$.

We can uniquely write a message $m \in \mathbb{Z}_{\geq 0}$ in k-ary digits, that is, we can write:

$$m = \sum_{i=0}^{\lfloor \log_k(m) \rfloor} a_i \cdot k^i,$$

for unique $a_i \in \{0, \dots, k-1\}$. We call the a_i the *k-ary digits* of m.

Algorithm 1.2. Input: A message $m \in \mathbb{Z}_{\geq 0}$, a graph $G = (V, E)$ as above and some starting vertex $v_0 \in V$.
Output: The hash of m with respect to G.

1. Compute the k-ary digits $a_0, \dots, a_r$ of m (here $r = \lfloor \log_k(m) \rfloor$).
2. Let e_0 be the $(k+1)$-th edge coming from v_0
3. Initialise $v_s = v_0$, $e_s = e_0$ and $i = 0$.
4. While $i \leq r$:
   (a) Let e_t be the a_i-th edge coming from v_s. If $e_t = e_s$, let it be the $(a_i + 1)$-th edge instead (computing modulo $k+1$).
   (b) Let v_t be the other edge coming from e_t, i.e. $e_t = (v_s, v_t)$.
   (c) Update $i = i + 1$, $e_s = e_t$, $v_s = v_t$
5. Return v_t.

In words, we make a ‘random’ walk through the graph based on our input message m and remember only the ending vertex of this walk. If the graph G over which we walk has the property that random walks quickly become indistinguishable from the uniform distribution over V, this hash function comes close to the ideal distribution outlined in the example above.

2 Pizer graphs

In this chapter, we will define the main objects of interest to us, the Pizer graphs. We start by defining Ramanujan graphs in generality, and motivating why these are objects worthy of study. Then we define the Pizer graphs themselves and give two algorithms we can use to compute then in their entirety. Finally, we present some data, acquired with said algorithms, in which we observe statistically a number of results whose mathematical explanation is the subject of the remaining chapters.

2.1 General Ramanujan graphs

We introduce expander graphs and Ramanujan graphs, and prove a mixing lemma on these graphs. We presume the reader already has some knowledge of graph theory For a quick recap of basic graph theory, see either appendix A of this thesis or chapter one of *Elementary Number Theory, Group Theory and Ramanujan Graphs* by Guillana Davidoff, Peter Sarnak and Alain Valette, [DSV03]. Most of the theory in this section can also be found in this source. Unless otherwise stated, our graphs will be finite and connected.

We will start with an extremely important definition.

Definition 2.1 (Isoperimetric constant)**.** The *isoperimetric constant* of a graph $G = (V, E)$ is

$$h(G) = \min_{S \subset V \mid |S| < |V|/2} \frac{|\delta S|}{|S|}$$

where

$$\delta(S) = \{(v, w) \in E \mid v \in S, w \notin S \text{ or } w \in S, v \notin S\}$$

is the *boundary* of S.

Definition 2.2 (Expander graph)**.** A family $G_1, G_2, \dots$ of k-regular undirected graphs, $G_m = (V_m, E_m)$, such that $|V_m| \to \infty$ as $m \to \infty$ is called a family of (combinatorial) ϵ-*expanders* if for each m,

$$h(G_m) \geq \epsilon$$

Remark. Before we get to discussing some properties of these graphs, perhaps it is worth taking a moment to discuss why one should think these objects are interesting. The isoperimetric constant is a measure of how well a graph mixes; we expect the boundary of a set of vertices to grow quite slowly with the size of the set. As such, we expect $h(G)$ to be very low and close to 0 if $|V|$ is large.

Expander graphs have very good mixing properties; the size of the boundary grows linearly with the size of the set. This means that if we start in some subset S of the vertices, S connects to many of vertices not in S. Random walks over expander graphs very quickly approach the uniform distribution, that is to say, the chance that a random walk in $n = O(\log(|V|)$ steps ends in a vertex v is approximately $\frac{1}{|V|}$ for *all* $v \in V$. Whilst getting a uniform sample from V can become hard to calculate as $|V|$ grows very large, if we can calculate edges easily, taking n random steps in a k-regular graph might be easy.

There is a more exact result, but before we get to that, we must first relate the isoperimetric constant to the eigenvalues of the matrix.

Recall that a k-regular graph G always has eigenvalue k, and in general all eigenvalues λ satisfy $|\lambda| \leq k$. With the non-trivial eigenvalues we mean those eigenvalues not equal to k (recall that the multiplicity of the eigenvalue k is the number of connected components, by e.g. Proposition A.5. Since our graphs are connected, k has multiplicity 1). We then get the following lemma, which is theorem 1.2.3 in [DSV03].

Lemma 2.3. *Let $G = (V, E)$ be a finite, connected, k-regular graph without loops. Let A be the adjacency matrix of G and μ_1 the largest non-trivial eigenvalue of A. Then*

$$\frac{k - \mu_1}{2} \leq h(G) \leq \sqrt{2k(k - \mu_1)}$$

Proof. For this proof, we define for any finite set X the space $l^2(X) = \mathrm{Maps}(X, \mathbb{C})$. Overly formally, one might think of this as the L^2 space on X.

We start with the first inequality. Choose some orientation of the edges of G, so that every edge $e \in E$ has an endpoint e^+ and origin e^- both in V and define the *boundary operator*

$$d : l^2(V) \to l^2(E), df(e) = f(e^+) - f(e^-).$$

We endow $l^2(V)$ with the Hermitian scalar product

$$\langle f, g \rangle = \sum_{x \in V} \overline{f(x)} g(x)$$

and $l^2(E)$ with the analogous product, so that we may consider the adjoint map $d^* : l^2(E) \to l^2(V)$, i.e. the map such that $\langle df, g \rangle = \langle f, d^* g \rangle$ for $f \in l^2(V), g \in l^2(E)$. Define a function $\delta : V \times E \to \{0, 1, -1\}$ given by

$$\delta(v, e) = \begin{cases} 1 & x = e^+ \\ -1 & x = e^- \\ 0 & \text{else} \end{cases}$$

We have

$$df(e) = \sum_{x \in V} \delta(x, e) f(x)$$

for $f \in l^2(V), e \in E$, by definition. It easily follows that, for $v \in V$ and $g \in l^2(E)$,

$$d^* g(v) = \sum_{e \in E} \delta(v, e) g(e).$$

We define the *laplacian* $\Delta = d^* \cdot d : l^2(V) \to l^2(V)$. Note that

$$\begin{aligned} \Delta f(v) &= (d^* \sum_{x \in V} \delta(x, .) f(x))(v) \\ &= \sum_{x \in V} \sum_{e \in E} \delta(x, e) \delta(v, e) f(x) \\ &= \sum_{e \in E} \delta(v, e)^2 f(x) + \sum_{x \in V, x \neq v} \sum_{e \in E} \delta(x, e) \delta(v, e) f(x) \\ &= k f(v) - \sum_{x \in V} A_{xv} f(x). \end{aligned}$$

Hence $\Delta = kI - A$, where I is the $|V|$ by $|V|$ identity matrix and A acts in the natural way on $l^2(V)$. Note that in the last equality, we use that A has no loops.

Since A is symmetric, it is diagonalizable and Δ takes the form

$$\Delta = \begin{pmatrix} 0 & & & \emptyset \\ & k-\mu_1 & & \\ & & \ddots & \\ \emptyset & & & k-\mu_{n-1} \end{pmatrix}.$$

Here $\mu_1 \geq \ldots \geq \mu_{n-1}$ are the eigenvalues of A not equal to k. If $f \in l^2(V)$ is such that $\sum_{x\in V} f(x) = 0$, i.e. if f is orthogonal to the constant functions (the eigenfunctions for Δ with eigenvalue 0), then

$$\|df\|_2^2 = \langle df, df\rangle = \langle \Delta f, f\rangle \geq (k-\mu_1)\,\|f\|_2^2\,.$$

Consider the function on V defined as follows. Fix $F \subset V$ and define $F^c = V\backslash F$. Set

$$f(x) = \begin{cases} |F^c| & x \in F \\ -|F| & x \notin F \end{cases}$$

Clearly, f has the following properties:

$$\sum_{x\in V} f(x) = |F^c||F| - F||F^c| = 0$$

$$\|f\|_2^2 = |F|\cdot|F^c|^2 + |F^c||F|^2 = |F|\cdot|F^c|\cdot|V|$$

$$df(e) = \begin{cases} \pm|V| & e \text{ connects a vertex in } F \text{ with a vertex in } F^c \\ 0 & \text{else} \end{cases}$$

As such, $\|df\|_2^2 = |V|^2|\delta F|$ and by the inequality above,

$$\|df\|_2^2 = |V|^2|\delta F| \geq (k-\mu_1)|F||F^c||V|,$$

yielding

$$\frac{|\delta F|}{|F|} \geq (k-\mu_1)\frac{|F^c|}{|V|}.$$

for any $F \subset V$. If $|F| \leq \frac{|V|}{2}$, we get the desired

$$\frac{\delta F}{|F|} \geq \frac{k-\mu_1}{2},$$

which implies (since the above holds for any $F \subset G$ with $|F| \leq \frac{|V|}{2}$)

$$h(G) \geq \frac{k-\mu_1}{2}.$$

We now move on to the second inequality. This is a more involved proof. We require some results on another, more mysterious operator. For any non-negative $f : V \to \mathbb{R}$, define

$$B_f = \sum_{e\in E} |f(e^+)^2 - f(e^-)^2|.$$

Fix some such f. Let $\beta_r > \ldots > \beta_1 > \beta_0 \geq 0$ be the values f takes and define

$$L_i = \{x \in V \mid f(x) \geq \beta_i\}.$$

This induces a a chain of subsets

$$L_r \subsetneq L_{r-1} \subsetneq \ldots \subsetneq L_1 \subsetneq L_0 = V.$$

Note that

$$\delta L_i = \{e \in E \mid f(e^+) \geq \beta_i \text{ and } f(e^-) < \beta_i \text{ or vice versa}\}.$$

Define

$$E_f = \{e \in E \mid f(e^+) \neq f(e^-)\}$$

and, for each $e \in E_f$, the pair $i(e)$ and $j(e)$ such that

$$\beta_{i(e)} = \max(f(e^+), f(e^-))$$

and

$$\beta_{j(e)} = \min(f(e^+), f(e^-)).$$

We have

$$B_f = \sum_{e \in E_f} |f(e^+)^2 - f(e^-)^2| = \sum_{e \in E_f} \beta_{i(e)}^2 - \beta_{j(e)}^2.$$

We introduce the terms $-\beta_{i(e)-1}^2 + \beta_{i(e)-1}^2 - \ldots - \beta_{j(e)+1}^2 + \beta_{j(e)+1}^2$ at each e to modify the above to:

$$B_f = \sum_{e \in E_f} \sum_{l=j(e)+1}^{i(e)} \beta_l^2 - \beta_{l-1}^2 = \sum_{i=1}^{r} |\delta L_i| (\beta_i^2 - \beta_{i-1}^e).$$

Now we wish to estimate B_f by $B_f \leq \sqrt{2k}\, \|df\|_2 \, \|f\|_2$. This is a matter of 'filling in':

$$\begin{aligned}
B_f &= \sum_{e \in E} |f(e^+) + f(e^-)| \cdot |f(e^+) - f(e^-)| \\
&\leq \left(\sum_{e \in E} (f(e^+) + f(e^-))^2 \right)^{\frac{1}{2}} \left(\sum_{e \in E} (f(e^+) - f(e^-))^2 \right)^{\frac{1}{2}} \\
&\leq \sqrt{2} \left(\sum_{e \in E} (f(e^+) + f(e^-))^2 \right)^{\frac{1}{2}} \|df\|_2 \\
&= \sqrt{2k} \left(\sum_{x \in V} f(x)^2 \right)^{\frac{1}{2}} \|df\|_2 \\
&= \sqrt{2k}\, \|f\|_2 \, \|df\|_2 .
\end{aligned}$$

by Cauchy-Schwarz and the fact that $(a+b)^2 \leq 2(a^2 + b^2)$.

Finally, we wish to relate B_f to $h(G)$. Suppose f is supported on at most half the vertices, that is

$$|\{v \in V \mid f(v) \neq 0\}| \leq \frac{|V|}{2}.$$

Then $B_f \geq h(G) \|f\|_2^2$:

Note that since there is $v \in V$ such that $f(v) = 0$, $\beta_0 = 0$. Since $|L_i| \leq \frac{|V|}{2}$ for $i > 0$, we have $|\delta L_i| \geq h(G)|L_i|$

by definition of $h(G)$. Thus

$$\begin{aligned}
B_f &\geq h(G)\sum_{i=1}^{r}|L_i|(\beta_i^2-\beta_{i-1}^2)\\
&= h(G)\left[|L_r|\beta_r^2+(|L_{r-1}|-|L_r|)\beta_{r-1}^2+\ldots+(|L_1|-|L_2|)\beta_1^2\right]\\
&= h(G)\left[|L_r|\beta_r^2+\sum_{i=1}^{r-1}(|L_i|-|L_{i+1}|)\beta_i^2\right].
\end{aligned}$$

Since $L_i \backslash L_{i+1}$ is precisely the set $\{v \in V \mid f(v)=\beta_i\}$, the term in brackets equals $\|f\|_2^2$.

We are finally ready to prove the second inequality. Let g be a real-valued eigenfunction of Δ with eigenvalue $k-\mu_1$, and set

$$V^+ = \{v \in V \mid g(v) > 0\}$$

and

$$f = \max(g, 0).$$

We may presume that f has support $|V^+| \leq \frac{|V|}{2}$ (since $g \neq 0$ and $\sum_{x\in V} g(x) = 0$, $|V^+| \neq \emptyset$ and $-g$ is also an eigenfunction). For $x \in V^+$, we have

$$\begin{aligned}
(\Delta f)(v) &= kf(v) - \sum_{x\in V} A_{vx}f(x)\\
&= kg(v) - \sum_{x\in V^+} A_{vx}g(x)\\
&\leq kg(v) - \sum_{x\in V} A_{vx}g(x)\\
&= (k-\mu_1)g(x).
\end{aligned}$$

Thus we get

$$\begin{aligned}
\|df\|_2^2 &= \langle \Delta f, f\rangle\\
&= \sum_{x\in V^+}(\Delta f)(x)g(x)\\
&\leq (k-\mu_1)\sum_{x\in V^+} g(x)^2\\
&\leq (k-\mu_1)\,\|f\|_2^2.
\end{aligned}$$

Combining the above with the second and third results on B_f, we get

$$\begin{aligned}
h(G)\,\|f\|_2^2 &\leq B_f\\
&\leq \sqrt{2k}\,\|df\|_2\,\|f\|_2\\
&\leq \sqrt{2k(k-\mu_1)}\,\|f\|_2^2.
\end{aligned}$$

Cancelling out $\|f\|_2^2 \neq 0$ yields the result.

□

We call $k-\mu_1$ the *spectral gap*. Good expanders have, by the above lemma, a large spectral gap, and equivalently graphs with a large spectral gap are good expanders. As such, we have the following alternate definition of expander graphs:

Definition 2.4. A family $G_1, G_2, \dots$ of k-regular undirected graphs, $G_m = (V_m, E_m)$, such that $|V_m| \to \infty$ as $m \to \infty$ is called a family of (spectral) ϵ-expanders if for every m and every non-trivial eigenvalue μ of G_m, we have $\mu \leq k - \epsilon$.

We also call a single k-regular graph G an ϵ-expander if all its non-trivial eigenvalues μ have $\mu \leq k - \epsilon$.

We have translated a graph-theoretic property into a linear-algebraic property. Rather than computing the rather esoteric isoperimetric constant of a graph, we need only compute the eigenvalues of a matrix. This latter problem is far more studied, and almost any programming language has an efficient built-in way to compute these eigenvalues.

We have the following limit lemmas on eigenvalues of graphs. These are theorems 1.3.1 and 1.3.3 in [DSV03]

Lemma 2.5. *Let $G_1, G_2, \dots$ be a family of k-regular, connected finite graphs $G_m = (V_m, E_m)$, such that $|V_m| \to \infty$ as $m \to \infty$. Then*

$$\liminf_{m \to \infty} \mu_1(G_m) \geq 2\sqrt{k-1}$$

where $\mu_1(G_m)$ is the largest non-trivial eigenvalue of G_m.

On the other side of the spectrum, we have a similar lemma, though with a more stringent requirement.

Lemma 2.6. *Let $G_1, G_2, \dots$ be a family of k-regular, connected finite graphs $G_m = (V_m, E_m)$, such that the girth $g_m$of G_m goes to ∞ as $m \to \infty$. Then*

$$\limsup_{m \to \infty} \mu(G_m) \geq 2\sqrt{k-1}$$

where $\mu(G_m)$ is the smallest (non-trivial) eigenvalue of G_m.

Thus, in the limit, the spectral gap cannot be larger than $k - 2\sqrt{k-1}$ and, if the G_i have increasing girth, the same restriction applies to the negative side of the spectrum. The significance of the following definition then becomes apparent.

Definition 2.7. A finite, k-regular graph X is called a *Ramanujan graph* if for every non-trivial eigenvalue μ of X, we have

$$|\mu| \leq 2\sqrt{k-1}$$

We can now state the mixing theorem on expander graphs we are interested in. This version and its proof are both lifted from [Gol01], lecture 10, theorem 5.

Theorem 2.8. *Fix $\epsilon > 0$ and $k \in \mathbb{Z}_{\geq 0}$. For every $\delta > 0$, there is an integer*

$$l_\delta = O(\log(1/\delta))$$

such that for every finite k-regular undirected ϵ-expander graph $G = (V, E)$, a random walk $X_0, X_1, \dots, X_{l_\delta}$ (of length l_δ) in G has, for every $v \in V$:

$$\left| \mathbb{P}(X_{l_\delta} = v) - \frac{1}{N} \right| \leq \delta$$

where $N = |V|$.

Remark. The way to think about this theorem is this: in $O(\log(\frac{1}{\delta}))$-steps, a walk in G is 'close to' the uniform distribution. In fact, if we take $\delta < \frac{1}{2N}$, we have that $\frac{1}{2N} \leq \mathbb{P}(X_l = v) \leq \frac{3}{2N}$. Thus a walk of $O(\log(N))$-steps is a constant factor removed from the uniform distribution over the vertices!

Additionally, the remarkable thing about this theorem is the fact that l is independent of the size of G; the logarithmic dependence on δ is true for any graph, but in the general version there is an additional dependence on the number of vertices.

Proof. Let A be the adjacency matrix of G and $A' = A/k$. Note that A' has entries in $[0,1]$ and the sum of any row or column is 1. A random walk over G has probability matrix A'. Let

$$\pi = \left(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right)$$

and $r(v)$ be the row of A' corresponding to a vertex v (i.e. $r(v)_w = A'_{v,w}$ is chance that a walk starting in v jumps to w).

Let $r_\perp(v) = r(v) - \pi$. Note that since the entries of $r(v)$ sum to 1, the entries of $r_\perp(v)$ sum to 0 and as such $r_\perp(v)$ is perpendicular to π.

