A panaroma of the fundamental group of the modular orbifold

A. Muhammed Uludağ* and Ayberk Zeytin**

Abstract. We give an overview of the category of subgroups of the modular group, incorporating both the tame part, i.e. finite index subgroups, and the non-tame part, i.e. the rest. We also discuss arithmetic related questions which exist in both the tame part (via Belyi's theorem) and the non-tame part.

2000 Mathematics Subject Classification: 14H30, 32G15; 11G32.

Keywords: modular group, modular graphs, fundamental group, étale coverings, çarks, binary quadratic forms, universal Teichmüller space

Contents

1 Introduction ............................................................... 2
2 Modular graphs ............................................................ 3
  2.1 Modular tile and the holy triality ................................. 6
  2.2 Special modular graphs ............................................ 7
    2.2.1 Congruence modular graphs ................................ 8
  2.3 Non-tame subgroups ................................................. 8
3 Arithmetic of modular graphs ......................................... 13
  3.1 Arithmetic of tame subgroups .................................... 13
    3.1.1 Galois action on algebraic fundamental groups ......... 14

*Work supported by a Galatasaray University Research Grant and the grant TUBITAK-114R073
**Work supported by TÜBİTAK Career Grant No. 113R017
1 Introduction

The fundamental group of the modular orbifold, $\mathcal{M}$, is the modular group, $\text{PSL}(2, \mathbb{Z})$. Due to the general topological facts, there is a one to one correspondence between subgroups of the modular group and covers of $\mathcal{M}$. Denote by $\text{Sub}(\text{PSL}(2, \mathbb{Z}))$ the category of all subgroups of the modular group with inclusions as morphisms and by $\text{FSub}(\text{PSL}(2, \mathbb{Z}))$ the category of all finite-index subgroups of $\text{PSL}(2, \mathbb{Z})$, with inclusions as morphisms. These two categories are equivalent to the base-pointed covering category $\text{Cov}^*(\mathcal{M})$ ($\text{FCov}^*(\mathcal{M})$) of (finite) covers of the modular orbifold, respectively. Denote by $\text{Cov}(\mathcal{M})$ and $\text{FCov}(\mathcal{M})$ the corresponding covering categories obtained by forgetting the base points. Our aim here is to give the reader a panorama of these categories.

The categories $\text{Cov}(\mathcal{M})$ and $\text{Cov}^*(\mathcal{M})$ and $\text{Sub}(\text{PSL}(2, \mathbb{Z}))$ contains uncountably many objects. On the other hand, objects in the categories $\text{FCov}(\mathcal{M})$ and $\text{FSub}(\mathcal{M})$ can be represented by modular graphs, i.e. a certain class of finite bipartite ribbon graphs or dessins d’enfants, usually thought of as graphs embedded on oriented topological surfaces. By a celebrated theorem of Belyi, [2], these graphs admit a Galois action, i.e. an action of the absolute Galois group, $G_{\mathbb{Q}}$. This action is faithful. The idea of studying $G_{\mathbb{Q}}$ through this interaction between $G_{\mathbb{Q}}$ and the modular group $\text{PSL}(2, \mathbb{Z})$ is the main motivation of the “dessins” program, [11, 10]. However, Deligne put already in 1989: “Grothendieck and his students developed a combinatorial description (“maps”) of finite covers of $\mathbb{P}^1_{\mathbb{C}}$... This did not help to understand the Galois action. One has just a few examples of non-solvable covers whose Galois conjugates have been computed”, [4].

Confirming Deligne’s convictions, it seems that the picture that emerged after almost 40 years of research is far from the original expectations expressed very lively in [11] and [10]. It turns out that understanding this Galois action on all of $\text{FCov}(\mathcal{M})$ appears to be a hopeless task. One turns attention to some manageable substructures of it. In the previous chapter, a certain substructure (the family of hypergeometric covers) of $\text{FCov}(\mathcal{M})$ is suggested to be studied from the arithmetic point of view. In the current chapter, we suggest to go beyond $\text{FCov}(\mathcal{M})$, and study some infinite covers in $\text{Cov}(\mathcal{M})$. It turns out that although of completely different nature, arithmetic is still present for infinite index subgroups, see for instance [24] and [28].
2 Modular graphs

Let $\mathbb{H}$ be the upper half plane with the standard action of the modular group $\text{PSL}(2, \mathbb{Z})$; i.e. the group of invertible $2 \times 2$ matrices with integer coefficients. This action sends $z \in \mathbb{H}$ to
\[
\begin{pmatrix} p & q \\ r & s \end{pmatrix} \cdot z \mapsto \frac{pz + q}{rz + s}.
\]
Its quotient space is called the \textit{modular orbifold} and denoted $\mathcal{M}$:
\[
\mathcal{M} = \text{PSL}(2, \mathbb{Z}) \backslash \mathbb{H}
\]
The modular orbifold has two orbifold points, $i$ with stabiliser $\langle S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \rangle \cong \mathbb{Z}/2\mathbb{Z}$ and $\exp 2\pi i/3$ with stabilizer $\langle L = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \rangle \cong \mathbb{Z}/3\mathbb{Z}$. Moreover, its orbifold fundamental group is
\[
\pi_1(\mathcal{M}) = \text{PSL}(2, \mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z} \ast \mathbb{Z}/3\mathbb{Z}.
\]

By the usual correspondence from topology, the subgroups of the fundamental group classify the coverings of $\mathcal{M}$. The underlying space of $\mathcal{M}$ is the projective line $\mathbb{P}^1$, the two orbifold points with $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$-inertia may be assumed to be respectively at 0 and 1. The translation $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is responsible for the puncture of the quotient which may be taken to be the point at $\infty$. There is a degree-six Galois covering of the modular orbifold by the projective line punctured at three points; $\mathbb{P}^1 \setminus \{0, 1, \infty\}$. Therefore every covering of $\mathbb{P}^1 \setminus \{0, 1, \infty\}$ is also a covering of $\mathcal{M}$. Note that every covering of $\mathcal{M}$ carries a canonical hyperbolic metric induced from the Poincaré metric on the upper half plane.

