Mathematica Slovaca

Gareth A. Jones

Maps on surfaces and Galois groups

*Mathematica Slovaca*, Vol. 47 (1997), No. 1, 1--33

Persistent URL: http://dml.cz/dmlcz/136695

Math. Slovaca, 47 (1997), No. 1, 1–33

Mathematica
Slovaca
©1997
Mathematical Institute
Slovak Academy of Sciences

# MAPS ON SURFACES AND GALOIS GROUPS

Gareth A. Jones

(*Communicated by Martin Škoviera*)

ABSTRACT. A brief survey of some of the connections between maps on surfaces, permutations, Riemann surfaces, algebraic curves and Galois groups is given.

My aim here is to give a brief survey of some of the connections between maps on surfaces, permutations, Riemann surfaces, algebraic curves and Galois groups. Some of these connections are very well-known, some are surprisingly old, and others are quite new. Taken together, they provide a good illustration of the essential unity of modern mathematics. Indeed, there are further links with other topics such as Teichmüller theory and conformal field theory, which are beyond my capacity to explain here.

Many people have discovered or re-discovered parts of this theory, but the first person with the imagination to see the whole picture was Grothendieck, inspired by a theorem proved by Belyĭ [2] in 1979. In its most basic form, the theory asserts that Galois groups of algebraic number fields have faithful representations on maps on surfaces, sometimes called *dessins d'enfants.* Grothendieck outlined this (and much more) in 1984, in his *Esquisse d'un Programme* [12], a research proposal full of tantalizing ideas and conjectures, sometimes precise and sometimes vague, but always highly original. The proposal was not successful, he abandoned the project, and for a while, that seemed to be the end of the story. However, in recent years, researchers from a number of countries, and from a wide variety of disciplines, have started to piece together the various clues Grothendieck has left, and a powerful general theory is now emerging. Here, I will concentrate mainly on the combinatorial and algebraic aspects, with particular emphasis on embeddings of graphs, and

AMS Subject Classification (1991): Primary 05C10.
Key words: surface, Riemann surface, algebraic curve, Galois group, bipartite graph.

This is an expanded version of a talk given at the Workshop on Graph Embeddings and Maps on Surfaces at Donovaly, Slovakia, 1994; the author is grateful to the organizers for the warmth of their hospitality, and to all the participants for the stimulating atmosphere of the Workshop.

especially bipartite graphs. The experts on *dessins d'enfants* should be familiar with the theorems stated here, but some of the examples I have provided may be new. For a rather more general introductory survey, see [22], and for a recent series of specialist papers, see [31].

## 1. Bipartite maps and permutations

It has long been known that maps on surfaces can be represented by permutations. There are many ways of doing this (almost as many as there are people doing it), but nevertheless, these methods all have many features in common. For simplicity, I will concentrate on one of the most basic situations, the embedding of a bipartite graph in an oriented surface, but I will occasionally indicate where generalizations are possible.

Let $X$ be a compact, connected, oriented surface without boundary. (From now on, I will simply use the word "surface" to denote such an object, except when the context obviously indicates otherwise.) Let $\mathcal{B}$ be a bipartite map on $X$, that is, a 2-cell embedding of a finite bipartite graph $\mathcal{G}$ in $X$. One can colour the vertices of $\mathcal{G}$ black and white, so that all edges connect vertices of different colours. (There are just two ways of doing this, differing by transposition of the colours, and the particular choice is usually unimportant.) At each black vertex, the chosen orientation of $X$ induces a cyclic permutation of the incident edges; since each edge meets a unique black vertex, these local rotations are the disjoint cycles of a permutation $g_0$ of the set $E$ of edges of $\mathcal{B}$. Similarly, the local rotations around the white vertices determine a permutation $g_1$ of $E$, as shown in Figure 1.

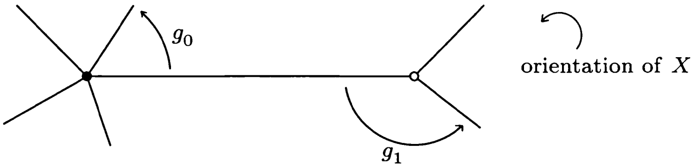

FIGURE 1.