Since $r_\perp(v) \in \pi^\perp$ and there is a basis consisting of eigenvectors of A (or A'), we have

$$\begin{aligned}\left\|(A')^l r(v) - \pi\right\|_\infty &\le \left\|(A')^l r(v) - \pi\right\|_2 \\ &= \left\|(A')^l (r_\perp(v) + \pi) - \pi\right\|_2 \\ &= \left\|(A')^l r_\perp(v)\right\|_2\end{aligned}$$

using that $A'\pi = \pi$

Define $\lambda = 1 - \frac{\epsilon}{k}$. All eigenvalues of A (except k) have absolute value $\le k - \epsilon$, and thus all eigenvalues of A' except 1 have absolute value $\le \lambda$. Thus, since the space of vectors perpendicular to π is spanned by the other eigenvectors of A', we have that $\|(A')w\| \le \lambda \|w\|_2$. Since $r_\perp(v)$ is perpendicular to π (and so is $Ar_\perp(v)$), we in particular have

$$\left\|(A')^l r_\perp(v)\right\| \le \lambda^l \left\|r_\perp(v)\right\|_2 \le \lambda^l \left\|r(v)\right\|_2 \le \lambda^l \left\|r(v)\right\|_1 = \lambda^l$$

Here we use that $r_\perp(v)$ is a projection of $r(v)$ and that $r(v)$ is a probability vector.

If $\lambda^l \le \delta$, then

$$|\mathbb{P}(X_l = v) - N^{-1}| \le \delta$$

This is equivalent to

$$l \ge \frac{\log(\delta)}{\log(\frac{1}{\lambda})}$$

which means that $l = O(\log(\frac{1}{\delta}))$. □

A final note before we get to the Pizer graphs. The above lemma will provide convergence to the uniform distribution in $O(\log(N))$ steps regardless of the expander coefficient ϵ. However, the estimations we make *do* depend on ϵ, and thus a better expander will have a quicker convergence and in the limit Ramanujan graphs will be the 'quickest'.

2.2 Pizer graphs

We will follow the conventions of Silverman with regards to elliptic curves, see [Sil09] for more detail.

Definition 2.9. Let p be a prime and $m \ge 0$ an integer coprime to p. Let $E_1, \dots, E_N$ be representatives of the isomorphism classes of supersingular elliptic curves over $\mathbb{F}_{p^2}$. The *Pizer graph of degree n over p* is the graph $\mathrm{Piz}_p(m)$ with vertex set $\{E_1, \dots, E_N\}$ and adjacency matrix $B_p(m)$ with entries:

$$B^p_{ij}(m) = \#\{C \subset E_i(\overline{\mathbb{F}}_p) \mid C \text{ a cyclic subgroup of order } n \text{ and } j(E_i/C) = j(E_j)\}$$

When the prime p we work over is clear, we will omit the p in the notation.

This corresponds to the isogeny graphs discussed in the introduction, since any isogeny is, up to composition with an isomorphism, wholly determined by its kernel. Thus, $B_{ij}(m)$ can also be seen as the number of degree-m isogenies from E_i and E_j, up to composition with automorphisms of E_j. Since we only consider curves up to isomorphism over $\overline{\mathbb{F}}_p$, we can and usually will use the set of j-invariants as the vertex set of $\mathrm{Piz}_p(l)$.

Remark. The Pizer graphs are directed. Our edges are isogenies, say $\phi : E_1 \to E_2$, which go in one direction. There is of course the dual map $\hat{\phi} : E_2 \to E_1$, which may initially seem to make our graph undirected. However, if E_1 has non-trivial automorphism group, we may run into trouble in the following way.

Let $\phi : E_1 \to E_2$ be an isogeny and $\sigma \in \mathrm{Aut}(E_1)$. We take duals to get

$$\widehat{\phi\sigma} = \hat{\sigma}\hat{\phi} = \sigma^{-1}\hat{\phi}.$$

Now, $\sigma^{-1}\hat{\phi}$ has the same kernel as $\hat{\phi}$, so these are the same edge in the Pizer graph, but

$$\ker(\phi\sigma) = \sigma^{-1}[\ker(\phi)].$$

If $\ker(\phi) \neq \sigma^{-1}(\ker(\phi))$, then ϕ corresponds to another edge in the graph than $\phi\sigma$. Thus we cannot cancel edges via duals to make our graphs undirected.

If $\sigma = \sigma^{-1} = \pm 1$, then of course $\sigma^{-1}(\ker(\phi)) = \ker(\phi)$, so if $\mathrm{Aut}(E_1) = \{\pm 1\}$, this problem does not arise. This is the case for all j-invariants except $j = 0$ and $j = 1728$, which have larger automorphism groups. Later on, we will see that these j-invariants are ordinary (i.e. not supersingular) modulo p if (and only if) $p \equiv 1 \mod 12$. In this case, we *can* view our graphs as undirected.

We will want to relate walks in the graph to isogenies of prime power degrees. We prove the following proposition:

Proposition 2.10. *Let $p \neq l$ be primes and E, E' be supersingular curves over $\mathbb{F}_{p^2}$.*

Composition of isogenies provides a surjection from the walks of length n in $Piz_p(l)$ between E and E' onto the isogenies of degree l^n between E and E' up to composition with isomorphisms.

Moreover, restriction of this map to walks without backtracking yields a bijection to isogenies with cyclic kernel.

Proof. Let $(e_1, \ldots, e_n)$ be a walk in $\mathrm{Piz}_p(l)$ from E to E'. The edges e_i correspond to degree-l isogenies $\phi_i : E_{i-1} \to E_i$, where $E_0 = E$ and $E_n = E'$. Composing these isogenies yields a map

$$\phi_n \circ \phi_{n-1} \circ \ldots \circ \phi_1 : E \to E'$$

of degree l^n. On the other hand, given a degree-l^n isogeny $\psi : E \to E'$, the kernel $\ker(\psi)$ contains a subgroup, say C_1, of degree l. We can consider the map

$$\phi_1 : E \to E/C_1$$

Write $E_1 = E/C_1$. We know that $\phi_1[\ker(\psi)] \simeq \ker(\psi)/C_1$ has degree l^{n-1}. If $n > 1$, $\phi_1[\ker(\psi)]$ also contains a subgroup of degree l, and we can consider

$$\phi_2 : E_1 \to E_1/C_2$$

Repeating the steps above n times, we get a sequence of morphisms

$$\phi_i : E_{i-1} \to E_i$$

where $E_0 = E$ and $E_n = E'$. These ϕ_i correspond to edges e_i in $\mathrm{Piz}_p(l)$, and $(e_1, \ldots, e_n)$ is a walk from E to E' whose composition is ψ. Note that the choice of subgroup C_1 is *not* always unique, which may yield different compositions (so this map is not per se injective).

It remains to prove that if the kernel of ψ is cyclic, the choice of subgroup *is* unique, and *the* corresponding walk in this case does not backtrack. The walk $(e_1, \ldots, e_n)$ backtracks if, for some i, e_i corresponds to the dual map of e_{i-1}.

Since our map ψ in this case has cyclic kernel, it contains a unique subgroup of order l, so the choice of C_1 as above is fixed. Since $\phi_1[\ker(\psi)]$ is the image of a cyclic group under a group morphism, it is cyclic also, and the choice of C_2 is also uniquely determined. Thus ψ can be uniquely decomposed (up to composition with isomorphisms) as a walk in the graph.

A walk with backtracking clearly has non-cyclic kernel: the composition of ϕ and $\hat{\phi}$ is $[l]$, with kernel $(\mathbb{Z}/l\mathbb{Z})^2$, a non-cyclic group.

It remains to see that a walk $(e_1, \ldots, e_n)$ *without* backtracking corresponds to an isogeny with cyclic kernel. Let $\phi_1, \ldots, \phi_n$ be the corresponding isogenies and suppose that

$$\psi = \phi_n \circ \ldots \circ \phi_1$$

has non-cyclic kernel.

Note that ϕ_1 has cyclic kernel (as it is of degree l). Let us write

$$\psi_i = \phi_i \circ \ldots \circ \phi_1$$

This is the 'truncated' version of ψ. Let m be such that ψ_{m-1} has cyclic kernel, but ψ_m has non-cyclic kernel. The kernel of ψ_{m-1} is (isomorphic to) $\mathbb{Z}/l^{m-1}\mathbb{Z}$. Since ψ_m must have non-cyclic kernel and this kernel must contain $\ker(\psi_{m-1})$, $\ker(\psi_m)$ must be isomorphic to $(\mathbb{Z}/l^{m-1}\mathbb{Z}) \times \mathbb{Z}/l\mathbb{Z}$. We see that

$$\psi_{m-2}(\ker(\psi_m)) \simeq (\mathbb{Z}/l\mathbb{Z})^2$$

An elliptic curve contains a unique subgroup of isomorphism type $(\mathbb{Z}/l\mathbb{Z})^2$, namely the l-torsion, and since $\psi_m = \phi_m \circ \phi_{m-1} \circ \psi_{m-2}$, $\phi_m \circ \phi_{m-1}$ must be $[l]$. But this means that

$$\phi_{m-1} = \hat{\phi}_m$$

This is in contradiction to the presumption that our walk was without backtracking. □

2.2.1 Computing Pizer graphs

We would like to get a better understanding of these graphs. One way to do this is simply to compute some of them and see if we can find statistical 'evidence' for their (cryptographic) properties. We discuss some algorithms for computing these graphs.

To compute $B_{ij}(m)$, one can compute the m-torsion of E_i and for every (cyclic degree-m) subgroup of this m-torsoin $C \subset E_i[l]$, compute $j(E_i/C)$. This leads to the following algorithm:

Algorithm 2.11. Input: Primes p and l, $l \neq p$.
Output: The adjacency matrix of the l-isogeny graph over $\mathbb{F}_{p^2}$.

1. Find a supersingular $j-$invariant j_1 over $\mathbb{F}_p$, and a corresponding curve E_1.

2. Record j_1 on the list of j-invariant to process and list of known j-invariants.

3. Let B be the 1-by-1 zero matrix.

4. While the list of to-process j-invariants isn't empty:
   (a) Let j' be a to-process j-invariant. Compute an elliptic curve E with j-invariant j'.
   (b) Compute the l-torsion of E, and find all $l+1$ subgroups of order l. Make a list of these groups, say $\{G_0, \dots, G_l\}$
   (c) For each $k \in \{0, \dots, l\}$:
      i. Compute the isogeny ϕ_k with kernel G_k. Let E_k be its co-domain and $j_k = j(E_k)$.
      ii. If j_k is on the list of known j-invariants, add 1 to B_{j',j_k}. Else, create a new row and column in B labeled by j_k with a 1 in the j'-th column and j_k-th row.
      iii. Add j_k to the list of to-process j-invariants.
   (d) Add j' to the list of finished j-invariants.
   (e) Remove from the list of to-process j-invariants all finished j-invariants.
5. Return B and the list indices of B.

The idea of the above algorithm is simple, and quite generally applicable to computations of complete graphs. We start with some vertex in the graph, calculate all the points to which it is connected, then jump to one of these vertices and repeat until we can only reach vertices we've already treated. Since we know that our graph is connected, we know we've truly found every vertex and edge in our graph.

Before we continue with some potential improvements, we should firstly know how we can do any of the above. More precisely:

1. Given j, how do we find a curve of j-invariant j?
2. How do we find a starting vertex, i.e. a supersingular elliptic curve E over $\mathbb{F}_{p^2}$?
3. How do we compute the l-torsion of a given elliptic curve?
4. Given a point or subgroup in an elliptic curve, how do we compute the accompanying isogeny, or the j-invariant of its codomain?

The first is easy, and is mentioned in Silverman, see [Sil09], proposition III.1.4c; the curve given by $y^2 = x^3 + 1$ has j-invariant 0 and the curve given by $y^2 = x^3 + x$ has j-invariant 1728. For j not equal to 0 or 1728, the curve given by

$$y^2 = x^3 - \frac{36}{j - 1728}x - \frac{1}{j - 1728}$$

has j-invariant j.

2.2.2 Finding supersingular j-invariants

Moving on to point 2. Computing a supersingular j-invariant in characteristic p can be achieved using *CM-theory of elliptic curves.* We follow the ideas of Bröker from [Brö09]. An important basic result of CM-theory is the following:

Theorem 2.12. *Let E be a CM-elliptic curve over some number field L, with endomorphism ring the maximal order in a quadratic imaginary number field K. Let $\mathfrak{p}$ be a prime of L lying above $p \in \mathbb{Z}$ such that E has good reduction at $\mathfrak{p}$. Then $E \mod \mathfrak{p}$ is supersingular if and only if p remains inert in K.*

See for instance [Lan87], theorem 13.12. In this book (chapter 10) one can also find the following result: given an elliptic curve as in the theorem above, there is an irreducible monic polynomial in $\mathbb{Z}[X]$ of degree equal to the class group of K, which we call P_K, such that the roots of P_K in $\mathbb{C}$ are the j-invariants of elliptic curves with endomorphism rings isomorphic to O_K.

Furthermore, P_K remains irreducible in $K[X]$ and $K[X]/P_K$ is the *Hilbert class field* of K. Let p be a prime that remains inert in K. The elliptic curves defined by the roots of P_K are supersingular curves mod p (or modulo the appropriate ideal above p). As such, P_K splits in $\mathbb{F}_{p^2}$ (as all supersingular j-invariants of characteristic p are defined here) and thus if P_K is of odd degree, it is guaranteed to have a root in $\mathbb{F}_p$. We have the following, which is lemma 2.3 in [Brö09].

Lemma 2.13. *Let $K = \mathbb{Q}(\sqrt{-d})$, where $d \in \mathbb{Z}_{\geq 2}$ is square-free. Let h_K be the class number of K. We have:*

$$h_K \text{ is odd} \iff d = 2 \text{ or } d \equiv 3 \mod 4 \text{ and prime}$$

Proof. Let D be the discriminant of K and $p_1, \ldots, p_r$ the *odd* prime divisors of D. We consider the *genus field*

$$G = K(\sqrt{p_1^*}, \ldots, \sqrt{p_r^*}).$$

Here, for odd primes p,

$$p^* = (-1)^{\frac{p-1}{2}} p = \begin{cases} p & p \equiv 1 \mod 4, \\ -p & p \equiv 3 \mod 4. \end{cases}$$

G is the largest unramified abelian extension of K that remains abelian as an extension of $\mathbb{Q}$. The Galois group $\mathrm{Gal}(G/K)$ is the Sylow-2 subgroup of the class group of K.

Thus $\mathrm{Gal}(G/K)$ is trivial if and only if h_K is uneven. Note that G always contains $\sqrt{p^*}$ for any odd prime divisor of d. Write

$$d = 2^b q_1 q_2 \ldots q_s,$$

where q_i are odd primes. Note that $b = 0, 1$. If d is not prime, $\mathbb{Q}(\sqrt{-d})$ does *not* contain $\sqrt{-q_1}$, so $K \neq G$. Thus if d is *not* prime, $\mathbb{Q}(\sqrt{-d})$ has *even* class number.

We see that d must be a prime then. The case $d = 2$ is trivial, so suppose d is an *odd* prime.

If $d \equiv 1 \mod 4$, G contains the real quadratic number field $\mathbb{Q}(\sqrt{d})$, so $[K : G] = 2$.

Else, if $d \equiv 3 \mod 4$, $D = -d$ and $G = K(\sqrt{-d}) = K$. □

Note that if K has class number 1, the j-invariant of E is defined over $\mathbb{Z}$ and computing $E \mod p$ is very easy. There are 9 quadratic imaginary fields with trivial class group. As such, to compute a supersingular curve mod p, we could start by checking if there is such a field K where p remains inert. This is a congruence relation modulo some fixed number, since if K has discriminant D, then p is inert if and only if $(\frac{D}{p}) = -1$, which by quadratic reciprocity is a congruence mod D.

If this fails, we can employ the following general method. It is more computational work, but will always produce a supersingular j-invariant.

Let us simply give the algorithm as Bröker does.

Algorithm 2.14. Input: a prime p
Output: a supersingular j-invariant in $\mathbb{F}_p$

1. If $p = 2$, return 0
2. If $p \equiv 3 \mod 4$, return 1728
3. Find small prime $q \equiv 3 \mod 4$ such that $-q$ is *not* a square mod p

4. Compute $P_K \in \mathbb{Z}[X]$ for $K = \mathbb{Q}(\sqrt{-d})$
5. Find a root j of $P_K \in \mathbb{F}_p[X]$
6. Return j

We briefly discuss these steps. Step 1 and 2 are simple cases; by CM-theory, 1728 is supersingular modulo an odd prime p if and only if $p \equiv 3 \mod 4$.

We construct q such that p is inert in $K = \mathbb{Q}(\sqrt{-q})$ and h_K is odd. Because p remains inert, P_K splits over $\mathbb{F}_{p^2}$ (as its root define supersingular j-invariants, and these all lie in $\mathbb{F}_{p^2}$). Thus if $h_K = \deg(P_K)$ is odd, P_K has a root in $\mathbb{F}_p$.

To find such a q, we can simply try small primes. The density of primes that are 3 mod 4 is $\frac{1}{2}$, and also the density of primes where $-q$ is a square mod p is $\frac{1}{2}$. Thus a random prime has chance $\frac{1}{4}$ to be sufficient for step 3, and we expect q to be rather small.

There is a worst-case upper bound; under assumption of the generalised Riemann hypothesis, there is effectively computable $c \in \mathbb{R}_{>0}$ such that a sufficient q exists with

$$q \leq c \log(4p^2)^2.$$

The computation of P_K is harder. This has degree equal to the class number. There is a theorem of Siegel that states that the class number of $\mathbb{Q}(\sqrt{-d})$ grows in $O(\sqrt{d})$. We can be somewhat more precise, and follow the introduction of [BM19]. For $d \in \mathbb{Z}_{\geq 1}$, write $h(d)$ for the class number of $\mathbb{Q}(\sqrt{-d})$. Using techniques from analytic number theory, one can prove that, as $d \to \infty$,

$$h(d) \ll \sqrt{d} \log(d).$$

There is also the following theorem, which provides a lower bound:

Theorem 2.15. *Define, for $d \in \mathbb{Z}_{\geq 1}$, $h(d)$ to be the class number of the imaginary quadratic number field $\mathbb{Q}(\sqrt{-d})$. For all $\epsilon > 0$, there is effectively computable constant $c(\epsilon)$ such that*

$$h(d) > c(\epsilon) d^{\frac{1}{2} - \epsilon}.$$

This is to say that the class number of $\mathbb{Q}(\sqrt{-d})$ grows *at least* as quickly as $\sqrt{d}$ as $d \to \infty$. The computability of $C(\epsilon)$ is a later result of Dorian Goldfeld in [Gol77]. This yields to

$$\lim_{d \to \infty} \frac{\log(h(d))}{\log(d)} = \frac{1}{2}$$

In order to compute P_K for a general quadratic order $\mathbb{Z}[\alpha]$, note that $\mathbb{Z}[\alpha] = \mathbb{Z} + \alpha\mathbb{Z}$ and this is closed under multiplication. In particular, as a lattice, it has endomorphism ring $\mathbb{Z}[\alpha]$, and hence the elliptic curve $\mathbb{C}/\mathbb{Z}[\alpha]$ also has endomorphism ring $\mathbb{Z}[\alpha]$. Since P_K is irreducible, it is the *minimum* polynomial of *any* j-invariant with endomorphism ring O_K. In particular, it is the minimum polynomial of $j_K = j(\mathbb{C}/O_K)$. If L is any field extension of $\mathbb{Q}$ containing all the roots of j_K, then:

$$P_K = \prod_{j \in \mathrm{Aut}(L) j_K} (X - j)$$

We will not give a full algorithm for the computation of this polynomial. One can find an efficient way to do so in [Brö08]. This algorithm uses the Galois action on field extensions of $\mathbb{Q}_p$; one can compute these actions with enough precision to deduce the coefficients (recall that $P_K \in \mathbb{Z}[X]$). If (the maximal order of) K has discriminant D, the outlined method has running time $O(|D| \log(|D|)^{8+\epsilon})$ for all $\epsilon > 0$.

Finally, we must compute a root of P_k mod p. We can use algorithm 14.15 from [vzGG13], which completes in $O(\deg(f) \cdot \log(p)^2))$ time.