Finite covers of the modular orbifold admit a combinatorial description by certain graphs (Linienzüge of Klein, \cite{Klein}), as follows. Consider the arc connecting the two elliptic points on the boundary of the standard fundamental domain of the $\text{PSL}(2, \mathbb{Z})$ action on $\mathbb{H}$, see Figure \[24\].

Then the $\text{PSL}(2, \mathbb{Z})$-orbit of this arc is a tree $\mathcal{F}$, which we call the \textit{Farey tree}. Now by its very definition, this tree admits a $\text{PSL}(2, \mathbb{Z})$-action, and a quotient graph by a subgroup of finite or infinite index $G < \text{PSL}(2, \mathbb{Z})$ has been named by us a \textit{modular graph}\footnote{Modular graphs are very similar to trivalent ribbon graphs. This new baptisation is our modest contribution to the very rich and diversified terminology concerning these combinatorial objects.} \cite{Linienzuege, ModularGraphs}. Thus the modular graph $G \backslash \mathcal{F}$ sits inside the curve $G \backslash \mathbb{H}$ in a standard way. In particular, the quotient \textit{orbi-graph} $\text{PSL}(2, \mathbb{Z}) \backslash \mathcal{F}$ is an arc connecting the two orbifold points of the modular orbifold $\text{PSL}(2, \mathbb{Z}) \backslash \mathbb{H}$. We call this the \textit{modular arc} and denote it by $\circ \rightarrow \bullet$. 


Figure 1: The standard fundamental domain for the action of $\text{PSL}(2, \mathbb{Z})$ on the upper half plane.

From another perspective, modular graphs are coverings (in the orbifold sense) of the modular arc $\circ \to \bullet$. The degree-$n$ covers of $\circ \to \bullet$ are obtained by gluing $n$ copies of $\circ \to \bullet$ at the endpoints, to obtain a connected graph, so that

1. endpoints of different types never meet,
2. type-$\circ$ vertices are of degree either 1 or 2, and type-$\bullet$ vertices are of degree either 1 or 3, and
3. for each vertex of degree 3, a cyclic order of the edges meeting at that vertex is given.

A systematic study of these graphs in connection with modular curves was made by Kulkarni, under the name cyclic trivalent graphs and tree diagrams. The Farey tree $\mathcal{F}$ is the universal cover of $\circ \to \bullet$.

Perhaps the simplest modular graphs that one may draw this way (as coverings of the modular interval) are finite planar trivalent trees, (with two types of terminal vertices and a puncture at infinity). Hence we may consider finite planar trivalent trees as natural objects in $\text{FSub}(\text{PSL}(2, \mathbb{Z}))$. Their Belyi maps are polynomials generalizing the Chebyshev polynomials.

One may imagine widening the puncture of the modular orbifold $\mathcal{M}$, until the modular orbifold “deformation retracts” to $\circ \to \bullet$ inducing an isomorphism of fundamental groups. Let us observe what happens to the coverings during this deformation retraction: at time 0 we start with punctured surfaces, at time $1 - \epsilon$ we see ribbon graphs and at time 1 we get the cyclic trivalent graphs. Meanwhile the universal cover of $\mathcal{M}$, i.e. the upper half plane, deformation
retracts to the Farey tree. One may imagine that the category \( \text{FCov}(\mathcal{M}) \) itself is deformation retracting to \( \text{FCov}(\circ\to\bullet) \). Naming the inverse limit of \( \text{FCov}^*(\circ\to\bullet) \) as the ribbon solenoid, this operation gives a deformation retract of the punctured solenoid to the ribbon solenoid, which is a compact space. So we obtain:

![Figure 2: Deformation retract of the ambient surface to a modular graph.](image)

**Theorem 2.1.** The pointed categories \( \text{FCov}^*(\mathcal{M}) \) and \( \text{FCov}^*(\circ\to\bullet) \) (= pointed finite modular graphs) are equivalent via deformation retraction, and they are both arrow-reversing equivalent to \( \text{FSub}(\text{PSL}(2, \mathbb{Z})) \) via the fundamental group functor.

The category of modular graphs admits a standard realization by piecewise geodesic and analytic arcs of \( \text{FCov}^*(\mathcal{M}) \), via

\[
G\backslash\mathcal{F} \subset G\backslash\mathbb{H}.
\]

Since there are countably many finite modular graphs, these categories admit countably many objects. The formula \( \#\{\text{vertices}\} - \#\{\text{edges}\} + \#\{\text{faces}\} \)

2A face on \( G\backslash\mathcal{F} \) is a finite left-turn closed path.


determines the Euler characteristic and hence the genus of the curve, whenever the graph is finite.

In fact, we haven’t made use of any finiteness property in the proof, and therefore conclude more generally that the categories \( \text{Cov}^*(\mathcal{M}), \text{Cov}^*(\circ\to\bullet) \) and \( \text{Sub}_{\text{PSL}(2, \mathbb{Z})} \) are equivalent as well. There are uncountably many infinite modular graphs, so there are as many infinite coverings of the modular orbifold, equivalently, as many subgroups of the modular graphs.

If one erases vertices of degree two from a modular graph then the remaining graph becomes trivalent, i.e. every vertex is of degree 3. Hence, trivalent ribbon graphs [20] are inside \( \text{FCov}(\circ\to\bullet) \) in a natural way. There are more modular graphs than trivalent ribbon graphs since the former are allowed to have terminal edges abutting at orbifold points of both types. On the other hand, “a trivalent ribbon graph (or an ideal triangulation) with a distinguished
oriented edge” encountered frequently in Teichmüller theory (both in the finite part, see [20, 6] and the infinite part, see [19, 13]) is same as the data “a modular graph with the choice of a base edge”. Note that the mapping class group acts freely on the set of embedded trivalent ribbon graphs, and the groupoid associated to this group action admits a nice presentation with modular graphs as objects and graph flips as generator of morphisms, see [20].