These two permutations generate a subgroup $G$ of the symmetric group $S^E$ of all permutations of $E$, called the *monodromy group* of $\mathcal{B}$ (the reason for this name will become clear later): the elements of $G$ are just the finite products of powers of $g_0$ and $g_1$. Our topological assumptions imply that $\mathcal{G}$ must be connected, so that $G$ acts transitively on $E$. (Note that $G$ is not, in general, a group of automorphisms of $\mathcal{B}$, or even of $\mathcal{G}$, since $g_0$ and $g_1$ do not preserve incidence.) It is straightforward to check that the faces of $\mathcal{B}$ correspond to the cycles of the permutation $g_\infty := (g_0g_1)^{-1}$: each cycle of length $l$ corresponds to a $2l$-gonal face. One can reverse this process, so that every 2-generator transitive subgroup of the symmetric group $S_N$ represents an oriented bipartite map: the edges are the $N$ symbols permuted, and the cycles of the two generators determine the sets of black and white vertices together with the local rotations of edges around them.

If bipartite maps $\mathcal{B}$ and $\mathcal{B}'$ are defined by pairs $g_i$ and $g_i'$ of permutations, then a morphism $\mathcal{B} \to \mathcal{B}'$ (preserving orientations and vertex-colours) is a function $\phi\colon E \to E'$ between their edge-sets such that $g_i\phi = \phi g_i'$ for $i = 0, 1$. In particular, an automorphism of $\mathcal{B}$ (as a bicoloured oriented map) is a permutation of $E$ commuting with each $g_i$, or equivalently with $G$. These form a group $\mathrm{Aut}_0\,\mathcal{B}$, the centralizer of $G$ in the symmetric group $S^E$, which preserves the orientation of $X$ and the colouring of the vertices. (This group should be distinguished from the possibly larger group $\mathrm{Aut}\,\mathcal{B}$ of automorphisms of the *uncoloured* oriented map $\mathcal{B}$, which may permute the vertex-colours; these are the elements of $S^E$ which conjugate the set $\{g_0, g_1\}$ to itself. Similarly, one could also allow automorphisms which reverse the orientation by inverting these generators $g_i$, but, in this paper, I will restrict attention to orientation-preserving automorphisms.) I shall call $\mathcal{B}$ *bipartite-regular* if $\mathrm{Aut}_0\,\mathcal{B}$ is as large as possible. There are several equivalent ways of expressing this more precisely: $\mathrm{Aut}\,\mathcal{B}$ acts transitively on $E$, $\mathrm{Aut}\,\mathcal{B}$ has order $N$ where $N = |E|$, or $G$ is a regular permutation group (transitive, of minimum order $N$), and under these conditions (though not otherwise), the groups $\mathrm{Aut}\,\mathcal{B}$ and $G$ are isomorphic. (This definition of regularity is a little weaker than the concept of regularity of an uncoloured map, which I will discuss in the next section.)

Examples.

(1) Let $\mathcal{G}$ be the complete bipartite graph $K_{m,n}$; this has $m$ black vertices $v_1, \ldots, v_m$ and $n$ white vertices $w_1, \ldots, w_n$, with a single edge $e_{ij} = v_iw_j$ between each pair $v_i$ and $w_j$, so that there are $N = mn$ edges in all. Let us use the numbering of the vertices to define the obvious local rotations of edges, as in Figure 2: we let $g_0$ have disjoint cycles $(e_{i1}, \ldots, e_{in})$, one for each black vertex $v_i$, and similarly, $g_1 = (e_{11}, \ldots, e_{m1}) \ldots (e_{1n}, \ldots, e_{mn})$. These permutations commute, and they generate a group $G = \langle g_0\rangle \times \langle g_1\rangle \cong C_n \times C_m$ of order $N$ which permutes the edge-set $E$ regularly; thus the bipartite map $\mathcal{K}_{m,n}$ which they define is bipartite-regular with $\mathrm{Aut}\,\mathcal{K}_{m,n} \cong C_n \times C_m$. Since the permutation $g_\infty = (g_0g_1)^{-1}$ has cycles of length $[m,n]$, the least common multiple of $m$ and $n$, the faces of this map are all $2[m,n]$-gons; the number of faces is therefore $2N/2[m,n] = (m,n)$, the highest common factor of $m$ and $n$, so $\mathcal{K}_{m,n}$ has Euler characteristic $(m+n) - mn + (m,n)$ and genus $\big((m-1)(n-1)+1-(m,n)\big)/2$. In particular, $\mathcal{K}_{n,n}$ has genus $(n-1)(n-2)/2$.