Hence running time we 'expect' to see is very good: we expect to find small q, and hence small class number h_K. The computation of P_K will thus be easy, and we are only left with $O(\log(p)^2)$ from the computation of a root of P_K mod p.

In the worst case, $q = O(\log(p)^2)$, so the class number h_K is

$$h_K = \deg(P_K) = O(\log(p)).$$

The computation of P_K takes

$$O(\log(p)^2 \log(\log(p))^{8+\epsilon}).$$

Finally, finding a root of P_K mod p takes

$$O(\deg(P_K) \log(p)^2) = O(\log(p)^3),$$

so the algorithm has a total worst-case running time of

$$O(\log(p)^3).$$

This compare favourably to 'naive' methods. The most naive method would be simply taking random j-invariants in $\mathbb{F}_{p^2}$ until one is supersingular. There are approximately $p/12$ supersingular j-invariants, so this has running time

$$O\left(\frac{p^2}{p/12}\right) = O(p).$$

The number of supersingular elliptic curves defined over $\mathbb{F}_p$ is approximately the class number of $\mathbb{Q}(\sqrt{-p})$. This is a consequence of the trace formula computed in Section 4.3 (for $m = p$).

This class number grows at order $O(\sqrt{p})$. Thus guessing in $\mathbb{F}_p$ has running time

$$O\left(\frac{p}{\sqrt{p}}\right) = O(\sqrt{p}).$$

Brökers algorithm has a far more attractive running time. One should of course remember that for small primes p, the implementation of Brökers algorithm is rather overkill, as $O(\sqrt{p})$ is a running time one can work with; it certainly was not the bottleneck for our implementation. If on the other hand one is interested in implementing, say, the hash function outlined in the introduction, one needs very large primes (of the order 2^{1024} or larger), and perhaps finding a starting vertex would be a significant part of the running time.

2.2.3 The torsion and Velù

The l-torsion then. For $l = 2$, this is easy. Since this is the only case we actually use, we will treat this in some more detail. If our curve is defined by $y^2 = f(x)$, then the 2-torsion is given by the points $(x_1, 0), (x_2, 0), (x_3, 0)$, where the x_i are the roots of f, and, under the correct choice of isomorphism class, all these points are defined over $\mathbb{F}_{p^2}$.

As mentioned under Bröker, there are only so many possible values for the trace of the p^2-power Frobenius. This is the first statement in [Brö09].

Theorem 2.16. *Let E be a supersingular elliptic curve over $\mathbb{F}_{p^2}$ and π the p^2-power Frobenius endomorphism of E. Write t for its trace. Then t can have the following values:*

1. $t = 2 \pm p$

2. $t = \pm p$ *if* $p \not\equiv 1 \mod 3$

3. $t = 0$ *if* $p \not\equiv 1 \mod 4$

Furthermore, there is always an elliptic curve E' *isomorphic to* E *(over* $\overline{\mathbb{F}}_p$*) such that the trace of Frobenius on* E' *is* $\pm 2p$*.*

Now, we can compute the discriminant of $\pi - [1]$ (here $[n]$ is the multiplication-by-n map on E): it is $t^2 - 4p^2$ (cf. the proof of lemma 3.7 of [Sch87]). As such, if $t = 2 \pm p$ or $t = 0$, $\pi - 1 \in \mathbb{Z}$ and thus $\pi \in \mathbb{Z}$, that is, there is $n \in \mathbb{Z}$ such that $\pi = [n]$. In fact, looking at degrees, $n = \pm p$. If we write $E[n]$ for the n-torsion of E, we get

$$E(\mathbb{F}_p) = \ker(\pi - [1]) = E[\pi - [1]] = E[p \pm 1]$$

Thus, we get the more general result that the m-torsion of E is defined over $\mathbb{F}_{p^2}$ if and only if $m \mid \pi - 1$. Which m-torsions are defined depends on whether $\pi = p$ or $\pi = -p$, but since $2 = \gcd(p-1, p+1)$, the 2-torsion is *always* defined over $\mathbb{F}_{p^2}$.

For $l > 2$, this is somewhat more difficult. Since the addition law on elliptic curves is given by polynomials in the coordinates, we can find roots by again solving polynomials. In other words, since the multiplication-by-l map is a morphism, it is given by rational functions, and has a global representation on the affine patch. In fact this to some degree the method we use for determining the 2-torsion. We will not go into much further detail, but note that these division polynomials quickly grow large and the roots might be defined over a large field. By the previously mentioned lemma 3.7 from [Sch87], one could determine these fields in more detail.

Finally, the final question we had regarding Algorithm 2.11 was how to actually compute the isogenies belonging to a certain point on an elliptic curve, or (the j-invariant of) the codomain. For this, there are the formulae of Velù, [V́71]. These are implemented in Sage, and are the standard manner of arithmetic on elliptic curves.

2.2.4 A second algorithm

The main sticking point of the above algorithm is, then, having to compute the entire l-torsion. Here, there is powerful tool that can give us a significant speedup; the *modular polynomial.*

For a definition, see [Sil94], exercise 2.18. The properties that are important for us are the following:

- For $n \in \mathbb{Z}_{\geq 0}$, $\Phi_n(X, Y) \in \mathbb{Z}[X, Y]$
- For $z_1, z_2 \in \mathbb{C}$, $\Phi_n(z_1, z_2) = 0$ if and only if there are elliptic curves E_1, E_2 with j-invariants z_1 and z_2 and an isogeny $\psi : E_1 \to E_2$ of degree n whose kernel is cyclic.
- The multiplicity of a root z_1 of $\Phi_n(X, X)$ equals the *number* of cyclic endomorphisms of elliptic curves E_1 over $\mathbb{C}$ of j-invariant z_1 of degree n.
- The above properties hold true if z_1 and z_2 are elements of a finite field $\mathbb{F}_q$, and the curves are defined over $\mathbb{F}_q$.

Calculating these modular polynomials is no easy feat; the algorithm in [BLS12] has running time $O(l^3 \log(l)^3 \log(\log(l)))$, for l a prime. The coefficients quickly become unwieldy: for instance,

$$\Phi_2(X, Y) = X^3 - 162000X^2 + 1488X^2Y - X^2Y^2$$

$$+8748000000X + 40773375XY - 157464000000000$$

and

$$\Phi_3(X,Y) = 1855425871872000000000X - 770845966336000000XY$$
$$+452984832000000X^2 + 8900222976000X^2Y + 2587918086X^2Y^2 + 36864000X^3$$
$$-1069956X^3Y + 2232X^3Y^2 - X^3Y^3 + X^4$$

Still, once one has this polynomial, for fixed $n = l$ prime, one can upgrade to the following:

Algorithm 2.17. Input: A prime p such that $p \neq l$.
Output: The l-isogeny graph over $\mathbb{F}_{p^2}$.

1. Find a supersingular j-invariant j over $\mathbb{F}_p$.
2. Add j onto the list of j-invariants to process and list of known j-invariants.
3. Let B be the 1-by-1 zero matrix.
4. While the list of to-process j-invariants isn't empty:
    - (a) Let j' be a to-process j-invariant. Compute $f = \Phi_l(j', X) \in \mathbb{F}_p[X]$.
    - (b) Find the roots of f over $\mathbb{F}_{p^2}$. say $\{j_0, \ldots, j_l\}$
    - (c) For each $k \in \{0, \ldots, l\}$:
        - i. If j_k is on the list of known j-invariants, add 1 to B_{j',j_k}. Else, create a new row and column in B labeled by j_k with a 1 in the j'-th column and j_k-th row.
        - ii. Add j_k to the list of to-process j-invariants.
    - (d) Add j' to the list of finished j-invariants.
    - (e) Remove from the list of to-process j-invariants all finished j-invariants.
5. Return B and the list indices of B.

Remark. Neither the above nor Algorithm 2.11 before makes use of the fact that the Pizer graph is 'mostly undirected'. That is to say, if there is a vertex from j_1 to j_2, then there is another vertex from j_2 to j_1; such a vertex represents an isogeny $\psi : E_1 \to E_2$, which has a dual isogeny of the same degree $\hat{\psi} : E_2 \to E_1$ (recall the discussion above Proposition 2.10). This 'mostly undirected' property can be useful; whenever we find an automorphism $j_1 \to j_2$, we can remember this. We know that one of the roots of $\Phi_l(j_2, X)$ is j_1, so we can calculate $\Phi_l(j_2, X)/(X - j_1)$ and find the remaining roots. This polynomial will have lower degree, and as such finding roots here will be much easier!

In the same vein, for Algorithm 2.11, if we fix a specific curve for each j-invariant, we can compute a point on the l-torsion to start. Let $E_1[l] = \{P_1, \ldots, P_{l^2}\}$ be the l-torsion of E_1 ($P_1 = O$ and let $\phi : E_1 \to E_2$ be the isogeny with kernel $\langle P_2 \rangle$. If $P_3 \notin \langle P_2 \rangle$, then $\phi(P_3)$ will be an order-l point on E_2, and we can compute l points in the l-torsion already. In fact, $E_1[l] = \langle P_2, P_3 \rangle$, so the isogeny with kernel $\phi(P_3)$ will correspond to the *dual isogeny* of ϕ.

2.3 Implementation and data

For the research of this thesis, we used the first algorithm, with no speedups. Furthermore, these computations were implemented in Sage, without specialized add-ons, on a home computer not dedicated to this program alone and thus very far from optimized. The code was also written by the author of this thesis, who is not an expert programmer.

Still, the spectra of the Hecke operators of all primes up to $3 \cdot 10^5$ are available. The time-usage is plotted below. By experimentation, the curve seems lie between $O(x^2)$ and $O(x^3)$ and happens to line up nicely with $C \cdot x^2\sqrt{x}$, though these are experimental results and should not be taken as fact. For clarification, a point on the blue line has coordinated (p, t), where p is a prime and t is the time (in seconds) our code took to compute the adjacency matrix of $\mathrm{Piz}_p(l)$.

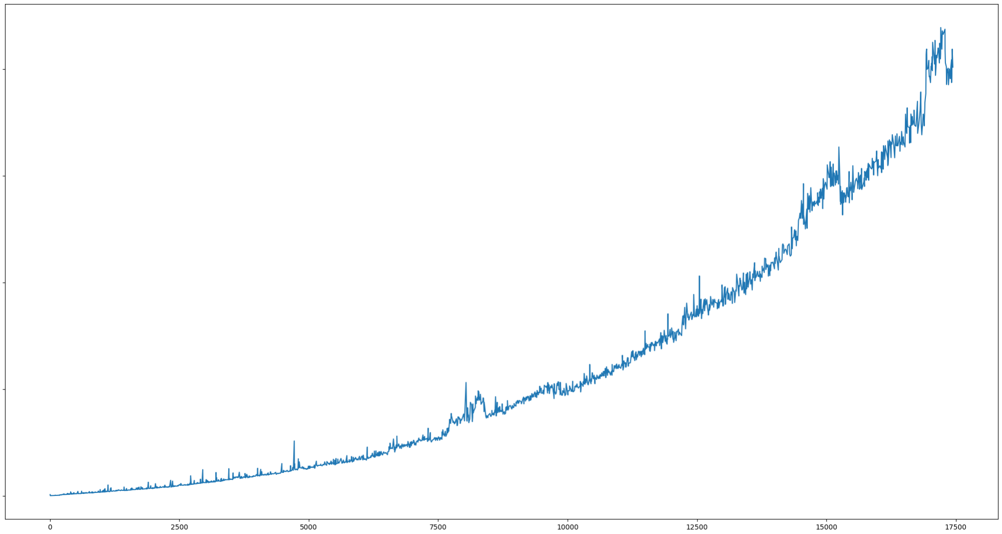

Figure 1: Time taken per prime

The code we used to implement the algorithms discussed above is included in the appendices. We used this code to calculate the spectra of the Pizer graphs $\mathrm{Piz}_p(2)$ for p between 5 and 38113. We could of course have analysed the entire adjacency matrix, but this would have taken even longer, and most of the properties we seek can be derived from the spectrum (to be clear, we do compute the entire matrix and only then the spectrum. However, storing each of the 3500 matrices, whose size scales quadratically in the size of the prime, would take a lot of storage. Additionally, Python does not write-to-file very quickly, thus in storing this data, we would also 'waste' a lot of time). We did save the full matrices for some small primes.

As such, we can give some examples. The first interesting example where $p \equiv 1 \mod 12$ is $p = 37$. In the picture, the labels are the j-invariants of the elliptic curves, and a is a root of $X^2 + 4X - 2$, that is, a generator of $\mathbb{F}_{37^2}$ over $\mathbb{F}_{37}$ (this is a standard generator Sage chooses).

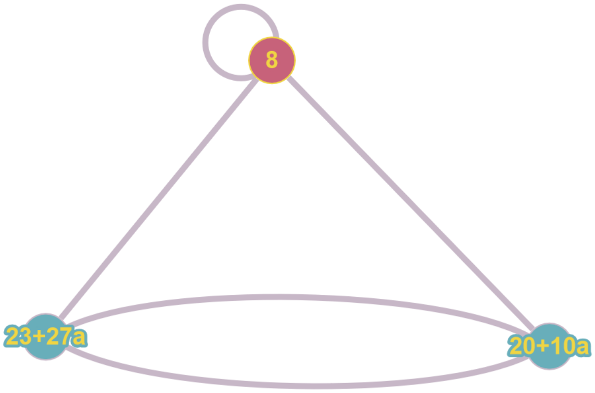

Figure 2: $\mathrm{Piz}_{37}(2)$

For an example where $\mathrm{Piz}_p(l)$ is directed, we chose $p = 67$. Again, a is a generator of $\mathbb{F}_{67^2}$ and a root for $X^2 + 4X - 2$. Note that there are 2 arrows from 1728 to 66, but only one from 66 to 1728. This is because the dual of the first arrow is the dual of the second arrow composed with an automorphism of (the curve with j-invariant) 1728. Still, the graph is 3-regular in the sense that the out-degree of each node is 3.

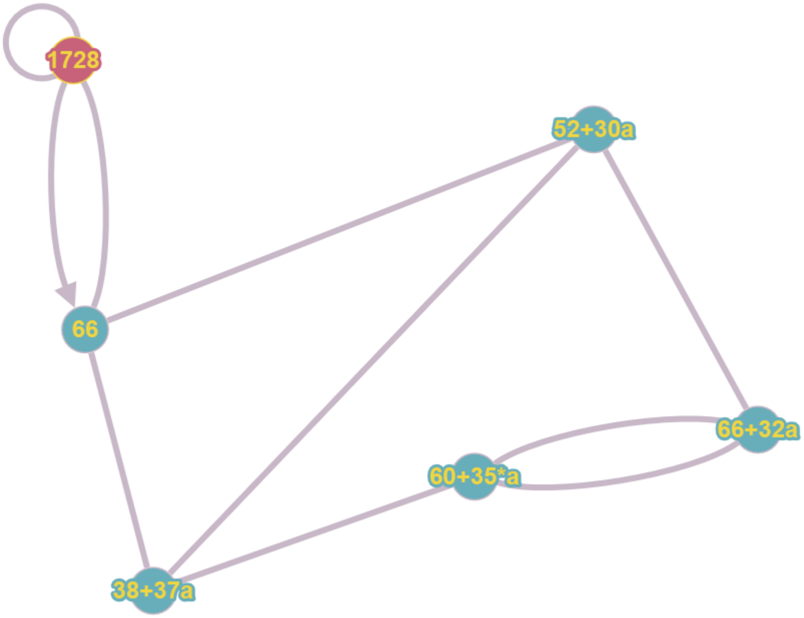

Figure 3: $\mathrm{Piz}_{67}(2)$

The reason for our interest in the Pizer graphs is of course because they are Ramanujan graphs. Below we see that this does indeed fit with our data: in blue is the value of the second largest eigenvalue of $\mathrm{Piz}_p(l)$, in red is the constant line $y = 2\sqrt{2}$. In chapter 3, we will prove the following:

Theorem 2.18. *For every pair of primes l and p, $Piz_p(l)$ is a Ramanujan graph. That is, for every eigenvalue λ of $Piz_p(l)$, we have $\lambda = l + 1$ or $\lambda \leq 2\sqrt{l}$. Furthermore, $\lambda = l + 1$ has (geometric and algebraic) multiplicity 1.*

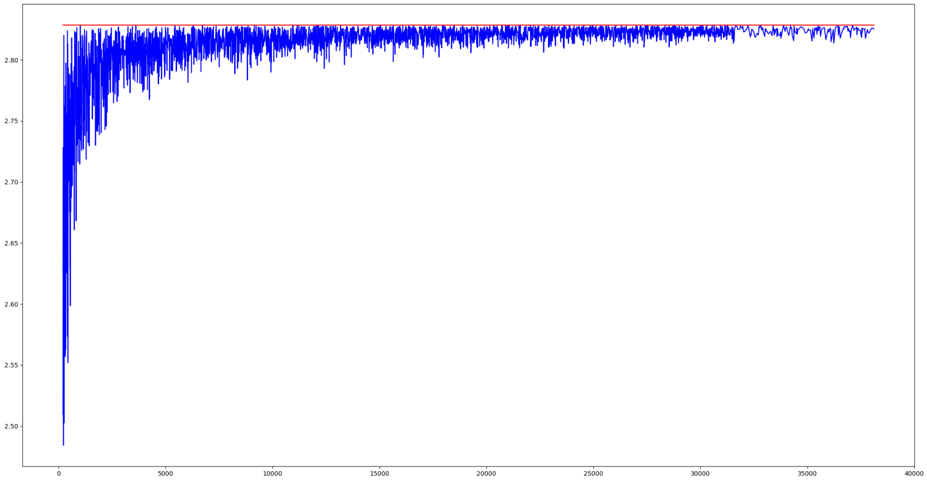

Figure 4: Value of the 2nd largest eigenvalue

We see that the statement is indeed true; the Pizer graphs have a large spectral gap.

When one is given a large amount of sets of data, a natural thing to do is to look for patterns between these sets. As such, we simply plotted all the spectra and made a .gif-file from them. Regrettably, moving pictures cannot as of yet be printed, and we hence cannot include this gif here. From this figure, it was clear that the eigenvalues followed a non-uniform distribution, that was still symmetric around 0. Below is plotted the ordered spectrum of $\mathrm{Piz}_{38113}(2)$; this spectrum is a tuple of real numbers $(\mu_1, \dots, \mu_n)$ such that $\mu_m \leq \mu_{m+1}$. A point on the blue line is of the form (m, μ_m) for some $1 \leq m \leq n$. In red is the straight line between $(1, \mu_1)$ and (n, μ_n). Note that 3 is not included in the spectrum, as it is obvious that it is an eigenvalue, and would only show as a strange 'jump' in the top-right of the graph.