Every triangulation of a surface has a dual graph which is an embedded trivalent ribbon graph. If we include degenerate triangulations this correspondence extends to modular graphs. This correspondence is bijective between the orientation-preserving homeomorphism classes of triangulations and trivalent ribbon graphs.

Conversely, any modular graph determines a covering of \( \mathbb{P}^1 \) branched only at 0, 1 and \( \infty \). Indeed, given a modular graph one chooses a base edge \( e_0 \) on the graph. Then, every finite loop based at \( e_0 \) on the graph determines an element, \( W \in \text{PSL}(2, \mathbb{Z}) \) in the following way: starting with the empty word \( W \) while tracing the loop if a degree two vertex is visited one adds an \( S \) to \( W \) and if a degree three vertex is visited then one writes \( L \) (resp. \( L^2 \)) if the next edge on the loop is on the left (resp. right) of the initial edge. Then, all such loops generate a subgroup, say \( G < \text{PSL}(2, \mathbb{Z}) \), and as a result of the construction explained above, the orbifold \( \text{PSL}(2, \mathbb{Z}) \backslash \mathbb{H} \) deformation retracts onto the modular graph. In fact, the Riemann surface \( G \backslash \mathbb{H} \) can be defined over \( \mathbb{Q} \).

### 2.1 Modular tile and the holy triality.

Every modular graph \( G \backslash \mathcal{F} \) has a standard piecewise analytic realization on the Riemann surface \( G \backslash \mathbb{H} \) with edges being geodesic segments with respect to the hyperbolic metric induced from the upper half plane. Equivalently, these edges are lifts of the modular arc by the canonical projection \( G \backslash \mathbb{H} \to \text{PSL}(2, \mathbb{Z}) \backslash \mathbb{H} \). If instead we lift the geodesic arc connecting the \( \mathbb{Z}/3\mathbb{Z} \)-elliptic point to the cusp to the surface \( \text{PSL}(2, \mathbb{Z}) \backslash \mathbb{H} \), then we obtain another graph, which is the associated ideal triangulation, i.e. a triangulation of the surface whose vertices are exactly at the punctures. Lifting the remaining geodesic arc gives rise to yet another type of graph, called a lozenge tiling. So there is a triality, not just duality, of these graphs, see Figure 3.

In fact, for some purposes it is more natural to study the \( \text{FSub} \) of the extended modular group \( \text{PGL}_2(\mathbb{Z}) \). This is the fundamental group of the quotient \( \mathcal{M}^* \) of the modular orbifold \( \mathcal{M} \) by the complex conjugation. We shall call \( \mathcal{M}^* \) the modular tile. Elements of \( \text{FCov}(\mathcal{M}^*) \) are surfaces tiled by copies of \( \mathcal{M}^* \) and they include bordered surfaces with some corners and

---

Footnote 3: Every terminal vertex is considered as an orbifold point. Hence the path from \( e_0 \) to a terminal vertex and back is a generator of finite order of the subgroup.
punctures, possibly lying on the boundary. Every element of $\mathbf{FCov}(\mathcal{M}^*)$ has a double which is in $\mathbf{FCov}(\mathcal{M})$. Any polygon, which is triangulated by its diagonals is an element of $\mathbf{FCov}(\mathcal{M}^*)$. Lifts of the each one of the three edges of the modular tile to its coverings gives rise to modular graphs, triangulations, and lozenges.

### 2.2 Special modular graphs

Now that we have introduced modular graphs, the next task is to classify them. There are several schemes that come to mind, the first being with respect to genus and the number of punctures/orbifold points of the ambient surface $G \setminus \mathbb{H}$. It seems natural to denote this slice of the covering category by

$$\mathbf{FCov}_{d,g,n,n_c;n_\circ}(\mathcal{M}).$$

Here $d$ is the number of edges of the modular graph ($=$ index of a corresponding subgroup), $g$ is the genus of the ambient Riemann surface, $n$ is the number of punctures, $n_c$ is the number of orbifold points of type 'circ', $n_\circ$ is the number of orbifold points of type '•'. Note that these invariants are not independent and they can be defined in a purely combinatorial manner. The genus-0 case is the case of modular graphs on the sphere. The lowest value that the number of punctures $n$ can attain is 1. When the genus is 0, the case $n = 1$ yields
loop-free modular graphs, i.e. planar bipartite trees with vertices of type ◦ being of degree 1 or 2 and those of type • being of degree 1 or 3. We may call them “modular trees”. If we erase vertices of degree 2, these are precisely the planar trivalent trees.

Reasoning geometrically or combinatorially, one may imagine other ways to classify the modular graphs and operations to produce new graphs from old ones. One possibility is to take a triangulation, simultaneously subdivide each edge into $n$ segments and then simultaneously subdivide every triangle into $n^2$ smaller triangles by connecting the endpoints of these segments in the obvious manner. In the previous chapter it is shown that this operation is of some relevance in the theory.

2.2.1 Congruence modular graphs. There is a well-known and important class of subgroups in the panorama of the modular group. These are defined as, for positive integer $N$

\[
\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z}) \mid c \equiv 0 \text{ Mod } N \right\}
\]

\[
\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z}) \mid c \equiv 0, \ a \equiv d \equiv 1 \text{ Mod } N \right\}
\]

\[
\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z}) \mid b \equiv c \equiv 0, \ a \equiv d \equiv 1 \text{ Mod } N \right\}.
\]

$\Gamma(N)$ is called the principal modular congruence group and $\Gamma_0(N)$ and $\Gamma_1(N)$ are said to be modular groups of Hecke type. The same definitions apply to the projectivizations. A subgroup of the modular group containing a principal congruence subgroup is called a congruence modular group, and the corresponding coverings of the modular orbifold are called modular curves. For details (which occupy a considerable volume in the mathematical literature) about these subgroups we refer the reader to [18].