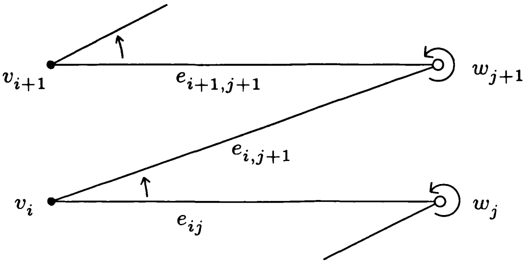

FIGURE 2. $\mathcal{K}_{m,n}$.

(2) Let $\mathcal{G}$ be the $n$-dimensional cube graph $Q_n$, where $n \geq 2$. The vertex-set $V$ is the $n$-dimensional vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$, or equivalently, an elementary abelian group of order $2^n$. The graph is the Cayley diagram for $V$ with respect to its standard basis $e_1, \ldots, e_n$: two vertices are joined by an edge if they differ (as vectors) in just one coordinate place, so each vertex $v$ is incident with $n$ edges $vw$, which we can label $j = 1, \ldots, n$ as $w = v + e_j$. By colouring each vertex black or white as the sum of its coordinates is $0$ or $1$ in $\mathbb{Z}_2$, we see that $\mathcal{G}$ is bipartite, with the black vertices forming a subgroup $V_0$ of index $2$ in $V$, that is, a subspace of codimension $1$. Around each black vertex, let the rotation $g_0$ be given by the cyclic ordering $(1, 2, \ldots, n)$ of the edge-labels, and around each white vertex, let $g_1$ use the inverse ordering $(n, n-1, \ldots, 1)$. Then all cycles of $g_\infty$ have length $2$, so we obtain a 4-gonal embedding $\mathcal{Q}_n$ of $Q_n$, each face being incident with four vertices $v$, $v + e_j$, $v + e_j + e_{j-1}$, $v + e_{j-1}$ in this cyclic order or its inverse, as in Figure 3. (The subscripts $j$ are regarded as elements of $\mathbb{Z}_n$, so that $e_0 = e_n$.)

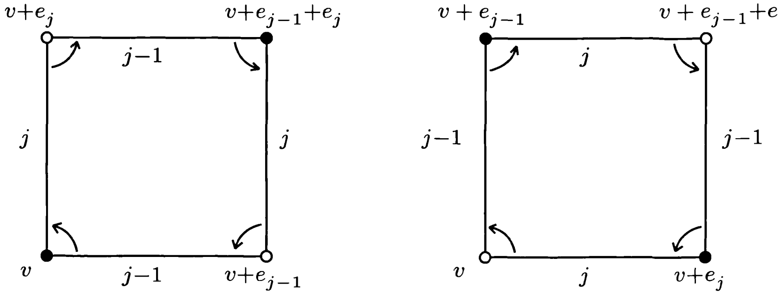

FIGURE 3. $Q_n$.