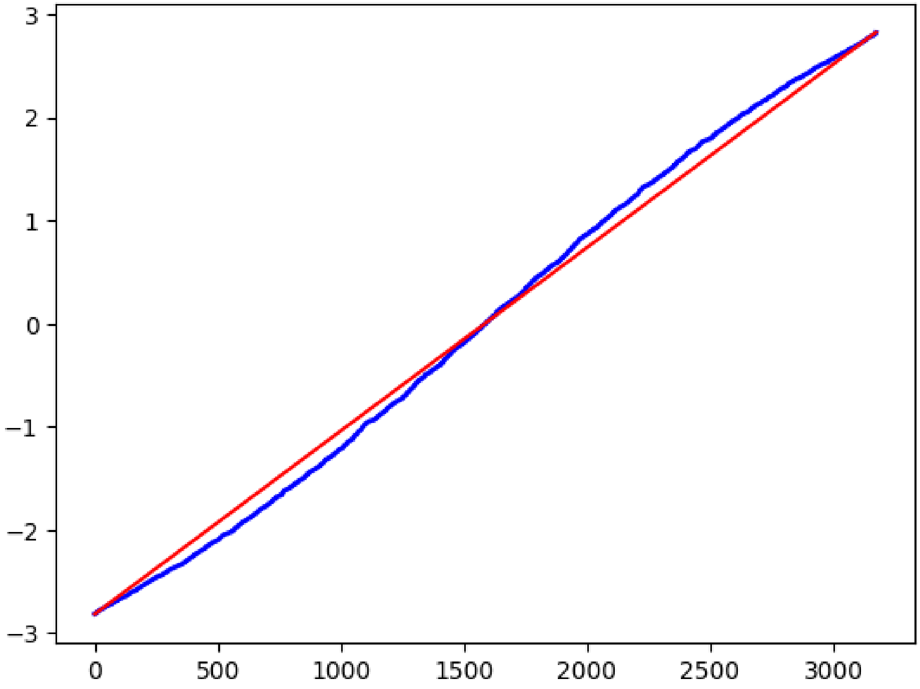

Figure 5: Spectrum at 38113

One notices that there is a clearly visible gap between the linear approximation and the spectrum. This was consistent for all other spectra, and indicates that the eigenvalues are not distributed according to the uniform distribution. The next thing to do is to simply count the occurrences of each eigenvalue.

Of course, these are algebraic numbers, so exact repetition will be quite rare, but if we round to 3 digits, we will see repetition. We computed over 4.5 million eigenvalues in total. Below is plotted our count. To clarify, a point (x, y) on the (blue) graph indicates that there were y instances of eigenvalue x occurring.

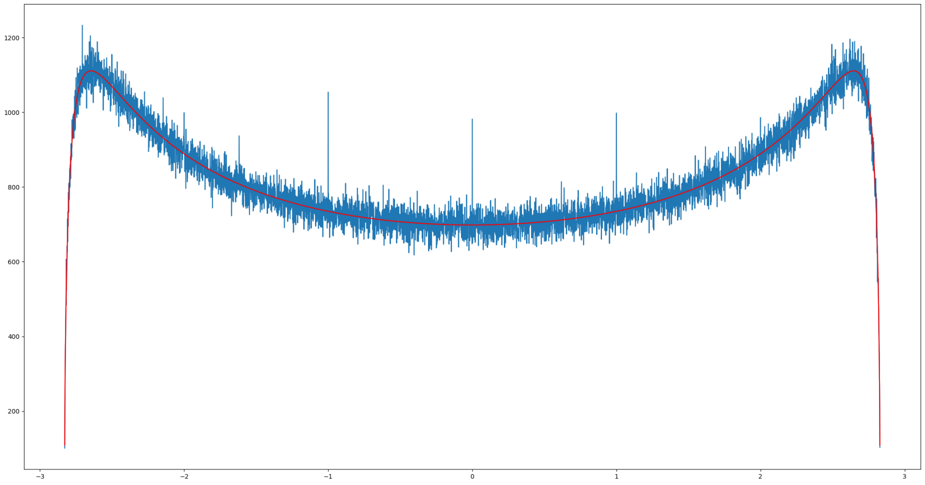

Figure 6: Occurrence of Eigenvalues, rounded to 3 digits

In red, there is plotted the function $\frac{3\sqrt{8-x^2}}{2\pi(9-x^2)}$. This comes from [McK81], and is the predicted density function for the probability distribution for the eigenvalues of large 3-regular graphs. The count follows this line very closely. This is intriguing; the proof in McKay's paper relies on an additional requirement that the total number of k-cycles grows sublinearly with the size of the graph, which is *not* the case of our graphs (as we see in Lemma 4.16). Still, this distribution seems to be maintained! We will, in chapter 4, get to proving the following.

Theorem 2.19. *Fix a prime l. Let $E_l(N)$ be the set of (non-trivial) eigenvalues of $Piz_p(l)$. Define*

$$F_l(x,p) = \frac{\#\{\mu \in E_l(p) \mid \mu \leq x\}}{\#E_l(p)}$$

Then, for all x in $[-2\sqrt{l}, 2\sqrt{l}]$, $F_l(x,p)$ converges to

$$F_l(x) = \int_{-2\sqrt{l}}^{x} \frac{(l+1)\sqrt{4l-x^2}}{2\pi((l+1)^2 - x^2)} dx$$

as $p \to \infty$.

The proof of the above theorem (discussed in chapter 4) relies on the relationship between quaternion algebras and isogeny graphs as discussed in Section 4.3. In particular, it relies on a formula for the trace, for which we must discuss the embedding theory of orders into quaternion algebras. For the special case $l = 2$, we do not need this; the proof of the below theorem is entirely CM-related.

Proposition 2.20. *For any prime p, we have*

$$Tr(Piz_p(2)) = 2 - \left(\frac{-7}{p}\right) - \frac{1}{2}\left(\left(\frac{-4}{p}\right) + \left(\frac{-2}{p}\right)\right).$$

In particular, the trace of $Piz_p(2)$ depends only on $p \mod 56$.

Proof. The trace of $\mathrm{Piz}_p(2)$ is, by the correspondence to our matrices, precisely the number of supersingular j-invariants with degree 2 endomorphisms, weighted by the *amount* of endomorphisms, i.e.

$$\mathrm{Tr}(\mathrm{Piz}_p(2)) = \sum_{j \in \mathbb{F}_{p^2} \mid j \text{ supersingular}} \frac{\#\{\phi \in \mathrm{End}(j) \mid \deg(\phi) = 2\}}{\#\,\mathrm{Aut}(j)}$$

Recall that a CM-j-invariant with corresponding CM-field K is supersingular modulo p if and only if p is inert in K. Recall also the *modular polynomials* defined above. We can factor $\Phi_2(X, X)$:

$$\begin{aligned}\Phi_2(X,X) &= -X^4 + 2978X^3 + 40449375X^2 + 17496000000X - 157464000000000\\ &= -(X+3375)^2(X-1728)(X-8000)\end{aligned}$$

As such, only the j-invariants -3375, 1728 and 8000 have endomorphisms of degree 2, and they have 2, 1 and 1 of these respectively. Over $\mathbb{C}$, they are CM-invariants, with corresponding fields $\mathbb{Q}(\sqrt{-7}), \mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-2})$ and discriminants -7,-4 and -8 respectively. Recall that a prime p is split in a quadratic number field of discriminant d if and only if $\left(\frac{d}{p}\right) = 1$.

Note that $\left(1 - \left(\frac{d}{p}\right)\right)$ is 0 if d is not a square mod p and 2 if it is (and 1 if $p \mid d$). Thus we get the trace formula:

$$\begin{aligned}\mathrm{Tr}\left(\mathrm{Piz}_p(2)\right) &= \left(1 - \left(\frac{-7}{p}\right)\right) + \frac{1}{2}\left(1 - \left(\frac{-4}{2}\right)\right) + \frac{1}{2}\left(1 - \left(\frac{-2}{p}\right)\right)\\ &= 2 - \left(\frac{-7}{p}\right) - \frac{1}{2}\left[\left(\frac{-4}{p}\right) + \left(\frac{-2}{p}\right)\right]\end{aligned}$$

We can use quadratic reciprocity to reduce the above formula to congruences mod 7,4 and 8 respectively. By the Chinese Remainder Theorem, these remainder properties can be combined into a requirement modulo $\mathrm{lcm}(4, 7, 8) = 56$.

□

Note that we are 'lucky' to be able to prove this theorem. It 'just so happens' that $\Phi_2(X, X)$ splits over $\mathbb{Z}[X]$ and the roots correspond to CM-curves. The trace formula from chapter 4 will allow us to get more results like this, though they will be harder to compute. Before we get to that, however, we should prove that we are actually interested in these curves, that is, that they are actually Ramanujan graphs.

3 The Ramanujan property

In this section, we will discuss a proof of first hypothesis, that is, that the Pizer graphs are actually Ramanujan graphs. The theorem we aim to prove is the following:

Theorem 3.1. *For all pairs of distinct primes l and p and every eigenvalue μ of $Piz_p(l)$ not equal to $l+1$, we have*

$$|\mu| \leq 2\sqrt{l}$$

Recall that $\mathrm{Piz}_p(l)$ is an $(l+1)$-regular graph for $l \neq p$, so $2\sqrt{l}$ is indeed the upper bound bound in the definition of Ramanujan graphs 2.7.

We will prove this by relating the Pizer graphs to Hecke operators on modular forms. We write $\mathcal{M}_k(N)$ (resp. $\mathcal{S}_k(N)$) for the modular (resp. cusp) forms of weight k on $\Gamma_0(N)$. The theorem above can be proven as a simple corrollary of the following theorem of Deligne (theorem 8.2 of [Del74]).

Theorem 3.2. *Let $l, N \in \mathbb{Z}_{\geq 1}$, l prime and $l \nmid N$. Write T_l for the l-th Hecke operator on $\mathcal{S}_k(N)$. For every eigenvalues λ of the l-th Hecke operator on $\mathcal{S}_k(N)$,*

$$|\lambda| \leq 2l^{\frac{k-1}{2}}$$

The proof of this general theorem uses techniques well beyond the scope of this thesis.However, to prove the first theorem, we do not need the full generality; we only need the case for weight $k = 2$. This proof is still not easy, however we *can* provide a sketch.

Before this, however, it would behoove us to explain why Theorem 3.2 implies Theorem 3.1. To this end, Section 3.1 defines the theta series (for some prime p) and compute the action of the Hecke operators on these series, which are elements of $\mathcal{M}_2(p)$.

The proof for the weight two case of Theorem 3.2 relies on the Riemann hypothesis for Abelian varieties:

Theorem 3.3. *Let A be an abelian variety defined over a finite field $\mathbb{F}_q$. Let σ_q be the q-power Frobenius map on A. Any eigenvalue λ of σ_q satisfies*

$$|\lambda| = \sqrt{q}.$$

In Section 3.2, we provide a sketch of the proof of this.

Finally, we need to relate the Hecke operators on modular forms to the Frobenius on abelian varieties. We use the Eichler–Shimura relation (which we do not prove) to conclude this in Section 3.3.

3.1 Pizer graphs and Hecke operators

The proof that the adjacency matrices of the Pizer graphs equal the Hecke operator, is not ultimately too enlightening. Fix a prime p. Let $E_1, \ldots, E_n$ be (representatives of isomorphism classes of) the supersingular curves over $\overline{\mathbb{F}}_p$. Recall the definition of $B(m)$; it is the matrix with entries $B_{ij}(m)$, where

$$B_{ij}(m) = \#\{C \subset E_i(\overline{\mathbb{F}}_p) \mid C \text{ a cyclic subgroup of order } m,\ E_i/C \simeq E_j\}.$$

we define $B_{ij}(m) = 0$ if m is not a non-negative integer. We define also $B(0)$ by defining

$$B_{ij}(0) = \frac{1}{\#\operatorname{Aut}(E_j)}.$$

Note that $B_{ij}(0)$ is $\frac{1}{2}$ unless $j(E_j) = 0$ or $j(E_j) = 1728$, in which case it is $\frac{1}{4}$ or $\frac{1}{6}$ respectively.

Let us write T_m for the m-th Hecke operator acting on $\mathcal{M}_2(p)$. One can verify that, for all $m \geq 0$,

$$\mathrm{Tr}(T_m) = \mathrm{Tr}(B(m)).$$

The details are beyond this thesis, but one can look at [Gro87], chapter 5.

Since $\dim(\mathcal{M}_2(p)) = n$ (see for instance [DS05], exercise 3.1.4 and theorem 3.5.1) and the T_m and $B(m)$ obey the same recurrence formula we know that inside $\mathrm{Mat}(n \times n, \mathbb{Q})$, the map $T_m \to B(m)$ is an isomorphism of subspaces, and thus there is a basis of $\mathcal{M}_2(p)$ such that $T_m = B(m)$ as matrices.

Rather than give an exact proof, then, we will try to convince the reader that this result is not unreasonable. To this end, we define the *theta series*.

Definition 3.4 (Theta series)**.** Write $q(z) = e^{2\pi i z}$. The i, j-th *theta series* is the function

$$f_{ij} : \mathcal{H} \to \mathbb{C},$$

given by the Fourier series

$$f_{ij}(z) = \sum_{m=0}^{\infty} B_{ij}(m) q(z)^m.$$

These are modular forms of weight two. Holomorphy is easy to prove; it suffices to prove that the above power series converges everywhere on $\mathcal{H}$ (i.e. that the f_{ij} are well-defined). But since (for $\mathrm{Im}(z) > 0$) $\left|e^{2\pi i z \cdot m}\right|$ is an exponentially decreasing function in m and $B_{ij}(m)$ is only polynomial in m (since the m-torsion is $(\mathbb{Z}/m\mathbb{Z})^2$, there are at most m^2 cyclic degree-m subgroups, so $B_{ij}(m) \leq m^2$), convergence is clear. Modularity by $\Gamma_0(N)$ is more complicated. One can look at [Ser73], section 6.5, though the notation here is quite different than the one used in this thesis.

Example 3.5. Let us compute the first few coefficients of the theta series over $p = 37$. To this end, we compute $B(0), B(1), B(2), B(3)$ and $B(4)$. Let us order E_1, E_2, E_3 as in Figure 2, i.e. such that $j(E_1) = 8$, $j(E_2) = 23 + 27a$ and $j(E_3) = 20 + 10a$ (where $a^2 = -4a + 2$ is such that $\mathbb{F}_{37}(a) = \mathbb{F}_{1369}$, as in Figure 2).

$B(0)$ and $B(1)$ are trivial:

$$B(0) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

$$B(1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$B(2)$ is simply the adjacency matrix of the graph in Figure 2.

$$B(2) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}.$$

Since $4 = 2^2$, $B_{ij}(4)$ is the number of paths of length 2 in the above-mentioned graph from E_i to E_j without backtracking.

$$B(4) = \begin{pmatrix} 0 & 3 & 3 \\ 3 & 2 & 1 \\ 3 & 1 & 2 \end{pmatrix}.$$

As expected, we find that $B(4) = B(2)^2 - 3 \cdot B(1)$.

Finally, for $B(3)$ we must compute the 3-torsion of E_1, E_2 and E_3. This is most easily done with a computer, and we find:

$$B(3) = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 0 & 3 \\ 1 & 3 & 0 \end{pmatrix}.$$

Filling this in the definition above, we get:

$$\begin{array}{rclcccccccccc}
f_{00}(z) & = & \frac{1}{2} & + & q(z) & + & q(z)^2 & + & 2q(z)^3 & & & + & \dots \\
f_{11}(z) & = & \frac{1}{2} & + & q(z) & & & & & + & 2q(z)^4 & + & \dots \\
f_{22}(z) & = & \frac{1}{2} & + & q(z) & & & & & + & 2q(z)^4 & + & \dots \\
f_{10}(z) = f_{01}(z) & = & \frac{1}{2} & & & + & q(z)^2 & + & q(z)^3 & + & 3q(z)^4 & + & \dots \\
f_{20}(z) = f_{02}(z) & = & \frac{1}{2} & & & + & q(z)^2 & + & q(z)^3 & + & 3q(z)^4 & + & \dots \\
f_{12}(z) = f_{21}(z) & = & \frac{1}{2} & & & + & 2q(z)^2 & + & 3q(z)^3 & + & q(z)^4 & + & \dots
\end{array}$$

Let us compute the action of the l-th Hecke operator T_l on f_{ij}:

$$(T_l f_{ij})(z) = \sum_{m=0}^{\infty} \left(B_{ij}(ml) + l B_{ij}(m/l) \right) q(z)^m.$$

Note that

$$B_{ij}(ml) + l B_{ij}(m/l) = \sum_{k=1}^{n} B_{ik}(l) B_{kj}(m).$$

This is a consequence of factoring and splitting isogenies; any lm-degree isogeny $\phi : E_i \to E_j$ can be split into a degree-m and degree-l part as we did in the proof of Proposition 2.10. We can now rearrange:

$$\begin{aligned}
(T_l f_{ij})(z) &= \sum_{m=0}^{\infty} (B_{ij}(ml) + l B_{ij}(m/l)) q(z)^m \\
&= \sum_{m=0}^{\infty} \left[\sum_{k=1}^{n} B_{ik}(m) B_{kj}(l) \right] q(z)^m \\
&= \sum_{k=1}^{n} B_{kj}(l) \sum_{m=0}^{\infty} B_{ik}(m) q(z)^m \\
&= \sum_{k=1}^{n} B_{kj}(l) f_{ik}(z).
\end{aligned}$$

The above shows that the space of theta series is invariant under the action of the Hecke operators. One can prove that the theta series span the entire space of modular forms, see for instance [Gro87], chapter 5. The space $\mathcal{M}_2(\Gamma_0(p))$ of modular forms of weight two over $\Gamma_0(p)$ has dimension n, as proven in chapter 3 of [DS05] (exercise 3.1.4 and theorem 3.5.1).

We would now like to state that there $i \in \{1, \dots, n\}$ such that $\{f_{i1}, \dots, f_{in}\}$. Regrettably, this is not known to be true; the computation of the dimension n_i of the space $\langle f_{i1}, \dots, f_{in} \rangle$ for some i is Hecke's basis problem, and is a major open problem.

Under presumption that there *is* an i such that $n_i = n$, i.e. such that $\{f_{i1}, \dots, f_{in}\}$ is a basis, the proof of the correspondence between Hecke operators and Pizer graphs becomes quick and constructive.

Theorem 3.6. *Let p be a prime. Presume that there is i such that $n_i = n$ as above. For every prime $l \neq p$, the matrix of T_l with respect to the basis $\{f_{i1}, \dots, f_{in}\}$ equals the adjacency matrix of $Piz_p(l)$.*

Proof. As computed, the action of T_l on f_{ij} is

$$T_l f_{ij} = \sum_{k=1}^{n} B_{kj}(l) f_{ik}(z).$$

Thus, with respect to the given basis, the r-th column of T_l is given by $(B_{kr}(l))_{k=1}^{n}$, which is of course the r-th column of the adjacency matrix of $\mathrm{Piz}_p(l)$. □

Now that we 'know' that the Hecke operators equal the adjacency matrix of our graphs, we almost understand why Theorem 3.2 implies that the Pizer graphs are Ramanujan. Note that the first-mentioned only investigates the *cusp forms*, rather than all modular forms. On the other hand, the Ramanujan property is a requirement on all eigenvalues except the one equal to $l + 1$.

As stated, we know that $\dim_{\mathbb{C}}(\mathcal{M}_2(p)) = n$. One can also prove that $\dim_{\mathbb{C}}(\mathcal{S}_2(p)) = n - 1$. We could now simply conclude that indeed our results line up: since T_l is the adjacency matrix of an $(l + 1)$-regular graph, it must have eigenvalue $l+1$. We know that its eigenvectors do not lie in $\mathcal{S}_2(p)$, since $2\sqrt{l} < l+1$ for all $l \neq 1$, thus the complement space of $\mathcal{S}_2(p)$, which is of dimension 1, must be the eigenspace of $l + 1$.

We can, however, be somewhat more constructive. The complement space is called the *Eisenstein space*. In our case, this space is one-of . A generator for this space is the Eisenstein series:

$$E_2^{(p)}(z) = \frac{p-1}{24} + \sum_{m=0}^{\infty} \sigma_1^{(p)}(m) q(z)^m.$$

Here we write

$$\sigma_1^{(p)}(n) = \sum_{d|n, p\nmid d} d.$$

For the construction of this series, see chapter 4 of [DS05] (specifically theorem 4.6.2).