Principal congruence subgroups are the kernels of the reduction maps onto $\text{SL}_2(\mathbb{Z}/N\mathbb{Z})$ and therefore they are normal subgroups. Hence they correspond to Galois coverings of the modular orbifold and provide us with a family of modular graphs with a transitively acting monodromy group.

2.3 Non-tame subgroups

The finite part $\text{FCov}(\mathcal{M})$ (or $\text{FSub}(\text{PSL}(2,\mathbb{Z}))$ of $\text{Cov}(\mathcal{M})$ (or $\text{Sub}(\text{PSL}(2,\mathbb{Z})))$ will be referred to as the tame part of the category because of the fact that these give rise to étale covers of the modular orbifold and conversely. All the remaining objects in the category will be called non-tame. As we noted above, the full category $\text{Cov}(\mathcal{M})$ admit uncountably many objects and what is beyond $\text{FCov}(\mathcal{M})$ appears to be a wild territory. However, there
is a subcategory of $\text{Cov}(\mathcal{M})$ which consists of infinite covers of finite topology (i.e. with finitely many punctures, boundary components and handles). We denote this sub-category by $\text{FGICov}(\mathcal{M})$ (Finitely Generated Infinite Covers). Once again there is the pointed category $\text{FGICov}^*(\mathcal{M})$. As the name suggests, the corresponding subgroup category $\text{FGISub}(\text{PSL}(2, \mathbb{Z}))$ consists of finitely generated subgroups of infinite index in the modular group. Unlike $\text{FCov}(\mathcal{M})$, it has no initial objects as the modular orbifold itself is not included in it. As a second remark, note that an infinite cover of an object of $\text{FGICov}(\mathcal{M})$ may or may not be inside $\text{FGICov}(\mathcal{M})$. On the other hand, this category is closed under finite covers. However, it is not true that any two objects have a common finite cover (i.e. the intersection of two finitely generated infinite subgroups is not always a finitely generated subgroup), so one cannot pass to the limit of $\text{FGICov}^*(\mathcal{M})$ to define some sort of proinite completion.

We see that one may consider two versions of $\text{FGICov}(\mathcal{M})^*$, one with finite covers and the other with finite or infinite covers as morphisms. The idea is that if we restrict to finite covers then the remaining category will become entirely “arithmetic”. Taking an object in this restricted category and considering the subcategory of Galois coverings of this object, one may define the fundamental group of this object to be the limit of the inverse system of Galois groups of these normal covers.

In order to demystify what have been said in the above paragraphs, let us consider some examples:

**Example 2.2.** The very first example is the upper half plane itself, corresponding to the trivial subgroup, which is indeed finitely generated of infinite index. Observe that this is the final object of $\text{FGICov}(\mathcal{M})^*$ (if we admit infinite coverings inside the category). The corresponding graph object is the bipartite Farey tree itself, see Figure 4. Observe that the action of the modular group can also be described as an action on the bipartite Farey tree. This is very useful in considering graphs of arbitrary subgroups.

**Example 2.3 (Finite Subgroups).** As the next example let us consider the elliptic subgroup $\{1, S\}$. In this case, the graph has infinitely many vertices, and $S$ acts on the Farey tree by rotation about one of the degree 2 vertices of the graph. Therefore $\Gamma_G$ is obtained by folding the Farey tree from the symmetry vertex, see Figure 5. In a similar fashion the matrix $L$ acts on the Farey tree as a rotation around a degree 3 vertex of degree $2\pi/3$, see Figure 5 for the graph corresponding to $\{I, L, L^2\}$. This depicts the general case for subgroups generated by one element of finite order. Indeed, every such element is conjugate to either $S$ or $L$.

\footnote{For a concrete example, take the two $\mathbb{Z}$-subgroups $\langle \gamma_1 \rangle$ and $\langle \gamma_2 \rangle$, where $\gamma_1$ and $\gamma_2$ are two distinct primitive elements of infinite order.}
Example 2.4 (Parabolic Subgroups). An element $W$ of $\text{PSL}(2, \mathbb{Z})$ is called parabolic if its absolute trace is 2. As was the case for finite subgroups, any such element is conjugate either to $(LS)^k$ or $(L^2S)^k$ for some integer $k$. The corresponding graph has a unique loop (called the spine) containing $2k$-many edges. There are $k$-many Farey tree components (called a Farey branch) attached to the degree 3 vertices of the spine (there are $k$ such vertices!). Observe that the Riemann surface $\langle W \rangle \backslash \mathbb{H}$ is a punctured disk. As was observed previously punctures on the Riemann surface are in one to one correspondence with finite left-turn closed paths on the graph. Therefore all Farey branches appear either on the left or on the right of the spine, depending on the orientation and they expand in the direction of the outer boundary, see Figure 6.

Example 2.5 (Hyperbolic Subgroups). The remaining class of elements in $\text{PSL}(2, \mathbb{Z})$ are hyperbolic elements, i.e. those elements which are of absolute
A Panorama of the fundamental group of the modular orbifold

Figure 6: The modular graph corresponding to $G = \langle (LS)^6 \rangle$.

trace strictly greater than 2. Modular graphs corresponding to such elements, named çark (pronounced chark), again have a unique loop, which is referred to as spine. For any hyperbolic element $W$ the Riemann surface $\langle W \rangle \setminus \mathbb{H}$ is an annulus. There is at least one Farey branch expanding in the direction of the two boundary components, referred to as inner and outer. The number of Farey branches is determined solely by the element $W$, see Figure 7 for the graph corresponding to the subgroup generated by $W = LSL^2S$.