Since $Q_n$ is bipartite and all the faces are 4-gons, this embedding of $\mathcal{Q}_n$ has the maximum number of faces, so it must be of minimum genus: there are $2^n$ vertices, $n2^{n-1}$ edges and $n2^{n-2}$ faces, so the characteristic is $(4-n)2^{n-2}$ and the genus is $1+(n-4)2^{n-3}$. The automorphism group $\operatorname{Aut}\mathcal{Q}_n$ has a normal subgroup $V_0 \cong C_2^{n-1}$ which permutes each of the two monochrome sets of vertices regularly; this is generated by the half-turns $h_j$ $(j=1,2,\ldots,n)$ about the mid-points of the faces $0$, $e_{j-1}$, $e_{j-1}+e_j$, $e_j$, which permute the vertices by $v \mapsto v+e_{j-1}+e_j$. This subgroup is complemented by a cyclic group of order $n$, generated by an automorphism which fixes $0$ and acts on its adjacent vertices by $e_j \mapsto e_{j+1}$. It follows that $\operatorname{Aut}_0\mathcal{Q}_n$ permutes the edges transitively, so that $\mathcal{Q}_n$ is bipartite-regular. When $n=3$, for example, $\mathcal{Q}_3$ is the cube, the black and white vertices are the vertices of two inscribed tetrahedra, and $\operatorname{Aut}\mathcal{Q}_3$ is the rotation group of each of them, isomorphic to $A_4$. In the case $n=4$, we obtain the torus map $\{4,4\}_{4,0}$, shown in Figure 4 with opposite edges identified; it is one of an infinite class of torus maps described by Coxeter and Moser in [10; §8.3]. This general construction of $\mathcal{Q}_n$ is based on that given by Biggs and White in [5; §5.6].

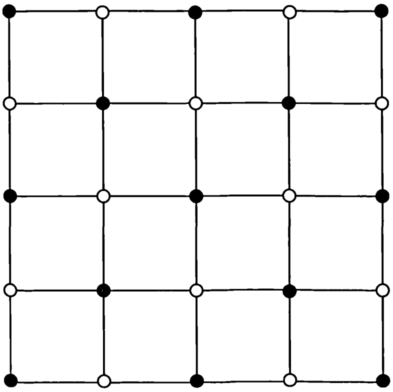

FIGURE 4. $Q_4 = \{4,4\}_{4,0}$.

(3) The map $\mathcal{Q}_n$ is not the only bipartite-regular embedding of $Q_n$. If we define the rotations $g_0$ and $g_1$ by using the *same* cyclic ordering $(1,2,\ldots,n)$ of the edge-labels, as in Figure 5, we find that $g_\infty$ now has cycles of length $n$, so the faces are $2n$-gons and the genus is $1+(n-3)2^{n-2}$. For instance, $\mathcal{Q}'_3$ is the torus map $\{6,3\}_{2,0}$ of [10; §8.4], shown in Figure 6 with opposite edges identified. One can obtain $\mathcal{Q}'_n$ from $\mathcal{Q}_n$ (and vice versa) by applying the Petrie operation: the underlying graph $Q_n$ is preserved, but the faces of $\mathcal{Q}_n$ are replaced with its Petrie polygons ([10]).

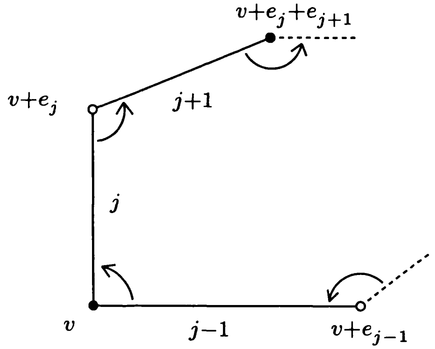

FIGURE 5. $Q'_n$.

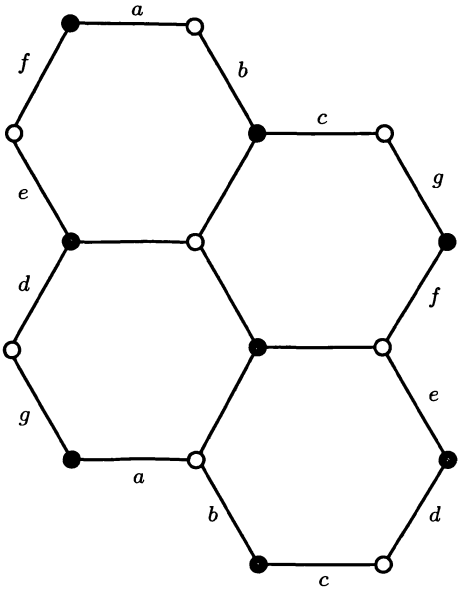

FIGURE 6. $Q'_3 = \{6,3\}_{2,0}$.