Note that this series is normalized, in the sense that the coefficient before $q(z)^1$ is 1. One can prove that $E_2^{(p)}(z)$ is an eigenvector of T_l for primes $l \neq p$ (e.g. proposition 5.2.3 of [DS05]). The l-th coefficient of $E_2^{(p)}(z)$ is $\sigma_1^{(p)}(l) = l+1$, so indeed

$$T_l E_2^{(p)}(z) = (l+1) E_2^{(p)}(z).$$

As such, we now see that it suffices to prove Theorem 3.2. In order to accomplish this, we must take a short detour through the theory of abelian varieties.

3.2 The Riemann hypothesis for abelian varieties

We prove the Riemann hypothesis for Abelian varieties, following the proof and notation from [Mil08]. Recall that an abelian variety over K is a complete, connected variety over K together with regular maps

$$+ : A \times A \to A$$
$$- : A \to A$$

and a point e on A all defined over K such that $A(L)$ becomes a group with identity element e for any field extension $K \subset L$. One can prove that such a group structure is always abelian and that any rational map

$$\alpha : A \to B$$

of abelian varieties is a group morphism composed with a translation, and furthermore that α is defined on *all* points of A.

We quickly recap what we mean with eigenvalues of regular maps. Let A, B be (abelian) varieties defined over a field K of characteristic l. For primes $p \neq l$, any regular map $\alpha : A \to B$ of defines a morphism of p-adic Tate modules

$$\alpha : T_p(A) \to T_p(B).$$

Consider now an endomorphism $\alpha : A \to A$. We can define the characteristic polynomial P_α of α by taking its characteristic polynomial as an endomorphism of $T_p(A)$ (it is not immediately clear that this is independent of the choice of p, but this is proven in theorem 10.9 and proposition 10.20 of [Mil08]). The eigenvalues of α are then the roots of its characteristic polynomial (which lies in $\mathbb{Z}[X]$). From this, we can also define the trace discriminant, etc. of α.

For $a \in A$, we write t_a for the translation-by-a map:

$$t_a : A \to A, b \mapsto a + b.$$

Recall the notation

$$\mathrm{End}^0(A) = \mathrm{End}(A) \otimes \mathbb{Q}.$$

This is a division algebra, since for non-zero $\alpha \in \mathrm{End}(A)$ there is a dual map $\hat{\alpha}$ such that $\alpha\hat{\alpha} = \hat{\alpha}\alpha = [\deg(\alpha)]$. Putting $d = \deg(\alpha)$, we get:

$$(\alpha \otimes 1) \cdot \left(\hat{\alpha} \otimes \frac{1}{d}\right) = [d] \otimes \frac{1}{d} = 1.$$

We define a map that will be crucial in our proof of the Riemann hypothesis: the *Rosati-involution.*

$$(\alpha\beta)^\wedge = \beta^\wedge\alpha^\wedge.$$

Definition 3.7. Let A be an abelian variety over a field K and D an ample divisor on A. Consider the map

$$\phi_D : A \to \mathrm{Pic}^0(A)$$

given by

$$a \mapsto [t_a^* D - D].$$

The *Rosati-involution* (associated to D) is the map

$$\dagger : \mathrm{End}^0(A) \to \mathrm{End}^0(A), \alpha^\dagger = \phi_D^{-1} \circ \alpha^\wedge \circ \phi_D.$$

This has the following basic properties (for $\alpha, \beta \in \mathrm{End}^0(A)$):

1. $(\alpha + \beta)^\dagger = \alpha^\dagger + \beta^\dagger$.

2. $(\alpha\beta)^\dagger = \beta^\dagger\alpha^\dagger$.

3. $(\alpha^\dagger)^\dagger = \alpha$

4. If $\alpha \in \mathbb{Q}$, then $\alpha^\dagger = \alpha$.

As a consequence of the above, $\dagger$ defines an endomorphism of the ring $\mathbb{Q}[\alpha]$ for any $\alpha \in \mathrm{End}(E)$.

The most important property of the Rosati-involution for our purposes is that it is *positive-definite*. More precisely:

Theorem 3.8. *Let A be an abelian variety defined over a field K and let $\dagger$ be the Rosati-involution on $End^0(A)$. For any non-zero $\alpha \in End^0(A)$, we have*

$$Tr(\alpha \circ \alpha^\dagger) > 0.$$

The proof of this theorem relies on a direct computation of $\mathrm{Tr}(\alpha \circ \alpha^\dagger)$, see [Mil86], section 17. This computation uses methods of algebraic geometry beyond the scope of this thesis.

We are now ready to prove our theorem. Recall our stated goal:

Theorem 3.9. *Let A be an abelian variety defined over a finite field $\mathbb{F}_q$. Let σ_q be the q-power Frobenius map on A. Any eigenvalue λ of σ_q satisfies*

$$|\lambda| = \sqrt{q}.$$

Let us relate the eigenvalues of a morphism to its action under the Rosati involution: This is lemma 1.3 of chapter 2 of [Mil08].

Lemma 3.10. *Let A be a variety over some (not necessarily finite) field K and $\alpha \in End(A)$ such that $\alpha^\dagger\alpha$ is some integer r. Then any eigenvalue λ of α has*

$$|\lambda| = \sqrt{r}.$$

Proof. Note that $\mathbb{Q}[\alpha]$ has finite dimension as a vector space over $\mathbb{Q}$, as α is a root of its characteristic polynomial. It is thus an Artin ring. Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ be the (finitely many) maximal ideals of $\mathbb{Q}[\alpha]$. Then

$$\mathbb{Q}[\alpha]/\bigcap_{i=1}^{n} \mathfrak{m}_i = \prod_{i=1}^{n} \mathbb{Q}[\alpha]/\mathfrak{m}_i = \prod_{i=1}^{n} K_i.$$

Here, the quotients $K_i = \mathbb{Q}[\alpha]/\mathfrak{m}_i$ are fields. Recall that $\bigcap_{i=1}^n \mathfrak{m}_i$ is the set of nilpotent elements of $\mathbb{Q}[\alpha]$. We prove that this intersection is trivial.

Consider $a \in \mathbb{Q}[\alpha], a \neq 0$. Then $\mathrm{Tr}(a^\dagger a) > 0$, so $b = a^\dagger a \neq 0$. $b^\dagger = b$, so $b^2 = b^\dagger b \neq 0$ and by induction $b^{2^n} = (b^{2^{n-1}})^\dagger b^{2^{n-1}} \neq 0$. Thus b is not nilpotent and neither is a. Hence $\bigcap_{i=1}^n \mathfrak{m}_i = \{0\}$ and $\mathbb{Q}[\alpha]$ is the product of finitely many fields.

Any automorphism $\tau : \mathbb{Q}[\alpha] \to \mathbb{Q}[\alpha]$ permutes the maximal ideals, and hence the factors $\mathbb{Q}[\alpha]/\mathfrak{m}_i$. That is to say, there is a permutation σ of $\{1, \ldots, n\}$ and morphisms $\tau_i : K_i \to K_{\sigma(i)}$ such that

$$\tau((a_1, \ldots, a_n)) = (b_1, \ldots, b_n),$$

where

$$b_{\sigma(i)} = \tau_i(a_i).$$

For $\tau = \dagger$, this permutation must be trivial: if $\sigma(i) = j \neq i$, then the vector e_i with a 1 on the i-th component and 0 elsewhere, has $\mathrm{Tr}(e_i e_i^\dagger) = \mathrm{Tr}(e_i e_j) = \mathrm{Tr}(0) = 0$, but $e_i \neq 0$.

Consider $\mathbb{Q}[\alpha] \otimes \mathbb{R}$. The above results hold here also, however now each K_i is isomorphic to $\mathbb{R}$ or $\mathbb{C}$. Any automorphism τ of $\mathbb{Q}[\alpha]$ is thus a product of automorphisms of $\mathbb{R}$ or $\mathbb{C}$ of finite order. Over $\mathbb{R}$, there is only the identity morphisms, and over $\mathbb{C}$, there is the identity and complex conjugation. The identity morphism is not positive-definite, since $\mathrm{Tr}(i \cdot i) = -1 < 0$. Hence $\dagger$ must be complex conjugation on any complex part of $\mathbb{Q}[\alpha] \otimes \mathbb{R}$.

As such, for any homomorphism $\rho : \mathbb{Q} \to \mathbb{C}$, $\rho(\alpha^\dagger) = \overline{\rho(\alpha)}$. Hence, $r = \rho(r) = \rho(\alpha^\dagger \alpha) = |\rho(\alpha)|^2$. Since $\rho(\alpha)$ is an eigenvalue of α (as it is a root of P_α), and any such eigenvalue gives rise to a morphism $\mathbb{Q}[\alpha] \to \mathbb{C}$, any eigenvalue λ of α has $|\lambda| = \sqrt{r}$. □

We can now finish our proof!

Proof (of Theorem 3.9). By the lemma above, it suffices to prove that $\sigma_q^\dagger \sigma_q = q$. To this end, let D be the divisor defining $\dagger$ and ϕ_D as in Definition 3.7.

Note that σ_q^* is given by

$$\sigma_q^* : \mathrm{Pic}^0(A) \to \mathrm{Pic}^0(A), [D'] \mapsto [\sigma_q^* D'].$$

If $D' = \mathrm{div}(f)$, we get:

$$\sigma_q^* D' = \mathrm{div}(f \circ \sigma_q) = \mathrm{div}(f^q) = qD'.$$

Additionally, for any morphism $\alpha : A \to A$ and $a \in A(\overline{K})$, we have:

$$\begin{aligned} \alpha \circ t_a(x) &= \alpha(a + x) \\ &= \alpha(a) + \alpha(x) \\ &= t_{\alpha(a)} \circ \alpha(x). \end{aligned}$$

Hence, we get, for $a \in A(K)$:

$$\begin{aligned} \sigma_q^\dagger \circ \sigma_q &= \sigma_q^* \circ \lambda \circ \sigma_q \\ &= \sigma_q^* [D_{\sigma_q(a)} - D] \\ &= [\sigma_q^* t_{\sigma_q(a)}^* D - \sigma_q^* D] \\ &= [(t_{\sigma_q(a)} \circ \sigma_q)^* D - \sigma_q^* D] \\ &= [(\sigma_q \circ t_a) D - \sigma_q^* D] \\ &= [t_a^* q D - qD] \\ &= q\lambda(a). \end{aligned}$$

□

3.3 Eichler–Shimura and finishing the proof

In order to finish our proof of Theorem 3.2, we need the Eichler–Shimura relation. The proof, and even the precise statement, of this relation is beyond the scope of this thesis. For a complete discussion, one can look at chapters 6, 7 and 8 of [DS05] (theorem 8.7.2 for the main result, the remainder for an explanation and proof). As such, we shall give only a sketch of this relation.

Consider the Jacobian $A(p)$ of $X_0(p)$. The Hecke operators define an endomorphism of this variety. Its tangent space is $\mathcal{S}_2(p)$. That is, there is a Hecke-equivariant (natural) isomorphism

$$T_e(A(p)) \simeq \mathcal{S}_2(p).$$

For any prime $l \neq p$, $A(p)$ has good reduction at l. Finally, the Eichler–Shimura relation (in the form we need) states that the following diagram commutes:

$$\begin{tikzcd}
A(p) \arrow[r, "T_l"] \arrow[d, "\text{mod } l"] & A(p) \arrow[d, "\text{mod } l"] \\
\tilde{A}(p) \arrow[r, "\sigma_l+\hat{\sigma}_l"] & \tilde{A}(p)
\end{tikzcd}$$

Here σ_l is the l-power Frobenius morphism and $\tilde{A}(p)$ is the reduction of $A(p)$ mod l. Note that, as linear maps on the q-adic Tate module of $\tilde{A}(p)$ (for some prime $q \neq l$), we have

$$\hat{\sigma}_l \sigma_l = l(\sigma_l^{-1}\sigma_l) = l.$$

As such,

$$\hat{\sigma}_l = l\sigma_l^{-1}.$$

Consider an eigenvalue λ of T_l acting on $T_e(A(p)) = \mathcal{S}_2(p)$. One can prove, as a consequence of the Eichler–Shimura relation, that λ is also an eigenvalue of $\sigma_l + \hat{\sigma}_l$ and thus of $\sigma_l + l\sigma_l^{-1}$. Since σ_l and $l\sigma_l^{-1}$ have the same eigenvectors, $\lambda = \beta + l\beta^{-1}$, where β is an eigenvalue of σ_l.

By the Riemann hypothesis, we have $|\beta| = \sqrt{l}$ and hence $|\beta|^{-1} = \frac{1}{\sqrt{l}}$. Using the triangle inequality we get

$$\begin{aligned}
|\lambda| &= |\beta + l\beta^{-1}| \\
&\leq |\beta| + l|\beta|^{-1} \\
&= 2\sqrt{l}.
\end{aligned}$$

4 Distribution of eigenvalues

In Figure 2.3 we saw the eigenvalues of $\mathrm{Piz}_p(2)$ seemed to follow some distribution as $p \to \infty$. In fact, there is a family of functions f_l such that the eigenvalues of $\mathrm{Piz}_p(l)$ are distributed according to a distribution with density function f_l (as $p \to \infty$).

This f_l appears not only as the distribution for our graphs, but is the defining distribution of eigenvalues for many families k-regular graphs. This is the main theorem of [McK81]:

Theorem 4.1. *Let $k \geq 2$ be an integer and let $X_1, X_2, \ldots$ be a family of k-regular graphs, such that the order $\#X_m \to \infty$ as $m \to \infty$. Let $F(X_i, x)$ be the cumulative distribution function of the eigenvalues of X_i. That is, if X_i has order $\#X_i = N_i$, then*

$$F(X_i, x) = \frac{1}{N_i} \cdot \#\{\lambda \in Eig(X_i) \mid \lambda \leq x\}.$$

Here $Eig(X_i)$ is the multiset of eigenvalues of X_i, counting is counting multiplicities.

Define $c_n(X_m)$ to be the number of cycles (closed walks without backtracking) of length n in X_m. If for all $n \geq 0$,

$$\lim_{m\to\infty} \frac{c_n(X_m)}{\#X_m} \to 0,$$

then, for all x,

$$\lim_{i\to\infty} F(X_i, x) = \int_{-2\sqrt{k-1}}^{x} f_k(t)dt$$

where $f_k : \mathbb{R} \to \mathbb{R}$ is defined as

$$f_k(t) = \begin{cases} \frac{k\sqrt{4(k-1)-t^2}}{2\pi(k^2-t^2)} & |t| \leq 2\sqrt{k-1}. \\ 0 & \textit{else.} \end{cases}$$

Remark. Note that '$c_n(X_m)/\#X_m \to 0$' means that the number of n-cycles in X_m grows sublinearly with the size of X_m. In particular, if the graphs have girth going to ∞, then $c_n(X_m) \to 0$ as $m \to \infty$ for any n, so graph families with increasing girth will in the limit follow this distribution. The proof of this theorem uses only graph theory.

Let us discuss its core ideas. Recall that a distribution is defined by the expected value of its moments (this is explained in the proof of Lemma 4.7). The r-th moment of the distribution of eigenvalues of a linear operator A is the sum of the r-th power of its eigenvalues; if A has eigenvalues $\lambda_1, \ldots, \lambda_n$, then the r-th moment is

$$\lambda_1^r + \ldots + \lambda_n^r = \mathrm{Tr}(A^r).$$

As such, to understand the distribution of eigenvalues of a matrix is to understand the traces of powers of this matrix. If A is the adjacency matrix of some graph G, then $\mathrm{Tr}(A^r)$ is the number of closed walks of length r in G (allowing backtracking), and hence we can determine the distribution of the eigenvalues of G by counting this number of walks. Instrumental in the proof of the above theorem is the following result (lemma 2.1 in McKay's paper):

Lemma 4.2. Let $G = (V, E)$ be a finite, k-regular graph and $r \in \mathbb{Z}_{\geq 1}$. Suppose $v \in V$ is such that the subgraph of G consisting of vertices distance at most $r/2$ away from v_0 contains no cycles. Then the number of closed walks of length r in X starting (and ending) in v is $\theta(r)$, where $\theta(r) = 0$ if r is odd and

$$\theta(2s) = \sum_{i=1}^{s} \binom{2s-i}{s} \frac{i}{2s-i} k^i (k-1)^{s-i}.$$

The proof of this lemma is not so hard (as the subgraph of vertices distance at most $r/2$ from v_0 is isomorphic to a subgraph of the k-regular tree, and any walks of distance r must remain within distance $r/2$ of v_0), but its impacts are profound.

For $r = 0$, there is exactly one closed circuit from v to v, so we should put $\theta(0) = 1$. Since in Theorem 4.1, the number of r-circuits without backtracking in X_i grows sublinearly, one can prove that the density of vertices of X_i for which the conditions of the above lemma are satisfied goes to 1 as $i \to \infty$. As such, we get:

$$\lim_{i\to\infty} \int x^r dF_i = \theta(r).$$

McKay proves that there is a function F such that

$$\int x^r dF = \theta(r)$$

for each r and constructs this function $F(x)$, which is indeed

$$F(x) = \int_{-2\sqrt{k-1}}^{x} f_k(t)dt.$$

We shall see that our graphs do *not* have sublinear growth of k-cycles (see the Trace Formula section below). As such, we cannot use McKay's proof here. The statement we will prove is the theorem below. Recall that $B_p(l)$ is the adjacency matrix of $\mathrm{Piz}_p(l)$.

Theorem 4.3. *Let $p, l \in \mathbb{Z}_{\geq 1}$ be distinct primes. Write $B'_p(l) = B_p(l)/\sqrt{l}$ and define*

$$F_p(x,l) = \frac{1}{\#\, Eig\left(B'_p(l)\right)} \#\{\lambda \in Eig\left(B'_p(l)\right) \mid \lambda \leq x\}.$$

Then, as $p \to \infty$, the (non-trivial) eigenvalues of $B'_p(l)$, which lay in [-2,2], are distributed according to the density function

$$f(x) = \frac{l(l+1)\sqrt{4-x^2}}{2\pi((l+1)^2 - lx^2)}.$$

The distributions in Theorem 4.1 and Theorem 4.3 are related by a change of variables. Note that the support of the density function in Theorem 4.1 is the interval $[-2\sqrt{k-1}, 2\sqrt{k-1}]$ whilst in 4.3 it is $[-2, 2]$. Recalling that $\mathrm{Piz}_p(l)$ is $(l+1)$-regular, we should put $k = l+1$ and thus 'compress' the first interval by replacing x with $\frac{x}{\sqrt{l}}$. Indeed:

$$\begin{aligned} f_{l+1}\left(\frac{x}{\sqrt{l}}\right) &= \frac{(l+1)\sqrt{4l - lx^2}}{2\pi((l+1)^2 - lx^2)} \\ &= \frac{\sqrt{l}(l+1)\sqrt{4-x^2}}{2\pi((l+1)^2 - lx^2)} \\ &= \frac{1}{\sqrt{l}} f(x). \end{aligned}$$

so by the chain rule

$$\int_{a\sqrt{l}}^{b\sqrt{l}} f_{l+1}(x)dx = \int_a^b f_{l+1}\left(\frac{x}{\sqrt{l}}\right) \cdot \frac{1}{\sqrt{l}} dx = \int_a^b f(x)dx.$$

Hence the eigenvalues of our Pizer graphs are indeed distributed in the same way as the regular graphs in McKay's paper.

Let us briefly discuss the structure of the proof we give. We follow the structure of [Ser97]. This method has the same strategy as that of McKay. We compute the trace of *polynomial* expressions of our adjacency matrices. McKay does this by estimating the number of length-n walks for all $n \geq 0$. Our graphs do not quite permit this method. Instead, we estimate the number of *cycles* of length n, and we use this result to compute the distribution.