Figure 7: The modular graph corresponding to $G = \langle LSL^2S \rangle$.

**Example 2.6 (Dihedral Subgroups).** A hyperbolic element $W \in PSL(2, \mathbb{Z})$ is called reciprocal if it is conjugate to its own inverse. It turns out that the element conjugating $W$ to $W^{-1}$, denoted $Z_W$, is determined uniquely up to multiplication by elements of the group $\langle W \rangle$. $Z_W$ is of order 2. This reflects itself as a special symmetry of the çark corresponding to $W$. Namely, if a hyperbolic element $W$ is reciprocal, then the two words, one obtained by tracing the spine of the çark clockwise and the other counterclockwise are same, see Figure 8 for the graph corresponding to the dihedral subgroup $\langle W, Z_W \rangle$.
Such subgroups are related to reciprocal geodesics in the modular orbifold which are of some interest from the point of view of analytic number theory, \cite{21}.

**Example 2.7** (Subgroups isomorphic to the modular group itself). Any pair of elements $S'$ and $L'$, of orders 2 and 3 respectively, generate a subgroup of $\text{PSL}(2, \mathbb{Z})$ which is isomorphic to $\text{PSL}(2, \mathbb{Z})$ itself, and this is a proper subgroup whenever the product $S'L'$ is not parabolic. In this case the modular graph $\langle S', L' \rangle \backslash \mathcal{F}$ will be the planar tree with only two terminal vertices, one of type $\circ$ and the other of type $\bullet$.

The above set of examples depict the general picture. That is to say, punctures on the Riemann surface correspond to finite left turn loops on the graph, infinite Farey branches correspond to boundaries of the Riemann surface that are homotopic to the circle, etc.

**Figure 9:** The modular graph corresponding to the pair of pants determined by the subgroup generated by $L^2S(LS)^2L^2SL$ and $SLSL^2SLS(L^2S)^4(LS)^2$. 

The modular graph of a dihedral subgroup

\[ \langle S', L' \rangle \backslash \mathcal{F} \]
To summarize the discussion, elements of $\text{FGICov}(M)$ are coverings of the modular orbifold by Riemann surfaces, with a finite number of punctures, orbifold points and boundary components. Note that the planar surfaces in $\text{FGICov}(M)$ are the Schottky uniformizations of the modular orbifold. Let us finally remark that for any object $X$ in $\text{FGICov}(M)$ the category $\text{FCov}(X)$ consisting of finite covers of $X$ is in fact a sub-category of $\text{FGICov}(M)$. Unlike the ambient category $\text{FGICov}(M)$, any two covers in $\text{FCov}(X)$ has a common cover in $\text{FCov}(X)$.

### 3 Arithmetic of modular graphs

This section is devoted to the study of various arithmetic questions around modular graphs. It turns out that the tame part of the category $\text{Cov}(M)$ admits an action of $G_\mathbb{Q}$. The questions here, as addressed in Section 1, are easy to explain. Unexpectedly, there is still very beautiful arithmetic questions around the non-tame part of $\text{Sub}(\text{PSL}(2, \mathbb{Z}))$ (in fact, concerning the finitely generated infinite index subgroups, $\text{FGISub}(\text{PSL}(2, \mathbb{Z}))$).

#### 3.1 Arithmetic of tame subgroups

The absolute Galois group $G_\mathbb{Q}$ of the field of rational numbers is the ultimate product of the classical Galois theory which is laid before us as a fundamental object of study. This is a non-abelian profinite group, which is the inverse limit of Galois groups of all finite Galois groups of the field of rationals. The main open problem of the so-called “inverse Galois theory” is to decide whether there exists a field extension of $\mathbb{Q}$ with a given finite group as its Galois group. We are far from having an explicit description of this inverse system of finite groups and we are unable to pin down an element of the absolute Galois group in an explicit manner, except complex conjugation. Note that being profinite, the absolute Galois group of $\mathbb{Q}$ has the structure of a compact, Hausdorff, totally disconnected topological group. Its size is that of the continuum.

The way to investigate a complicated group is to construct and study its action on some familiar class of objects. An action of the group $G_\mathbb{Q}$ is called a Galois action. If the action is on a vector space and is continuous, it is usually called a Galois representation. A continuous action of a profinite group on a discrete space or a finite-dimensional complex vector space concerns only a finite quotient of the group and cannot reveal its essential structure if the group is infinite.

In order to study an infinite piece of the absolute Galois group, one idea is to take all iterates $\{f_i\}_{i=1}^{\infty}$ of a given polynomial $f$, and study the Galois action on the tree $T_f$ whose vertices are the points $f_i^{-1}(0)$. In case all the $f_i$s
are separable, and \( f \) is of degree \( d \), this is a complete rooted \( d \)-ary tree having a profinite automorphism group \( \text{Aut}(T_f) \), with a continuous homomorphism \( G_\mathbb{Q} \to \text{Aut}(T_f) \). Of course, this representation is never faithful. A systematic study of this action has started quite recently under the name Arboreal Galois Representations (see \[3\]). We shall not pursue this direction with a dynamical system flavor here, nevertheless trees will enter the scene in a different guise. Note that the critical points of \( f \) clutter so the representation we get is in a certain quotient of the fundamental group of the complement of a terrible set.

A technique which proved to be very strong and fruitful to construct and study Galois representations is to use algebraic geometry: \( p \)-adic étale cohomology groups attached to an algebraic variety are finite-dimensional \( \mathbb{Q}_p \)-vector spaces with a continuous action of \( G_\mathbb{Q} \), and the image of such representations may be infinite, although “small” in the sense that such representations always have a “big” kernel. The main open problem of this field is the characterization of representations of geometric origin among all linear representations (Fontaine-Mazur conjecture \[15\]).