These are bounded by closed zig-zag paths which turn alternately left and right on $Q_n$ at the vertices, so this operation is equivalent to replacing the generating pair $g_0, g_1$ with the pair $g_0, g_1^{-1}$. It therefore preserves both the monodromy group $G = \langle g_0, g_1 \rangle$ and the automorphism group (the centralizer of $G$), so in particular it follows that $Q'_n$ is bipartite-regular, with $\mathrm{Aut}\, Q'_n \cong \mathrm{Aut}\, Q_n$. (It is a curious fact that $Q'_n$ and $Q_{n+1}$ have the same genus, so they can be drawn on the same surface. For instance, Figure 7 shows how to convert $\mathcal{Q}_3'$ into $\mathcal{Q}_4$ by replacing each of its four hexagons with two squares and four half-squares. It would be interesting to have a more general explanation of this phenomenon.)

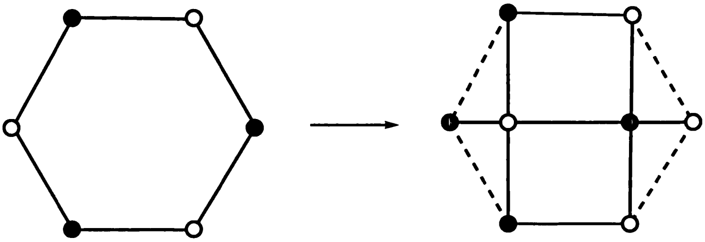

FIGURE 7.

## 2. Maps and permutations

When using permutations to describe maps, one is not restricted to embeddings of bipartite graphs. Let $\mathcal{M}$ be an oriented map on $X$, where the underlying graph $\mathcal{G}$ is finite but not necessarily bipartite. One can form a bipartite map $\mathcal{B}$ from $\mathcal{M}$ by first colouring all the vertices of $\mathcal{M}$ black, and then inserting a white vertex of valency 2 in each edge of $\mathcal{M}$. Thus each edge $e$ of $\mathcal{M}$ yields two edges of $\mathcal{B}$, which can be regarded as directed edges of $\mathcal{M}$ pointing in either direction along $e$. By applying the method of the previous section to $\mathcal{B}$, we obtain a transitive group $C = \langle c_0, c_1 \rangle \leq S_{2N}$ of permutations of these directed edges, where $c_0$ and $c_1$ are the rotations around the black and white vertices, as shown in Figure 8.

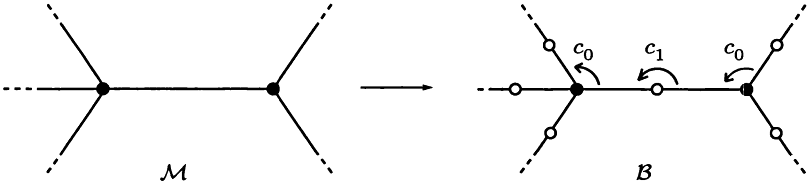

FIGURE 8.