Explicitly, for all $n \geq 0$, we find a polynomial $P_n \in \mathbb{Z}[X]$ of degree n such that $P_n(B'_p(l)) = B'_p(l^n)$ for primes $l \neq p$ and $n \in \mathbb{Z}_{\geq 0}$, and relate these to the $f(x)$ above.

Using the theory of quaternion algebras, we compute the Eichler–Selberg trace formula for the Brandt matrices. Serre uses the general version, for Hecke operators acting on modular forms of general weight and general character. We are able to provide a sketch of the proof of our specific case, and because we use a more specific and hence shorter formula, our estimations are easier, but the methods are the same.

We finish our proof by computing a limit, specifically

$$\lim_{p \to \infty} \frac{\operatorname{Tr}(B'_p(l^m))12}{p-1}.$$

This, together with the measure theory above, proves our result. Note the similarities in Serre's method and McKay's method. Both rely on the fact that in order to know a distribution, one needs only study its moments. McKay can immediately use this, together with the result on the number of closed walks in our graphs, to conclude that there is some distribution from which the eigenvalues are drawn. He then constructs this distribution using the *Chebyshev polynomials*. Serre does this also, though he uses the normalised version; the P_n is the normalised version of the n-th Chebyshev polynomial.

4.1 Measure theory of eigenvalues

In this section, we relate the distribution of the traces of moments of linear operators to the distribution of their eigenvalues. To do so, we briefly discuss some theory of measures on topological spaces. We presume the reader is familiar with the basics of measure theory. Recall the definition of a Radon measure:

Definition 4.4. A *Radon Measure* is a measure μ on (the Borel-σ-algebra of) a Hausdorff topological space X such that:

- For all $x \in X$ there is a neighbourhood U of x such that $\mu(U) < \infty$
- For all open sets $U \subset X$, $\mu(U) = \sup\{\mu(K) \mid K \subset U \text{ is compact}\}$

We can use such a measure to define a linear map

$$C(X, \mathbb{R}) \to \mathbb{R}$$

given by

$$f \mapsto \int f(x)\mu(x).$$

$C(X, \mathbb{R})$ here denotes the continuous functions from X to $\mathbb{R}$. If $f(x) \geq 0$ for all $x \in X$, then $\int f(x)\mu(x) \geq 0$. We will write

$$\langle f, \mu \rangle = \int f(x)\mu(x).$$

For finite non-empty sets $S \subset X$, we write

$$\delta_S = \frac{1}{\#S} \sum_{s \in S} \delta_s.$$

Here δ_s is the Dirac measure supported on s. Explicitly,

$$\langle f, \delta_S \rangle = \frac{1}{\#S} \sum_{s \in S} f(s).$$

We can now precisely state what we mean when we say that a sequence of values are *distributed* according to a measure (or distribution).

Definition 4.5 (Equidistribution)**.** Let X be a Hausdorff space and let $S_1, S_2, \ldots$ be a sequence of finite non-empty subsets of X. We say that the S_i are *μ-equidistributed* or *(evenly) distributed according to μ*, for some Radon measure μ on X if, in the space of measures with weak topology, we have

$$\lim_{i\to\infty} \delta_{S_i} = \mu.$$

Explicitly, for all functions $f \in C(X, \mathbb{R})$, we must have

$$\lim_{i\to\infty} \frac{1}{\# S_i} \sum_{s\in S_i} f(s) = \langle f, \mu \rangle \, .$$

The fact that this is indeed an appropriate definition is the subject of the following lemma:

Lemma 4.6. *Let X and μ be as above and let $S_1, S_2, \ldots$ be subsets of X evenly distributed according to μ. Let also $A \subset X$ be given such that its boundary ∂A is measurable and has $\mu(\partial A) = 0$. Then*

$$\lim_{i\to\infty} \frac{\#(S_i \cap A)}{\# S_i} = \mu(A).$$

For a more general result and proof, see theorem III.1.3 of [Bou04]. This lemma should intuitively make sense, since by definition, we have:

$$\lim_{i\to\infty} \frac{1}{\# S_i} \cdot \#(S_i \cap A) = \lim_{i\to\infty} \frac{1}{\# S_i} \sum_{s\in S_i} \mathbb{1}_A(s) = \langle \mathbb{1}_A, \mu \rangle = \mu(A).$$

We will now apply the above concepts to the distribution of eigenvalues. As such, consider for $i \in \mathbb{Z}_{\geq 1}$ some linear map $H_i \in \mathrm{Mat}(\mathbb{R}, n_i \times n_i)$, and suppose that all H_i have eigenvalues all lying in some interval $\Omega = [a, b] \subset \mathbb{R}$. Define S_i to be the set of eigenvalues of H_i. We then have the following proposition:

Lemma 4.7. *In the situation above, the following are equivalent:*

1. *The S_i are μ-equidistributed (on Ω)*
2. *For all polynomials $P(X) \in \mathbb{R}[X]$, we have*
$$\lim_{i\to\infty} \frac{Tr(P(H_i))}{n_i} = \langle P, \mu \rangle$$
3. *For all $m \geq 0$, there is a polynomial $P_m(X) \in \mathbb{R}[X]$ with $\deg(P_m) = m$ and*
$$\lim_{i\to\infty} \frac{Tr(P_m(H_i))}{n_i} = \langle P_m, \mu \rangle$$

Proof. Recall that if A is a finite-dimensional matrix, over some field K, with eigenvalues $\lambda_1, \ldots, \lambda_n$ and $P \in K[X]$, then $\mathrm{Tr}(P(A)) = \sum_{i=1}^{n} P(\lambda_1)$.

The implication $1 \implies 2$ is immediate, as $\mathbb{R}[X] \subset C(\Omega, \mathbb{R})$. Furthermore, by the Stone-Weierstrass theorem, $\mathbb{R}[X]$ is dense in the set of continuous functions, which proves $2 \implies 1$.

The equivalence of 2 and 3 is a consequence of the fact that such P_m's give a $\mathbb{R}$-basis for $\mathbb{R}[X]$, and taking the trace and integration are both $\mathbb{R}$-linear operations. □

To relate all this to the proof, the H_i's are our (normalized) adjacency matrices, the measure is

$$\mu_l(x) = \frac{l(l+1)\sqrt{4-x^2}}{2\pi((l+1)^2 - lx^2)}dx$$

It is then our task to find polynomials P_m for $m \in \mathbb{Z}_{\geq 0}$ such that we understand $P_m(B_p(l))$. We will first define related matrices, and compute a trace formula for these matrices.

4.2 Our polynomials and measures

We give the measures and polynomials we shall use in the final proof of this theorem, and give the necessary results. This is taken from section 2 of [Ser97]. Central in our construction is the observation that the $B_p(m)$ satisfy a recurrence relation: if $l \neq p$ is a prime, then for all $k \geq 2$

$$B_p(l^k) = B_p(l)B_p(l^{k-1}) - l \cdot B_p(l^{k-2})$$

To see this, note that these matrices count isogenies of degree l^k with cyclic kernel. These correspond to walks in $\mathrm{Piz}_p(l)$ of length k without backtracking, which implies this relation (see Lemma A.4).

Recall that we write

$$B'_p(m) = \frac{B_p(m)}{\sqrt{m}}.$$

These matrices then satisfy the relation

$$B'_p(l^n) = B'_p(l)B'_p(l^{n-1}) - B'_p(l^{n-2}).$$

As such, if we define $P_0 = 1, P_1 = X$ and $P_n = X \cdot P_{n-1} - P_m$ for $n \geq 2$, we see that

$$P_n(B'_p(l)) = B'_p(l^n).$$

The P_n are monic polynomials in $\mathbb{Z}[X]$. They are the normalized Chebyshev polynomials of the second kind (see for instance definition 1.4.4 of [DSV03]) defined by the equality

$$P_n(2\cos(\phi)) = \frac{\sin((n+1)\phi)}{\sin(\phi)}.$$

We give the first few such polynomials:

$$\begin{aligned}
&P_0(X) = 1 \\
&P_1(X) = X \\
&P_2(X) = X^2 - 1 \\
&P_3(X) = X^3 - 2X \\
&P_4(X) = X^4 - 3X^2 + 1 \\
&P_5(X) = X^5 - 4X^3 + 3X
\end{aligned}$$

Their generating series is

$$\sum_{n=0}^{\infty} P_n(x)t^n = \frac{1}{1 - xt + t^2}.$$

By the recurrence relation, they satisfy

$$P_m \cdot P_n = P_{m+n} + P_{m+n-2} + \ldots + P_{|m-n|+2} + P_{|m-n|}.$$

Define the measures

$$\mu_l = \frac{l(l+1)\sqrt{4-x^2}}{2\pi((l+1)^2 - lx^2)} dx$$

for $l > 1$ and

$$\mu_\infty = \lim_{l\to\infty} \mu_l = \frac{\sqrt{4-x^2}}{2\pi} dx.$$

Note that $\mu_l = \frac{l(l+1)}{(l+1)^2 - lx^2} \mu_\infty$.

We now relate the μ_l and the P_n. From the generating series of the P_n, we get that

$$\sum_{n=1}^{\infty} P_{2n} t^{-n} = \frac{t(t+1)}{(t+1)^2 - tx^2}.$$

Thus

$$\mu_l = \mu_\infty \cdot \sum_{n=1}^{\infty} P_{2n} l^{-n}.$$

We compute

$$\langle P_n, \mu_\infty \rangle = \int_{-2}^{2} P_n(x) \frac{\sqrt{4-x^2}}{2\pi} dx.$$

Applying the change of variables $x = 2\cos(\phi)$, this becomes

$$\frac{1}{\pi} \int_0^\pi P_n(2\cos(\phi)) 2\sin^2(\phi) d\phi = \frac{2}{\pi} \int_0^\pi \sin((m+1)\phi)\sin(\phi) d\phi = \begin{cases} 1 & m = 0, \\ 0 & m > 0. \end{cases}$$

As such $\langle P_m P_n, \mu_\infty \rangle = \mathbb{1}_{m=n}$. We get

$$\langle P_m, \mu_l \rangle = \sum_{n=0}^{\infty} \langle P_m P_{2n}, \mu_\infty \rangle l^{-n} = \sum_{n=0}^{\infty} \mathbb{1}_{m=2n} l^{-n} = \begin{cases} \frac{1}{l^{m/2}} & m \text{ is even}, \\ 0 & m \text{ is odd}. \end{cases}$$

4.3 Brandt matrices and their trace

In order to finish our proof, we need a formula on the traces of the adjacency matrices of our graphs. This we do by their relation to quaternion algebras. This subsection is mostly taken from [Gro87].

4.3.1 Brandt matrices

Let us start by recalling the quaternionic definition of the Brandt matrices. Fix some prime p and let B be the (unique up to isomorphism) quaternion algebra over $\mathbb{Q}$ ramified at p and ∞ (and nowhere else). Let $R \subset B$ be a maximal order and consider the following equivalence relation on left-ideals of R:

$$I \sim J \iff \text{there is } b \in B^* \text{ such that } J = Ib.$$

The set of left ideal classes (modulo the relation above) is finite and independent of the choice of maximal order R. Let $\{I_1, \ldots, I_n\}$ be a set of representatives of left ideal classes, chosen such that $I_1 = R$.

Write

$$R_i = \{b \in B \mid I_i b \subset I_i\}.$$

for the *right order* of I_i. This is a maximal order of B and each maximal order of B is represented in the set $\{R_1, \ldots, R_n\}$, up to conjugacy.

The groups $\Gamma_i = R_i^* / \langle \pm 1 \rangle$ are finite and thus it makes sense to define

$$w_i = \#\Gamma_i$$

Their product $W = \prod_{i=1}^n w_i$ is independent of the choice of R and equals the denominator of $\frac{p-1}{12}$. As for their sum, we have the following general result:

Theorem 4.8 (Eichler Mass Formula). *Let B be a quaternion algebra over $\mathbb{Q}$ of discriminant D and $O \subset B$ a maximal order. Let $\{I_1, \ldots, I_r\}$ be representatives the left ideal classes of O, R_i the right order of I_i and $w_i = \#(R_i/\{\pm 1\})$. Then*

$$\sum_{i=1}^{r} \frac{1}{w_i} = \frac{\varphi(D)}{12}$$

Here $\varphi(D) = \#(\mathbb{Z}/D\mathbb{Z})^*$ is the Euler-totient function.

This is theorem 25.1.1 in [Voi21]. Its general proof is beyond the scope of this thesis, but for $D = p$ a prime, this is actually equivalent to the following (exercise 5.9 of [Sil09], an easy consequence of theorem V.4.1(c)).

Proposition 4.9. *Let p be a prime. If $E_1, \ldots, E_n$ are the supersingular curves over $\mathbb{F}_{p^2}$, then*

$$\sum_{i=1}^{n} \frac{1}{\# \mathit{Aut}(E_i)} = \frac{p-1}{24}$$

We define also

$$M_{i,j} = I_j^{-1} I_i = \left\{ \sum_k a_k \cdot b_k \mid a_k \in I_i^{-1}, b_k \in I_j \right\}.$$

Here I_i^{-1} is the inverse ideal of I_i:

$$I^{-1} = \{b \in B \mid IbI \subset I\}.$$

We write $N(b)$ for the norm of an element $b \in B$. We also define a norm on the set of $M_{i,j}$ by letting $N(M_{i,j})$ be the positive rational number such that $\frac{N(b)}{N(M_{i,j})}$ is an integer for all $b \in \mathcal{M}_{i,j}$ and

$$\gcd\left(\left\{ \frac{N(b)}{N\left(M_{i,j}\right)} \mid b \in M_{i,j} \right\} \right) = 1.$$

For $b \in M_{i,j}$, write

$$N_{i,j}(b) = \frac{N(b)}{N(M_{i,j})}.$$

Define

$$B_{ij}(m) = \#\{b \in M_{i,j} \mid N_{i,j}(b) = m\} \cdot \frac{1}{2w_j}.$$

This is the number of element in $M_{i,j}$ of norm m up to multiplication with units (since $w_j = \frac{1}{2}\#R_j^*$).

We now have the tools to define the Brandt matrices.

Definition 4.10 (Brandt matrices). Let $p, m \in \mathbb{Z}_{\geq 0}$ be given, p prime. Let the $B_{ij}(m)$ be given as above. The m-th Brandt matrix of p is the matrix.

$$B_p(m) = \begin{pmatrix} B_{11}(m) & \ldots & B_{n1}(m) \\ \vdots & \ddots & \vdots \\ B_{1n} & \ldots & B_{nn}(m) \end{pmatrix}.$$

Remark. To see why these matrices are the adjacency matrices of our Pizer graphs, recall that there are exactly n supersingular elliptic curves over $\mathbb{F}_{p^2}$ up to isomorphism, corresponding to the maximal orders of B up to conjugation. We can now choose representatives of curves $E_1, \ldots, E_n$ such that

$$\operatorname{End}(E_i) \simeq R_i.$$

There is then a bijection between isogenies $\phi : E_i \to E_j$ and elements of $M_{i,j}$, and if ϕ corresponds to b, then $\deg(\phi) = N_{i,j}(b)$. As such, if we consider only isogenies up to conjugation with isomorphisms, we must divide by $\frac{1}{\# \operatorname{Aut}(E_j)}$, and $\# \operatorname{Aut}(E_j) = \# R_j^* = 2w_j$, so we see that the definition above yields the same matrices as Definition 2.9.

4.3.2 Optimal embeddings and class numbers

In the proof of our trace formula, we will need the theory of embeddings of imaginary quadratic orders into quaternion algebras. As this is not standard, we give a short overview. This is discussed in the proof of proposition 1.9 of Gross, but the discussion in chapter 30 of [Voi21] is more complete (both than Gross' discussion and ours). We start with the definition of embeddings.

Definition 4.11. [Embedding] Let O be an order in a number field K and R a maximal order in a quaternion algebra B over $\mathbb{Q}$. An *embedding* of O into R is an injective ring morphism

$$\psi : O \to R.$$

An embedding is called *optimal* if, for the induced map

$$\psi : K \to B,$$

we have

$$\psi(K) \cap R = \psi(O).$$

We write $\operatorname{Emb}(O, R)$ for the set of embeddings, and $\operatorname{OptEmb}(O, R)$ for the set of *optimal* embeddings, of O into R.

Note that if an embedding $\psi : O \to R$ (for some order $O \subset K$) is not optimal, there is a unique order $O' \subset K$ containing O such that the induced embedding $\psi : O' \to R$ *is* optimal.

We are interested in the case that K is quadratic and imaginary and B is the quaternion algebra ramified at p and ∞ as above. In this case, since orders are defined uniquely by their discriminants and, by the above, embeddings are uniquely determined by the order at which they are optimal, we have

$$|\operatorname{Emb}(O_D, R)| = \sum_{df^2 = D} |\operatorname{OptEmb}(O_d, R)|.$$

The sum here is over divisors $d \mid D$ such that $\frac{D}{d}$ is a square.

Furthermore, writing $K = \mathbb{Q}(\sqrt{-d})$, there is an embedding

$$K \to B$$

if and only if there is $b \in B$ such that $b^2 = -d$, which is the case if and only if

$$\operatorname{Tr}(b) = 0 \text{ and } N(b) = -d.$$

More generally, if $O = \mathbb{Z}[\alpha]$ is an order in K and R some (maximal) order in B, then there is an embedding

$$O \to B$$

if and only if there is $r \in R$ such that

$$\mathrm{Tr}(r) = \mathrm{Tr}(\alpha), N(r) = N(\alpha).$$

Any such r defines an embedding of O into R.

There is an action of $\Gamma = R^*/\{\pm 1\}$ on $\mathrm{Emb}(O, R)$, given by:

$$(\gamma\psi)(z) = \gamma^{-1}\psi(z)\gamma.$$

If ψ is optimal, so is $\gamma\psi$, so this also induces an action on $\mathrm{OptEmb}(O, R)$.

The reason we are interested in these embeddings is their relation to class groups. For $d < 0$, write $h(d)$ for the class number of O_d, the order of discriminant d in an imaginary quadratic number field. As in the previous discussion, let $R_1, \ldots, R_n$ be the maximal orders of B and $\Gamma_i = R_i^*/\{\pm 1\}$. The following is a result of Eichler, see theorem 30.4.7 of [Voi21]:

Theorem 4.12. *Let B, the R_i and Γ_i be as above. Define (for $d < 0$)*

$$h_i(d) = \#[\Gamma_i \backslash \mathit{OptEmb}(O_d, R_i)].$$

(That is, $h_i(d)$ is the number of optimal embeddings of O_d into R_i up to conjugation with R_i^*). Then we have

$$\sum_{i=1}^{n} h_i(d) = \begin{cases} (1 - \left(\frac{d}{p}\right)) \cdot h(d) & \text{if } p^2 \nmid d, \\ 0 & \text{if } p^2 \mid d. \end{cases}$$

4.3.3 The Eichler–Selberg trace formula

The trace formula of the Brandt matrices we will give is a sum of *Hurwitz class numbers*. The definitions of these numbers are lifted from Gross, see chapter 1 of [Gro87] (the part just before the trace formula, proposition 1.9)

Definition 4.13. For $d < 0$, let O_d be the order of discriminant d and rank 2 over $\mathbb{Z}$. We define:

$$\begin{aligned} h(d) &= \#\mathrm{Cl}(O_d), \\ u(d) &= \#(O_d^*/\{\pm 1\}). \end{aligned}$$

Let $D \in \mathbb{Z}_{\geq 1}$ be given. The *Hurwitz Class Number* $H(D)$ is given by

$$H(D) = \sum_{df^2 = -D} \frac{h(d)}{u(d)}.$$

The sum here is over all divisors d of $-D$ such that $\frac{-D}{d}$ is a square.