A natural next step is to study the Galois actions on (quotients of) algebraic fundamental groups. These are called “large” representations because they may be faithful. Étale fundamental groups were introduced and studied by Grothendieck in the late sixties \[1\]. In this setup, if \( k \) is a field, then the (étale) fundamental group of \( \text{Spec}(k) \) is \( \text{Gal}(k) \). In the arithmetic case one cannot employ the usual topological tools to define the fundamental group. However, there is a good notion of covering, i.e. étale covering, and one uses the category of all étale coverings of the space in question, in order to define its fundamental group.

### 3.1.1 Galois action on algebraic fundamental groups.

Let \( \overline{\mathbb{Q}} \subset \mathbb{C} \) be the algebraic closure of \( \mathbb{Q} \) in \( \mathbb{C} \) and let \( X_\mathbb{Q} \) be a smooth geometrically irreducible variety over \( \mathbb{Q} \) (it is more natural to consider \( X \) to be a stack). Let \( p \in X \) be a rational point (in the scheme-theoretic sense) and let \( \overline{p} \) be a geometric point of \( X_{\overline{\mathbb{Q}}} := X \otimes_{\mathbb{Q}} \overline{\mathbb{Q}} \) above \( p \). One associates étale fundamental groups to \((X, p)\) and to \((X_{\overline{\mathbb{Q}}}, \overline{p})\) together with an exact sequence
\[
1 \to \pi_1(X_{\overline{\mathbb{Q}}}, \overline{p}) \to \pi_1(X_\mathbb{Q}, p) \to G_\mathbb{Q} \to 1
\] (3.1)
which defines an outer action \( \rho_{X,p} : G_\mathbb{Q} \to \text{Out}(\pi_1(X_{\overline{\mathbb{Q}}}, \overline{p})) \). The group \( \pi_1(X_{\overline{\mathbb{Q}}}, \overline{p}) \) is canonically isomorphic to the profinite completion of the topological fundamental group of \( X_{\mathbb{C}} \), that is, \( \pi_1(X_{\overline{\mathbb{Q}}}, \overline{p}) = \hat{\pi}_{1,\text{top}}(X_{\mathbb{C}}, p) \). Here \( X_{\mathbb{C}} \) denotes the complexification of \( X \).

The simplest interesting situation is \( X = \mathbb{P}^1_{\mathbb{Q}} \setminus \{0, \infty\} =: \text{Spec} \mathbb{Q}[x, 1/x] \). In this case the exact sequence is
\[
1 \to \pi_1(\mathbb{P}^1_{\overline{\mathbb{Q}}} \setminus \{0, \infty\}, \overline{p}) \to \pi_1(\mathbb{P}^1_{\overline{\mathbb{Q}}} \setminus \{0, \infty\}, p) \to G_\mathbb{Q} \to 1
\] (3.2)
Since the topological fundamental group of $\mathbb{P}^1_\mathbb{C}\setminus\{0,\infty\}$ is $\mathbb{Z}$, the algebraic fundamental group $\pi_1(\mathbb{P}^1_\mathbb{Q},\overline{p})$ is isomorphic to the profinite completion of integers $\hat{\mathbb{Z}}$. In fact, we can describe the system of all coverings of $X$ explicitly, by the power maps $\varphi_n: X \to X$. The Galois group acts on the monodromy group $\mathbb{Z}/n\mathbb{Z}$ of the covering $\varphi_n$ through the cyclotomic character $\sigma \in G_Q \rightarrow \chi(\sigma) \in \hat{\mathbb{Z}}$, such that $\chi(\sigma): [i] \in \mathbb{Z}/n\mathbb{Z} \rightarrow [i\chi(\sigma)] \in \mathbb{Z}/n\mathbb{Z}$. (If we represent $\mathbb{Z}/n\mathbb{Z}$ as the group of $n$th roots of unity in $\mathbb{C}$ generated by a primitive $n$th root $\xi$, the action is described by $\xi^i \rightarrow \xi^{i\chi(\sigma)}$). For a given element $\sigma$, this defines an automorphism of each monodromy group, in a way compatible with the system of coverings $\{\varphi_n\}$. To sum up, the above exact sequence becomes

$$1 \rightarrow \hat{\mathbb{Z}} \rightarrow \pi_1(\mathbb{P}^1_\mathbb{Q}\setminus\{0,\infty\}) \rightarrow G_Q \rightarrow 1 \quad (3.3)$$

inducing a surjective homomorphism $G_Q \rightarrow \text{Aut}(\hat{\mathbb{Z}}) \simeq \hat{\mathbb{Z}}^\times$. In fact, this is the abelianization map of $G_Q$.

The simplest non-abelian example of a Galois action on a fundamental group is the case of the dihedral triangle orbifold with signatures $(2,2,\infty)$. However, this requires some preparation to orbifolds (or stacks), so let's make the situation non-abelian by removing one more point from $\mathbb{P}^1_\mathbb{Q}\setminus\{0,\infty\}$. So let $Y = \mathbb{P}^1_\mathbb{Q}\setminus\{0,1,\infty\}$ (=: Spec $\mathbb{Q}[x, 1-x, 1/x]$). Since the topological fundamental group of $\mathbb{P}^1_\mathbb{Q}\setminus\{0,1,\infty\}$ is free of rank 2, the group $\pi_1(\mathbb{P}^1_\mathbb{Q}\setminus\{0,1,\infty\},\overline{p})$ is the free profinite group $\hat{F}_2$ of rank 2. To sum up, the above exact sequence becomes

$$1 \rightarrow \hat{F}_2 \rightarrow \pi_1(\mathbb{P}^1_\mathbb{Q}\setminus\{0,1,\infty\}) \rightarrow G_Q \rightarrow 1. \quad (3.4)$$

So any element of $G_Q$ determines an automorphism of $\hat{F}_2$, up to an inner automorphism. There is a standard way to lift the outer Galois action to an outer action $\text{Aut}(\pi_1(X,\overline{p})) \simeq \text{Out}(\hat{F}_2)$. In other words, there is a homomorphism $G_Q \rightarrow \text{Out}(\hat{F}_2)$.

Belyi's theorem implies that this representation is faithful, and the characterization of its image is the main question of Grothendieck-Teichmüller theory (GT) initiated by Belyi, Drinfeld and Ihara (see [23] and [17]). As a result of decades of intense research by prominent mathematicians, several sets of equations that must be respected by this image have been found. At the current level of research, it seems an intractable question to see whether these equations are independent or whether they are sufficient to characterize the image.

We shall probably never have a satisfactory understanding of the group $\text{Aut}(\hat{F}_2)$ and the associated Galois action. One may say that the Galois representations on fundamental groups are “too large”. A remedy to this problem is to consider representations on smaller quotients of fundamental groups, such
as the pro-nilpotent or pro-l fundamental groups. At this point starts the motivic side of the theory, initiated by Wojtkowiak and Deligne (see [5] and [12]). Although this is a very rich territory, with connections to multiple zeta values, polylogarithms, etc. it is still “linear” and “far from the anabelian dream imagined by Grothendieck.” For a discussion of the “non-linear” theory, see [17]. The hypergeometric Galois actions that we proposed in the previous chapter might provide a possibility to go beyond the linear theory.

3.1.2 Galois action on modular graphs. If $G$ is a finite index subgroup of PSL$(2, \mathbb{Z})$, then the projection $\pi_G$ is a Belyi map, i.e. it is a covering map from the Riemann surface $G\backslash \mathbb{H}$ to the modular orbifold $\mathcal{M}$. The latter is conformally equivalent to $\mathbb{C}$. Compactifying by adding a point at infinity, we get the Riemann sphere and the orbifold covering

$$\pi_G : G\backslash \mathbb{H} \to \mathcal{M}$$

can be viewed as a branched covering

$$\pi_G : G\backslash \mathbb{H} \to \mathbb{P}^1(\mathbb{C})$$

branched at elliptic points and at infinity. Branching at elliptic points is restricted whereas there is no restriction on the branching above the point at infinity. This is no loss of generality as there is a degree 6 covering from $\mathbb{P}^1 \setminus \{0, 1, \infty\}$ to the modular orbifold. By Belyi’s theorem this branched covering admits a model whose defining polynomials have coefficients from a number field (i.e. a finite extension of $\mathbb{Q}$)

In particular, $G\backslash \mathbb{H}$ is defined over a number field. Such curves will be called arithmetic curves. This allows us to define an action of the absolute Galois group on $\text{FCov}(\mathcal{M})$, or equivalently on the category of modular graphs.

By Belyi’s theorem every arithmetic curve arises as a covering of $\mathbb{P}^1 \setminus \{0, 1, \infty\}$ and hence the modular orbifold: arithmetic curves are precisely the compactifications of the covering curves inside $\text{FCov}(\mathcal{M})$. As we already noted above, every arithmetic curve will appear in $\text{FCov}(\mathcal{M})$ in infinitely many ways, i.e. $\mathbb{P}^1(\mathbb{C})$ will appear as many times as the number of trivalent graphs drawn on the sphere. The last example illustrates why we consider the action of the Galois group on the coverings (curve plus the covering), and not just on the curves. The arithmetic curve here is just $\mathbb{P}^1(\mathbb{C})$ which is defined over $\mathbb{Q}$ and so the action is trivial, whereas the action becomes faithful when we consider the coverings as well.

The action of $G_\mathbb{Q}$ preserves (see [14]):

1. the number of edges in the modular graph, equivalently, the degree of the branched covering;

---

A converse to this claim was first stated by Belyi whose proof relied entirely on the article by Weil, [25]. Experts in the area believed that some explanation was necessary, see [26].
A Panorama of the fundamental group of the modular orbifold

(2) the genus of the arithmetic curve;
(3) the degree distribution of vertices of type ◦, equivalently, the branching behavior above ◦ (recall that this degree can be only 1 or 2);
(4) the degree distribution of vertices of type ●, equivalently, the branching behavior above ● (recall that this degree can be only 1 or 3);
(5) the degree distribution of right-turn cycles, equivalently, the branching behavior above the point at infinity (there is no restriction on this);
(6) monodromy group.

As the aim is to understand \( G_\mathbb{Q} \), one looks for families of arithmetic curves supporting a faithful action of \( G_\mathbb{Q} \). It appears that there are many candidates for this. Motivated by the above list one may first try fixing the genus, \( g \).

In fact, for any non-trivial element \( \sigma \) of \( G_\mathbb{Q} \) one can easily find an algebraic number, say \( \alpha \), on which \( \sigma \) acts non-trivially. Then, if \( E \) is an elliptic curve with \( j \) invariant \( \alpha \), then \( E \) and \( E^\sigma \) are distinct elliptic curves. Hence genus 1 arithmetic curves admit a faithful \( G_\mathbb{Q} \) action. By Belyi’s theorem every such curve corresponds to a modular graph of genus 1. The conclusion is that the Galois action on the set of modular graphs of genus 1 is faithful. One can also show that the same is true for genus 0 curves. More generally, we have:

**Theorem 3.1** \((\mathbb{8})\). Fix a non-negative integer \( g \). The action of the absolute Galois group on the set of genus \( g \) modular graphs is faithful.

In a different vein, the Galois action sends planar trees to planar trees by the above conservation properties, and we have the result:

**Theorem 3.2** \((\mathbb{22})\). The action of the absolute Galois group on the set of modular graphs which are planar trees is faithful.