Note that $u(d) = 1$ for all d except $d = -3$ or $d = -4$, when it is 3 or 2 respectively.

Example 4.14. We compute $H(D)$ for some small D:

$$\begin{aligned} H(3) &= \tfrac{h(-3)}{u(-3)} && = \tfrac{1}{3}, \\ H(4) &= \tfrac{h(-1)}{u(-1)} + \tfrac{h(-4)}{u(-4)} && = \tfrac{1}{2}, \\ H(5) &= \tfrac{h(-5)}{u(-5)} && = 0, \\ H(6) &= \tfrac{h(-6)}{u(-6)} && = 0, \\ H(7) &= \tfrac{h(-7)}{u(-7)} && = 1. \end{aligned}$$

Let p be a prime number. We define the modified Hurwitz class numbers $H_p(D)$ as follows:

$$H_p(D) = \begin{cases} 0 & p \text{ splits in } O_{-D}. \\ H(D) & p \text{ remains inert in } O_{-D}. \\ \frac{1}{2}H(D) & p \text{ ramifies in } O_{-D}, \text{ but does not divide its conductor.} \\ H_p(\frac{D}{p^2}) & p \text{ ramifies in } O_{-D} \text{ and does divide its conductor.} \end{cases}$$

Also define

$$H_p(0) = \sum_{i=1}^{n} \frac{1}{2w_i} = \frac{p-1}{24}.$$

We can now finally give our trace formula. This is proposition 1.9 from [Gro87].

Theorem 4.15. *For all $m \geq 0$, we have*

$$Tr(B_p(m)) = \sum_{s^2 \leq 4m} H_p(4m - s^2).$$

The sum here is over integers s of any sign.

Proof. For $m = 0$, this is the mass formula (Theorem 4.8). We assume $m \geq 1$.

$B_{ii}(m)$ is the number of elements in R_i with norm m modulo multiplication with units in R_i^*. Recall that $\#R_i^* = 2w_i$. Define the set

$$A_i(s, m) = \{b \in R_i \mid \mathrm{Tr}(b) = s, N(b) = m\}.$$

This set is finite, and if $s^2 > 4m$, it is empty: the discriminant of an element b in R_i is negative and equals $\mathrm{Tr}(b)^2 - 4N(b)$.

Thus

$$\mathrm{Tr}(B_p(m)) = \sum_{i=1}^{n} \sum_{s^2 \leq 4m} \frac{\#A_i(s, m)}{\#R_i^*}.$$

By swapping indices, we may rewrite this to

$$\sum_{s^2 \leq 4m} \sum_{i=1}^{n} \frac{\#A_i(s, m)}{\#R_i^*}.$$

It thus suffices to prove that

$$\sum_{i=1}^{n} \frac{\#A_i(s, m)}{\#R_i^*} = H_p(4m - s^2).$$

If $4m - s^2 = 0$, this again follows from the mass formula. Presume then that

$$D = 4m - s^2 > 0.$$

As mentioned in Section 4.3.2, any element $b \in A_i(s,m)$ gives an embedding of the order O_{-D} into R_i. Furthermore, if $O_{-D} = \mathbb{Z}[\alpha]$ and we have an embedding

$$\psi : O_{-D} \to R_i,$$

then $\psi(\alpha) \in A_i(s,m)$ and $(\gamma\psi)\alpha = \gamma^{-1}\psi(\alpha)\gamma$. Thus if we let Γ_i act on $A_i(s,m)$ by conjugation, the orbits of this action correspond to embeddings $O_{-D} \to R_i$ up to the action of Γ_i. In other words:

$$\#[\Gamma_i \backslash A_i(s,m)] = \#[\Gamma_i \backslash \operatorname{Emb}(O_{-D}, R_i)] = \sum_{df^2=-D} h_i(d).$$

Now, the stabiliser of an element $b \in A_i(s,m)$ is trivial, unless the corresponding embedding of O_{-D} (i.e. the one where $\psi(\alpha) = b$) extends to $\mathbb{Z}[i]$ or $\mathbb{Z}[\zeta_6]$ (writing $\zeta_n = e^{2\pi i/n}$), in which case this stabiliser has order 2 or 3 respectively. This corresponds with our definition of the $u(d)$, and we find:

$$\#A_i(s,m) = w_i \sum_{df^2=-D} \frac{h_i(d)}{u(d)}$$

Combining this with Theorem 4.12 above, we get our desired result:

$$\begin{aligned}
\operatorname{Tr}(B_p(m)) &= \sum_{i=1}^{n} \sum_{s^2 \leq 4m} \frac{\#A_i(s,m)}{\#R_i^*} \\
&= \sum_{s^2 \leq 4m} \sum_{i=1}^{n} \frac{\#A_i(s,m)}{\#R_i^*} \\
&= \sum_{s^2 \leq 4m} H_p(4m - s^2).
\end{aligned}$$

□

4.4 Finishing the proof

To finish our proof, we must show that, for all $n \geq 0$,

$$\lim_{p\to\infty} \frac{\operatorname{Tr}(P_n(B_p'(l)))}{\dim(\mathcal{S}_2(p))} = \langle P_n, \mu_l \rangle .$$

Since $P_n(B_p'(l)) = B_p'(l^n)$, we can use our trace formula. We will prove the following:

Lemma 4.16. *For all $m \geq 0$, we have:*

$$\lim_{p\to\infty} \frac{12\, Tr(B_p'(m))}{p-1} = \begin{cases} \frac{1}{\sqrt{m}} & m \textit{ is a square.} \\ 0 & \textit{else.} \end{cases}$$

Note that l^n is a square if and only if n is even, so this is the result we desire. Furthermore, taking $m = 1$ in the above, $B_p'(1) = B_p(1)$ is the identity morphism on $\mathcal{S}_2(p)$ and thus $\operatorname{Tr}(B_p(1)) = \dim(\mathcal{S}_2(p))$. As such,

$$\lim_{p\to\infty} \frac{12 \dim(\mathcal{S}_2(p))}{p-1} = 1$$

and we really measure the *proportion* of eigenvalues. Finally, since $(B_p(l^k))$ counts the number of paths of length k without backtracking in $\operatorname{Piz}_p(l)$, $\operatorname{Tr}(B_p(l^k))$ is the number of *closed* walks of length k without backtracking. We now see that this number does not grow sublinearly, but rather linearly, in p:

$$\lim_{p\to\infty}\frac{\operatorname{Tr}(B_p(l^k))}{p}=\begin{cases}\frac{1}{12} & k \text{ is even.}\\ 0 & \text{else.}\end{cases}$$

Let us continue with the proof.

Proof (of Lemma 4.16). Since $B'_p(m)=\frac{1}{\sqrt{m}}B_p(m)$, it follows from the trace formula that

$$\operatorname{Tr}(B'_p(m))=\frac{1}{\sqrt{m}}\sum_{s^2\leq 4m}H_p(4m-s^2).$$

Remark that, for all $D\neq 0$

$$\lim_{p\to\infty}\frac{H_p(D)}{p-1}=0.$$

After all,

$$H_p(D)\in\left\{0,H(D),\frac{1}{2}H(D),H_p(D/p^2)\right\},$$

depending on the splitting behaviour of p in O_{-D}. In the case

$$H_p(D)=H_p(D/p^2),$$

there is still some $D'\leq D$ such that

$$H_p(D)\in\left\{0,H(D'),\frac{1}{2}H(D')\right\}.$$

In any case,

$$H_p(D)\leq\sup_{d\leq D}H(D).$$

Hence $H_p(D)$ is not increasing in p and the above limit holds. Recall that for $D=0$, we defined

$$H_p(0)=\frac{p-1}{24}.$$

Note that $4m=s^2$ has a solution over $\mathbb{Z}$ if and only if m is a square. Thus, if m is *not* a square, we have:

$$\begin{aligned}\lim_{p\to\infty}\frac{\operatorname{Tr}(B'_p(m))}{p-1}&=\lim_{p\to\infty}\frac{1}{\sqrt{m}}\sum_{s^2\leq 4m}\frac{H_p(4m-s^2)}{p-1}\\&\leq\frac{1}{\sqrt{m}}\lim_{p\to\infty}\sum_{D=1}^{4m}\frac{H_p(D)}{p-1}\\&=\frac{1}{\sqrt{m}}\sum_{D=1}^{4m}\lim_{p\to\infty}\frac{H_p(D)}{p-1}\\&=0.\end{aligned}$$

Next suppose that m is a square. In this case, $4m=s^2$ has *two* solutions; $s=\pm 2\sqrt{m}$. We split the trace and then proceed the same as before:

$$\operatorname{Tr}(B'_p(m))=\frac{1}{\sqrt{m}}(2H_p(0)+\sum_{s^2<4m}H_p(4m-s^2)).$$

Thus we get

$$\begin{aligned}
\lim_{p\to\infty} \frac{12\,\mathrm{Tr}(B'_p(m))}{p-1} &= \frac{12}{\sqrt{m}} \lim_{p\to\infty} \frac{2H_p(0)}{p-1} + 12 \lim_{p\to\infty} \sum_{s^2<4m} \frac{H_p(4m-s^2)}{p-1} \\
&= \lim_{p\to\infty} 2H_p(0) \frac{12}{\sqrt{m}(p-1)} \\
&= \lim_{p\to\infty} \frac{2(p-1)}{24} \frac{12}{\sqrt{m}(p-1)} \\
&= \frac{1}{\sqrt{m}}.
\end{aligned}$$

□

5 Final thoughts

In this final section of our thesis, we provide some closing remarks on the various proofs and relations we have seen, and posit some avenues for further research.

To start, we discussed several algorithms used for computing Pizer graphs. For these algorithms, we used at various points simple graph theory (connectedness of our graph for the correctness of our first algorithm), CM-theory (for computing an initial supersingular j-invariant) and the more mysterious modular polynomial for our second algorithm, arising in the theory of modular forms.

This section also presented some figures, created from data acquired without specialized equipment or extensive knowledge of programming, showing that computing these graphs is not particularly hard (though doing so efficiently is). This data allows us to think about the graphs in a different manner than as abstract mathematical objects, and observe already well-known properties statistically, providing some intuition.

We observed the graphs approach the Ramanujan bound, that is, the largest non-trivial eigenvalue of $\mathrm{Piz}_p(l)$ approaches $2\sqrt{l}$. We might wonder how swift this convergence occurs, that is, if we write $\mu_p(l)$ for the greatest (in absolute value) non-trivial eigenvalue of $\mathrm{Piz}_p(l)$, how does $2\sqrt{l} - \mu_p(l)$ grow in p (or l)? From Figure 2.3, we might observe that $2\sqrt{2} - \mu_p(2)$ grows sublinearly. Is this indeed the case? The growth of the largest non-trivial eigenvalue is, to our knowledge, less well studied. This would indeed have practical applications, as the estimations made in the mixing lemma (Theorem 2.8) become better with larger spectral gap.

Beyond computational analysis, we also outlined the proofs for two important properties of the spectra of isogeny graphs. First of all, chapter 3 we sketched a proof for the fact that the Pizer graphs are Ramanujan. This proof required us to first understand the relation between the Pizer graphs and (Hecke operators on) modular forms. Secondly, we showed that the eigenvalues of $\mathrm{Piz}_p(l)$ obey some distribution as p tends to infinity, for which we needed the relation with quaternion algebras.

On this last result, if we may speculate somewhat; it is interesting that these deeply number-theoretic graphs obey a distribution law also obeyed by many other families of regular graphs. In our proof, we saw hardly any graph theory, whereas in the proof of McKay there is hardly anything else. We might posit some greater connection. Recall that McKay's paper proves this distribution for regular graphs of increasing size and sub-linear growth of the number of cycles. We could posit that any such family is always related to some facet of number theory, even if the initial definition of such a family is not number-theoretical.

We can easily motivate the relevance of our first result to an application-oriented mathematician. It is relevant to know that these Pizer graphs are Ramanujan, because this implies that they are optimal expanders, and thus have good mixing properties. The motivation for the relevance of the distribution is harder. There does not seem to be an obvious way to use this, neither to enhance schemes nor to break them. The primary motivation for this result is that it is an interesting result with an interesting proof. If this is not sufficient for our application-oriented friend, we will remark that this proof lays bear yet another layer of deep mathematical structure behind these graphs. Through CM-theory, they are intertwined with the classic theory of number fields. Through Hecke operators, there is a relation with complex analysis. Through the Brandt matrices, they are intimately connected to quaternion algebras also.

This structure may have consequences for the cryptographic application of these graphs. On the one hand, it allows us to prove correctness of cryptographic schemes, and perhaps we can use it to construct speedups for computation. On the other, it may be exploited by those with malicious intent to formulate attacks on these schemes. In short, structure can be both a blessing and a curse in cryptography.

For example, the CSIDH scheme (a Diffie-Hellman key-exchange system, see [CLM+18]). This relies on the subgroup of $\mathrm{Piz}_p(l)$ of elliptic curves defined over $\mathbb{F}_p$, and the action of a (commmutative) class group of a quadratic order on this graph. This allows CSIDH several speedups relative to SIDH (which uses the whole graph). On the other hand, there is Kuperbergs algorithm ([Kup03]). This is an algorithm attacking the hidden-subgroup problem for cyclic groups. The fact that the class group used in CSIDH is mostly cyclic, makes this scheme vulnerable to this attack, and as a consequence the number of bits required for a secure implementation of CSIDH is a matter of some debate.

As we mentioned in the introduction, Mestre suggested a way to compute data on modular forms by passing to the Pizer graphs ([Mes86]). The relation between these objects is thus reciprocal; it is not the case that one is always better off thinking of these graphs as Hecke operators or vice versa. It is this interplay of different viewpoints that make these graphs so interesting.

Bibliography

References

[BJS14] Jean-François Biasse, David Jao, and Anirudh Sankar. A quantum algorithm for computing isogenies between supersingular elliptic curves. In *Progress in cryptology—INDOCRYPT 2014*, volume 8885 of *Lecture Notes in Comput. Sci.*, pages 428–442. Springer, Cham, 2014.

[BLS12] Reinier Bröker, Kristin Lauter, and Andrew V. Sutherland. Modular polynomials via isogeny volcanoes. *Math. Comp.*, 81(278):1201–1231, 2012.

[BM19] Ajit Bhand and M. Ram Murty. Class numbers of quadratic fields. *Hardy-Ramanujan J.*, 42:17–25, 2019.

[Bou04] Nicolas Bourbaki. *Integration. I. Chapters 1–6.* Elements of Mathematics (Berlin). Springer-Verlag, Berlin, 2004. Translated from the 1959, 1965 and 1967 French originals by Sterling K. Berberian.

[Brö08] Reinier Bröker. A p-adic algorithm to compute the Hilbert class polynomial. *Math. Comp.*, 77(264):2417–2435, 2008.

[Brö09] Reinier Bröker. Constructing supersingular elliptic curves. *J. Comb. Number Theory*, 1(3):269–273, 2009.

[CLG09] Denis X. Charles, Kristin E. Lauter, and Eyal Z. Goren. Cryptographic hash functions from expander graphs. *J. Cryptology*, 22(1):93–113, 2009.

[CLM$^+$18] Wouter Castryck, Tanja Lange, Chloe Martindale, Lorenz Panny, and Joost Renes. *CSIDH: an efficient Post-Quantum Commutative Group Action.* Cryptology ePrint Archive. IACR, 2018. https://eprint.iacr.org/2018/383.

[Del74] Pierre Deligne. La conjecture de Weil. I. *Inst. Hautes Études Sci. Publ. Math.*, (43):273–307, 1974.

[DS05] Fred Diamond and Jerry Shurman. *A first course in modular forms*, volume 228 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2005.

[DSV03] Giuliana Davidoff, Peter Sarnak, and Alain Valette. *Elementary number theory, group theory, and Ramanujan graphs*, volume 55 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2003.

[Gol77] Dorian M. Goldfeld. The conjectures of Birch and Swinnerton-Dyer and the class numbers of quadratic fields. In *Journées Arithmétiques de Caen (Univ. Caen, Caen, 1976)*, pages 219–227. Astérisque No. 41–42. 1977.

[Gol01] Oded Goldreich. Randomized methods in computation. https://www.wisdom.weizmann.ac.il/~oded/rnd-sum.html, 2001. Accessed: 16-03-2022.

[Gro87] Benedict H. Gross. Heights and the special values of L-series. In *Number theory (Montreal, Que., 1985)*, volume 7 of *CMS Conf. Proc.*, pages 115–187. Amer. Math. Soc., Providence, RI, 1987.

[Kup03] Greg Kuperberg. A subexponential-time quantum algorithm for the dihedral hidden subgroup problem. 2003.

[Lan87] Serge Lang. *Elliptic functions*, volume 112 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1987. With an appendix by J. Tate.

[McK81] Brendan D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40:203–216, 1981.

[Mes86] J.-F. Mestre. La méthode des graphes. Exemples et applications. In *Proceedings of the international conference on class numbers and fundamental units of algebraic number fields (Katata, 1986)*, pages 217–242. Nagoya Univ., Nagoya, 1986.

[Mil86] James S. Milne. Abelian varieties (1986b), 1986. Available at https://www.jmilne.org/math/articles/1986b.pdf.

[Mil08] James S. Milne. Abelian varieties (v2.00), 2008. Available at www.jmilne.org/math/.

[Sch87] René Schoof. Nonsingular plane cubic curves over finite fields. *J. Combin. Theory Ser. A*, 46(2):183–211, 1987.

[Ser73] J.-P. Serre. *A course in arithmetic*. Graduate Texts in Mathematics, No. 7. Springer-Verlag, New York-Heidelberg, 1973. Translated from the French.

[Ser97] Jean-Pierre Serre. Répartition asymptotique des valeurs propres de l'opérateur de Hecke T_p. *J. Amer. Math. Soc.*, 10(1):75–102, 1997.

[Sil94] Joseph H. Silverman. *Advanced topics in the arithmetic of elliptic curves*, volume 151 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994.

[Sil09] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, Dordrecht, second edition, 2009.

[V7́1] Jacques Vélu. Isogénies entre courbes elliptiques. *C. R. Acad. Sci. Paris Sér. A-B*, 273:A238–A241, 1971.

[Voi21] John Voight. *Quaternion algebras*, volume 288 of *Graduate Texts in Mathematics*. Springer, Cham, [2021] ©2021.

[vzGG13] Joachim von zur Gathen and Jürgen Gerhard. *Modern Computer Algebra*. Cambridge University Press, 3 edition, 2013.

A Graph Theory

This is a short recap of some elementary elements of graph theory. This is a part of mathematics where a good intuition can be far better than an exact definition. We will attempt to give both.

One should think of a graph as a collection of points and lines between pairs of those points. In general, one should think of oriented lines, so that there might be a line from point A to point B, but not the other way around. For instance, if we consider a graph where the points are people and the edges are lines of communication, the head of state of your nation probably has a way to reach you, for instance by emergency notification, but you most likely do not have a direct way to communicate to him.

On the other hand, it is possible to get a message to your head of state. You might know someone who works in a governmental department, who knows their boss, who know the minister, who can directly relay the message to your head of state. As such, you might be in some way still connected to your head of state. You can probably not get a message to Charlemagne, however, as he has been dead for over a millennium, so you are not connected to Charlemagne.

There might be multiple ways to get a message to your friend. You might send them an e-mail, or text them, or send them a letter (to readers in the distant future; these are different manners of communication, which depending on how we have done, are either very old-fashioned or extremely advanced to you). As such, there can be multiple edges between you and your friend.

We now get to formal definitions. Keep in the back of your mind the intuition of points and lines, or the concrete example of persons and lines of communication.

Definition A.1. A *graph* G is a pair of sets $G = (V, E)$, elements of V being the *vertices*, and elements of E the *edges*, together with a map $M : E \to V \times V$. We say that a pair of vertices $v, w \in V$ are *neighbours* if $(v, w) \in \mathrm{im}(M)$ or $(w, v) \in \mathrm{im}(M)$.
A graph is *finite* if V and E are finite sets. In this case, the *size* of G is $\#V$
G is *undirected* if, for all $v, w \in V$, $\#M^{-1}(v, w) = \#M^{-1}(w, v)$.
A graph is said to be *simple* if it is undirected, M is injective and no element $v \in V$ is a neighbour of itself.
A *path* of length n in G is a finite sequence of edges $P = (e_1, e_2, \dots, e_n)$ where $M(e_1) = (v_1, w_1), M(e_2) = (v_2, w_2), \dots, M(e_n) = (v_n, w_n) \in E$ such that for all $i \in \{1, \dots, n-1\}$, $w_i = v_{i+1}$. Such a path is a path *from* v_1 *to* w_n. We say such a path *connects* v_1 and w_n.
We say that two elements $v, w \in V$ are *connected* if there is a path of any length from v to w. G is a *connected graph* if all pairs of vertices are connected.

Definition A.2 (Adjacency Matrix)**.** Let $G = (V, E)$ be a finite graph. Then the *adjacency matrix* of G is the matrix A indexed by elements in V, where

$$A_{v,w} = \#M^{-1}(v, w) = \#\{e \in E \mid M(e) = (v, w)\}$$

Remark. The adjacency matrix is thus a record of how many lines are between each pair of points. Going back to our example, there might be 1 line from your head of state to you, none from you to Charlemagne and 3 between you and your friend, and the appropriate entries in the matrix would be 1, 0 and 3, though that 3 would be in two opposite positions in the matrix.

We also see that the total graph of all humans and their connections is finite (as there have only been finitely many humans in a finite part of the universe), but it not a connected graph (as you are not connected to Charlemagne), nor is it undirected. One could consider the connected part of the graph of which you are part, which would be all the people you could send a message to, possibly via many intermediaries.

Clearly A has entries in $\mathbb{Z}_{\geq 0}$ and dimension equal to $\#V$. Some definitions translate nicely; a graph G is undirected if its adjacency matrix A is symmetric, and simple if, on top of symmetry, A has entries in $\{0,1\}$ and $\mathrm{Tr}(A) = 0$. The number of self-loops equals $\mathrm{Tr}(A)$, and the number of paths of length n between v and w is the (v,w)-entry of A^n.

We also define a regular graph

Definition A.3. An undirected graph G with adjacency matrix A is called *k-regular* for some $k \in \mathbb{Z}_{\geq 0}$ if for each row or column T in A, we have $\|T\|_1 = k$.

That is, the sum over each column in A is k. In terms of the graph, this means that every vertex has exactly k edges coming in and exactly k going out! For undirected graphs, there are multiple definitions. One can demand that both rows *and* columns sum to k, or only one. We will use the row-sum definition. In graph terms, this means that each vertex has exactly k vertices going out, though, due to self-loops, it may have a different number coming in.

In this case, we can also get some information on walks without backtracking from the adjacency matrix. Define A_m to be the matrix indexed by V, such that $A_m(v,w)$ is the number of walks of length m without backtracking. Obviously, A_1 is the adjacency matrix of G.

Lemma A.4. *Let G be a finite simple k-regular graph and A_m as above. Define I to be the identity matrix in dimension $|V|$. We have the following relations:*
$A_2 = A_1^2 - k \cdot I$
$A_{r+1} = A_r \cdot A_1 - (k-1)A_{r-1}$.
Furthermore, the A_m's form a commutative family of matrices.

Proof. Since A_1^2 counts the number of paths of length 2 between vertices, we must count how many paths there are of length 2 with backtracking. But a walk of length 2 with backtracking is simply a walk up an edge and then down the same edge. As such, between a pair $v, w \in V$, there are none if $v \neq w$ and (by k-regularity) k if $v = w$. Thus the matrix counting these backtracking walks is kI, and $A_2 = A_1^2 - kI$.

A_rA_1 counts the number of paths of length $r+1$ where in the first r steps there is no backtracking. Consider the (v,w)-th entry of A_rA_1 and a walk between v and w, say $(e_0, \ldots, e_r)$. If the endpoint of e_{r-1} was w, we do backtrack at the last step. Else, we do not. In the first case, we have a path of length $r-1$ between v and w of length $r-1$, and we can choose another vertex from w except e_{r-1}. There are $A_{r-1}(v,w)$ such walks, and we can choose from $k-1$ edges each, so indeed $A_{r+1}(v,w) = A_rA_1(v,w) - (k-1)A_{r-1}(v,w)$ for all (v,w).

Note that $(A_mA_n)(v,w)$ is the number of walks of length $m+n$ between v and w, where there is no backtracking in the fist m or the last n steps. Consider now the reverse of this walk. This is a walk of length $m+n$ from w to v where there is no backtracking in the first n or last m steps. Inversion of walks thus proves that $(A_mA_n)(v,w) = (A_nA_m)(w,v)$. But since G is undirected, $(A_mA_n)(v,w) = (A_mA_n)(w,v)$. Thus indeed the A_m commute. □

This matrix has a natural action on the space of functions

$$l^2(V) = \{f : V \to \mathbb{C}\},$$

given by

$$(Af)(v) = \sum_{w \in V} A_{vw} \cdot f(w).$$

We can now consider eigenvalues of this matrix. The following is proposition 1.1.2 in [DSV03].

Proposition A.5. *Let A be the adjacency matrix of a finite k-regular graph $G = (V,E)$.*

1. *k is an eigenvalue of A.*

2. *For all eigenvalue λ of A, we have $|\lambda| \leq k$.*

3. *The multiplicity of k equals the number of connected components of A. In particular, k has multiplicity 1 if and only if A is connected.*

Proof. For the first point, consider the constant function $C(v) = 1$. Then, by k-regularity:

$$(AC)(v) = \sum_{w \in V} A_{vw} \cdot C(w) = \sum_{w \in V} A_{vw} = k.$$

Hence C is an eigenvector of A with eigenvalue k.

For the second, let f be an eigenfunction with eigenvalue μ. Let x be such that

$$|f(x)| = \max_{v \in V} |f(v)|.$$

Suppose that $f(x) > 0$ (if not, replace f with $-f$). Then we have:

$$\begin{aligned} f(x)|\mu| \quad &= |f(x)\mu| \\ &= |(Af)(x)| \\ &= |\textstyle\sum_{v \in V} A_{xv} f(v)| \\ &\leq f(x) |\textstyle\sum_{v \in V} A_{xv}| \\ &= f(x)k. \end{aligned}$$

Finally, write m for the multiplicity of k, r for the number of connected components and let $G_1, \ldots, G_r$ denote the connected components of G and $G_i = (V_i, E_i)$. Consider first the functions

$$f_i : V \to \mathbb{C}, f_i(v) = \mathbb{1}_{v \in V_i}.$$

These are indicator functions that are 1 on V_i and 0 elsewhere. Let $v \in V$ be given and say $v \in V_j$. We compute

$$(Af_i)(v) = \sum_{w \in V} A_{vw} f_i(w) = \sum_{w \in V_j} A_{vw} f_i(w).$$

If $i = j$, this is $\sum_{w \in V_j} A_{vw} = k$ and we get $(Af_i)(v) = k = k \cdot 1$. Else, this sum is 0, so we get $(Af_i)(v) = 0 = k \cdot 0$. In any case, $(Af_i)(v) = kf_i(v)$, so the f_i are all eigenfunctions of A with eigenvalue k. Hence $m \geq r$.

For $m \leq r$, it suffices to prove that k has multiplicity 1 if A is connected (i.e. if $r = 1$, then $m = 1$), since we reduce to the connected components G_i if G is not connected. Let then f be an eigenfunction with eigenvalue k, and as above, take $x \in V$ such that $|f(x)| = \max_{v \in V} |f(v)|$. We have that

$$f(x) = \frac{(Af)(x)}{k} = \sum_{v \in V} \frac{A_{xv}}{k} f(v).$$

Thus $f(x)$ is a convex combination of numbers of absolute value at most equal to it. Thus $f(x) = f(v)$ for all $v \in V$ and f is constant. Thus k has multiplicity 1. □

To close, we relate eigenvalues of a graph to its bicolourability. This is not immediately relevant for our thesis, but is an often-studied part of graph theory which relates to the spectrum. First, the definition of colourability.

Definition A.6. Let $n \in \mathbb{Z}_{\geq 1}$. We call a graph $G = (V, E)$ *n-colourable* if there is a partition of the vertices

$$V = V_0 \cup V_2 \cup \ldots \cup V_{n-1}$$

such that if vertices v and w have $A_{vw} > 0$ and $v \in V_i$, then $w \notin V_i$.
For $m = 2$, we call such graphs *bicolourable.*

Intuitively, one wishes to mark, usually by colouring, the vertices of a graph such that no two vertices of the same colour are connected. There is a famous result on colourability, which states than any planar graph, that is any graph that can be drawn on flat paper without edges intersecting, is 4-colourable. Since it is known that there are planar graphs that are not 3-colourable, this is an optimal result.

One can relate the n-colourability of a graph to its spectrum in great generality, as is done in paragraphs 1.6 and 1.7 of [DSV03]. We are, for now, interested only in bicolourability. We have the following result, which is proposition 1.1.4 in the previous.

Proposition A.7. *Let $G = (V, E)$ be a finite, connected k-regular graph. The following are equivalent:*

1. *G is bipartite*
2. *The spectrum of G is symmetric about 0*
3. *$-k$ is in the spectrum of G*

Proof. We write A for the adjacency matrix of G.

$1 \to 2$: Let $V = V_0 \cup V_1$ be a partition as in the above definition. Let f be an eigenfunction of A with eigenvalue μ. Define

$$g(v) = \begin{cases} f(v) & v \in V_1, \\ -f(v) & v \in V_2. \end{cases}$$

We claim that g is an eigenfunction for $-\mu$. Let $v \in V$ be given, say $v \in V_i$.

$$(Ag)(v) = \sum_{w \in V} A_{vw} g(w) = \sum_{w \in V_{1-i}} A_{vw} g(w) = -\mu g(v).$$

$2 \to 3$: Obvious, since k is in the spectrum.

$3 \to 1$: Let f be an eigenfunction with eigenvalue $-k$ and $x \in V$ with $|f(x)| = \max_{v \in V} |f(v)|$ and $f(x) > 0$ (again, we may replace f with $-f$). We have

$$f(x) = \frac{(Af)(x)}{-k} = \sum_{v \in V} \frac{A_{xv}}{k} \cdot (-f(y)).$$

As above, $f(x)$ is a convex combination of the $-f(y)$ with absolute value at most $|f(x)|$. Thus, if $A_{xv} \neq 0$, we must have that $-f(y) = f(x)$. To the same effect, for any z adjacent to a y adjacent to x, we must have $-f(z) = f(y)$. Since X is connected and $f(x) > 0$, there *cannot* be $v \in V$ with $f(v) = 0$. We may then define the partition

$$V_0 = \{v \in V \mid f(v) < 0$$

$$V_1 = \{v \in V \mid f(v) > 0,$$

which provides a bicolouring of G. □

For instance, in chapter 3, we have seen that the non-trivial eigenvalues of the Pizer graphs $\mathrm{Piz}_p(l)$, which are $(l+1)$-regular, have absolute value $\leq 2\sqrt{l} < l + 1$ (since $l \geq 2$), so the Pizer graphs are at least not bicolourable!

B SageMath Scripts

Below is the Sage code used to compute Pizer graphs. The figures in chapter 2 were aquired with this code (the actual data-processing was done in the standard python library matplotlib/pyplot).

```
import random
import time
from sage.plot.scatter_plot import ScatterPlot
import numpy as np
from sage.schemes.elliptic_curves.ell_finite_field import is_j_supersingular
from sage.schemes.elliptic_curves.ell_finite_field import supersingular_j_polynomials
working_direc = 'YourFolderHere'#name of data folder
def printToFile(A,name):
    fileName = working_direc+name+'.txt'
    with open(fileName,'w') as f:
        f.write(str(A))
        f.close()

def addZeroColumn(A):
    #Adds column of zeroes to A
    return(np.c_[A,np.zeros(A.shape[0])])

def addZeroRow(A):
    #Adds row of zeroes to A
    return(np.r_[A,[np.zeros(A.shape[1])]])

def addZeroDimen(A):
    #Adds row *and* column of zeroes to A
    if(A.shape == (0,0) or A.shape == (0,)):#empty matrix is weird
        return(np.array([[0]]))
    A = addZeroColumn(A)
    return(addZeroRow(A))



def PrimesTo(n):
    #Just gives a list of all primes p<n
    P = Primes()
    X = [2]
    while(P.next(X[-1])<n):
        X.append(P.next(X[-1]))
    return(X)

def leg(lam):
    #See Silverman, III.1b
    top = lam^2-lam+1
    top = top^3
    bot = lam^2*(lam-1)^2
    return(2^8*top/bot)

def curveFromjInv(j,K,return_coef = False):
    #Returns curve y^2=x^3+Ax+B of j-invariant j
    if(j == 0):#j = 0 and j = 1728 are 'weird' j-invariants
        E =  EllipticCurve(K,[0,1])
    elif(j == 1728):
        E = EllipticCurve(K,[1,0])
    else:#see Silverman, proof of prop. III.1.4c,
        B = -1/(j-1728)
        A = 36*B
        E =  EllipticCurve(K,[1,0,0,A,B]).short_weierstrass_model()
    if(return_coef):
        return(E,E.a_invariants()[-2:])
    else:
        return(E)

def supersingularFromPoly(K,return_coef = False):
    #Based on Silverman, V.4.1b
    #Very inefficient at finding just 1 j-invariant
    count_time = time.time()
    p = K.characteristic()
    if(p<5):# p = 2,3 potentially weird
        return(supersingularjInv(K,return_coef=return_coef))
    m = (p-1)/2
    coef = [GF(p)(choose(m,x))^2 for x in range(m+1)]
    P.<t> = PolynomialRing(GF(p^2))
    t = P.gen()
    f = sum(coef[i]*t^i for i in range(m+1))#this is H_p(t) in Silverman
    rts = f.roots(multiplicities=False)
    return(rts[0],curveFromjInv(j,K,return_coef=return_coef))

def supersingularjInv(K,return_coef = False,keepTime = False):
    #Returns a supersingular j-invariant of K
    count_time = time.time()
    notSS = True
    j = 0
    p = K.characteristic()
    if p is 0:
        raise ValueError('Characteristic 0 not implemented!')
    #have criterion for detecting if j = 0 or j = 1728 are supersingular
    if(Mod(p,3) == 2):
        j = 0
    elif(Mod(p,4) == 3):
        j = 1728
    else:#if neiher are (if p = 1 mod 12), just try 'random' elts in F_p
        while(notSS):
            j += 1
            notSS = not(is_j_supersingular(K(j)))
    if keepTime:
        print('finding j cost',time.time()-count_time,'seconds')
    return(j,curveFromjInv(j,K,return_coef = return_coef))

def imagejInv(E,P):
    #Finds j-invariant of E/<P>, via Velu
    #Only works for E in reduced Weierstrass
    a4,a6 = E.a_invariants()[-2:]
    b2,b4,b6,b8 = E.b_invariants()
    x = P.xy()[0]
    t = 3*x^2+a4
    u = 4*x^3+b2*x^2+2*b4*x+b6
    w = u+t*x
    A = a4-5*t
    B = a6-b2*t-7*w
    j = -1728*(4*A)^3/(-16*(4*A^3+27*B^2))
    return(j)
```

```
def getAdjMatrix(p,l = 2, mtrx = True,keepTime = False, justMat = False):
    #Computes the adjacency matrix of PG_l(p)
    #Currenlty only implemented for l = 2
    if l != 2:
        raise ValueError('Higher-degree Pizer graphs not implemented!')
    if keepTime:
        start_time_mat = time.time()
    Fp = GF(p)
    k.<a> = GF(p^2)
    P.<x>=  PolynomialRing(k)
    j,E = supersingularjInv(Fp)
    finishedJs = []#this is where we store our finished j-invariants
    matrixJs = [j]#this is the list for matrix-indices
    toDo = [j]#our to-do list
    adjMat = np.array([[0]])
    while(toDo!=[]):
        j = toDo[0]
        n = matrixJs.index(j)#remember which index we came from
        E = curveFromjInv(j,k)
        Tf = E.torsion_polynomial(l)/4#if E: y^2 = x^3+Ax+B Tf = x^3+Ax+B
        xCo = Tf.roots(multiplicities=0)
        twoTors = [E(x,0) for x in xCo]#two-torsion points, from torsion polynomial
        isogjInvs = [imagejInv(E,x) for x in twoTors]
        for isJ in isogjInvs:#fills in the matrix
            if isJ in matrixJs:
            #If isJ is already known, fill in the right place in the matrix
                m = matrixJs.index(isJ)
            else:
            #Else, we add a row of 0s (since the new isJ can't be connected
            #To already-treated invariants)
                adjMat = addZeroDimen(adjMat)
                m = adjMat.shape[0]-1
                matrixJs.append(isJ)
            adjMat[n][m]+=1#add 1 edge between the nth and mth j-invariant
        finishedJs += [j]
        toDo += isogjInvs
        toDo = [elt for elt in toDo if elt not in finishedJs]#remove treated j-inv
    if(mtrx):#As stands, adjMat is a numpy double list, which isn't a sage matrix
        adjMat = matrix(adjMat)
    if keepTime:
        print('time elapsed:',time.time()-start_time_mat,' seconds')
    if justMat:
        return(adjMat)
    return(adjMat,matrixJs)

print('starting up!')
X = [p for p in PrimesTo(100000) if p>3]#primes to compute
timeList = []#track total time taken per prime
#can add more lists for different data
for p in X:
    start_time = time.time()
    print('this is for',p)
    A,jInvs = getAdjMatrix(p,keepTime = False)
    timeList.append(time.time()-start_time)
    print(A)
    print(jInvs)
    #printToFile(timeList,'timeTaken')
    print('time elapsed:',time.time()-start_time,' seconds')
```

We should mention that Sage has an in-built function to compute isogenies of prime degree belonging to a elliptic curve. If E is an elliptic curve and l is a prime

```
E.isognies_prime_degree(l)
```

will return a list of all isogenies $\phi_1, \ldots, \phi_n$ of degree l and domain E. So, if one knows the supersingular j-invariants $j_1, \ldots, j_n$ over $\mathbb{F}_{p^2}$, one can do the following:

```
K.<a> = GF(p^2)
listjs = [j1,...,jn]
A = list(matrix(nrows = n,ncols = n))#can't edit matrix entries
for j in listjs:
    E = curveFromjInv(j,K)
    col = listjs.index(j)
    for phi in E.isogenies_prime_degree(l):
        codoj = phi.codomain().j_invariant()
        row = listjs.index(codoj)
        A[col][row]+= 1
A = matrix(A)#convert  back to matrix
```

This is for instance how we computed $B(3)$ in example 3.5. If one doesn't know the list of j-invariants, the following short program can very inefficiently compute $B_p(l)$ for (very) small p and l.

```
K.<a> = GF(p^2)
listjs = []
for j in K:
    if curveFromjInv(j,K).is_supersingular():
        listjs.append(j)
print(listjs)
n = len(listjs)
A = list(matrix(nrows = n,ncols = n))
for j in listjs:
    E = curveFromjInv(j,K)
    col = listjs.index(j)
    for phi in E.isogenies_prime_degree(l):
        codoj = phi.codomain().j_invariant()
        row = listjs.index(codoj)
        A[col][row] += 1
matrix(A)
```