There are a myriad of questions related to the arithmetic surrounding the theory built so far, and there is a vast literature about these questions. We refer the interested reader to \([\mathbb{9}]\) for further reading. We do not pursue here this “finite” side of the theory further, which is devoted to the study of \( \text{FCov}(\mathcal{M}) \) of \textit{finite} coverings of the modular orbifold, and we turn our attention to the study of some special infinite covers.

### 3.2 Arithmetic of non-tame subgroups

At this stage, the reader will probably expect us to define a Galois action on \( \text{FCov}(\mathcal{X}) \), where \( \mathcal{X} \) is an object in \( \text{Cov}(\mathcal{M}) \) exactly as in the finite case. Likewise for the ambient category \( \text{FGICov}(\mathcal{M}) \). This requires an interpretation of the elements and morphisms of this category as arithmetic objects over
some field, which is possibly some transcendental extension of $\mathbb{Q}$ depending on $X$. We firmly believe that such an interpretation does exist and is fruitful, however, we don’t know how to make this. Instead we will give an alternative arithmetic interpretation of hyperbolic $\mathbb{Z}$-covers from \cite{24}.

3.2.1 Hyperbolic subgroups and binary quadratic forms. Let $W = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ be a hyperbolic element in $\text{PSL}(2, \mathbb{Z})$, i.e. $|p + s| > 2$. The action of $W$ on $\mathbb{H}$ has two fixed points, determined by roots of the following integral equation:

$$rx^2 + (s - p)x - q = 0.$$  

Since solutions of the above equation remains fixed when one takes non-zero multiples of the coefficients, the above equation can be reduced to:

$$\frac{r}{\delta}x^2 + \frac{s - p}{\delta}x - \frac{q}{\delta} = 0,$$

where $\delta$ is the greatest common divisor of $r, s-p$ and $q$. Note that hyperbolicity of the matrix implies that the equation has two real roots. There is a unique geodesic in $\mathbb{H}$ joining these two fixed points, called the geodesic of $W$. Indeed, the action of $W$ on $\mathbb{H}$ is nothing but a translation along this geodesic.

The projectivization of the left hand side of Equation 3.5, namely

$$\frac{r}{\delta}x^2 + \frac{s - p}{\delta}xy - \frac{q}{\delta}y^2$$

is then an indefinite binary quadratic form, i.e. a homogeneous of degree two element of $\mathbb{Z}[x, y]$ of positive discriminant. This form will be denoted by $f_W$. Conversely, given any triple $(a,b,c)$ with $b^2 - 4ac$ being positive and the greatest common divisor of $a$, $b$ and $c$ being 1, by solving a suitable Pell equation, one can find a hyperbolic element of $\text{PSL}(2, \mathbb{Z})$, $W$, so that

$$f_W = ax^2 + bxy + cy^2.$$ 

On the other hand, $\text{PSL}(2, \mathbb{Z})$ acts on the set of indefinite binary quadratic forms by change of variable. This action transforms into the conjugation action in the language of subgroups. We have:

\textbf{Theorem 3.3} (\cite{24} Theorem 3.1). \textit{There is a one to one correspondence between $\text{PSL}(2, \mathbb{Z})$-classes of binary quadratic forms and conjugacy classes of subgroups of $\text{PSL}(2, \mathbb{Z})$ generated by one hyperbolic element.}\n
As described in Example \cite{25} such modular graphs have a unique loop (called the spine), and finitely many Farey branches expanding in the direction of both connected components of the boundary of the annulus, $(W) \setminus \mathbb{H}$. 
Given two forms $f_1 = a_1x^2 + b_1xy + c_1y^2$ and $f_2 = a_2x^2 + b_2xy + c_2y^2$ there is a product defined by Gauß, [7], if and only if either $\Delta_1/\Delta_2 = (b_2^2 - 4a_1c_1)/(b_2^2 - 4a_2c_2)$ or $\Delta_2/\Delta_1$ is a perfect square. In particular the set of all forms of discriminant $\Delta$ is a group under this operation. The function mapping a binary quadratic form $f = ax^2 + bxy + cy^2$ of discriminant $\Delta = b^2 - 4ac$ to the narrow ideal class generated by $(1, \omega)$, where $\omega = \frac{b+\sqrt{\Delta}}{2a}$ in the quadratic number field $\mathbb{Q}(\sqrt{\Delta})$ is a group isomorphism. Therefore, given a square-free positive integer $\Delta$, finding the class number of the number field $\mathbb{Q}(\sqrt{\Delta})$, i.e. the size of the class group of $\mathbb{Q}(\sqrt{\Delta})$, is equivalent to finding the number of inequivalent $\text{PSL}(2, \mathbb{Z})$ classes of indefinite binary quadratic forms of same discriminant.

The latter set of problems, which are now older than 200 years, the class number problems of Gauß, and are all stated in Disquisitiones Arithmeticae, [7]. We must mention that Gauß was not able to solve the analogous problems for negative discriminant but guessed almost all of class number one.

There are further arithmetic problems concerning indefinite binary quadratic forms:
(1) minimum problem: find the smallest positive integer $N$ which is attained, and

(2) representation problem: given an integer $N$ determine whether the equation $f = ax^2 + bxy + cy^2 = N$ has integral solutions, and if exists how many.

Both of the solutions have beautiful interpretations and algorithmic solutions in terms of corks, see [28]. Moreover the corresponding geodesics and above arithmetic problems are visualized in the application InfoMod, see Figure [10] by the authors and H. Ayral.

Acknowledgements. We are thankful to Athanase Papadopoulos for inviting us to publish in this volume and for his comments on the previous versions of this manuscript.

References


A Panorama of the fundamental group of the modular orbifold