This permutation group $C$ is the monodromy group of $\mathcal{B}$, also called the *cartographic group* of $\mathcal{M}$. By our construction, $c_1^2 = 1$. Any transitive permutation group $C = \langle c_0, c_1 \mid c_1^2 = 1, \dots \rangle$ arises in this way: the vertices, edges and faces of $\mathcal{M}$ are the cycles of $c_0$, $c_1$ and $c_\infty := (c_0c_1)^{-1}$, with incidence given by non-empty intersection. (To allow for the cases where the involution $c_1$ has fixed-points, one has to allow $\mathcal{M}$ to have "free edges", or "half-edges" $e$, incident with only one vertex of $\mathcal{M}$; in forming $\mathcal{B}$, one places a white vertex of valency 1 at the other end of $e$, giving a single edge of $\mathcal{B}$ fixed by $c_1$.) Morphisms and automorphisms are defined as before as functions and permutations commuting with the generators $c_i$, so in particular the automorphism group $\mathrm{Aut}\,\mathcal{M}$ of $\mathcal{M}$ is the centralizer of $C$ in $S_{2N}$. We say that $\mathcal{M}$ is *regular* (as an oriented but uncoloured map) if $\mathrm{Aut}\,\mathcal{M}$ has maximal order $2N$, so that it acts transitively on the directed edges, in which case $\mathrm{Aut}\,\mathcal{M} \cong C$. (Note that if $\mathcal{M}$ happens to be bipartite, then $\mathrm{Aut}\,\mathcal{M}$ may be strictly larger than $\mathrm{Aut}\,\mathcal{M}$, since there may be automorphisms which transpose the two sets of vertices.)

EXAMPLES.

(1) The most familiar examples of maps are the regular solids, all of which are regular maps on the sphere: the tetrahedron has automorphism group isomorphic to the alternating group $A_4$, the cube and the octahedron have $\mathrm{Aut}\,\mathcal{M} \cong S_4$, while the icosahedron and the dodecahedron have $\mathrm{Aut}\,\mathcal{M} \cong A_5$. This last group is the smallest of the non-abelian finite simple groups; Malle, Saxl and Weigel [25] have recently shown that every non-abelian finite simple group $S$ can be generated by two elements, one of them of order 2, so $S \cong \mathrm{Aut}\,\mathcal{M}$ for some regular map $\mathcal{M}$.

(2) For some non-regular examples, again of genus 0, consider the maps $\mathcal{M}$ in Figure 9 (the first contains two half-edges).

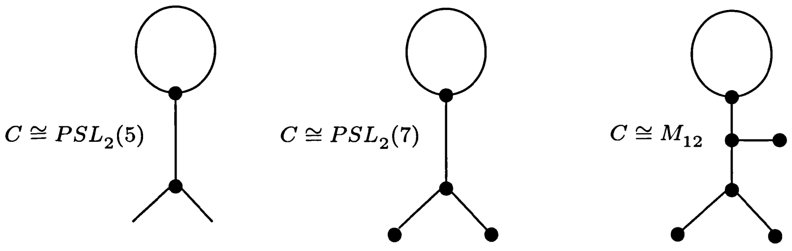

FIGURE 9.

The corresponding bipartite maps $\mathcal{B}$ have 6, 8 and 12 edges, and the cartographic groups $C$ are isomorphic to the projective special linear groups $PSL_2(5)$ ($\cong A_5$) and $PSL_2(7)$, and the Mathieu group $M_{12}$, simple groups of orders 60, 168 and 95040. For example, Figure 10 shows the construction of $\mathcal{B}$ for the first map $\mathcal{M}$; the edges of $\mathcal{B}$ are labelled with the elements $0, 1, 2, 3, 4, \infty$ of the projective line $PG_1(5)$ over the field $\mathbb{Z}_5$, so that $c_0$, $c_1$ and $c_\infty$ induce the projective transformations $z \mapsto 1/(1-z)$, $z \mapsto 1/z$ and $z \mapsto z+1$ which generate $PSL_2(5)$.

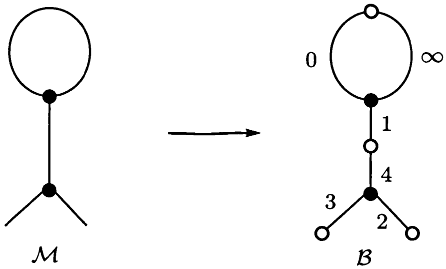

FIGURE 10.

(3) If we ignore their vertex-colouring, we can treat our bipartite maps $\mathcal{K}_{m,n}$ in the above way: each edge $e_{ij}$ yields two directed edges $(v_i, w_j)$ and $(w_j, v_i)$, directed towards $v_i$ and $w_j$ respectively, so the cartographic group $C$ is a subgroup of $S_{2N} = S_{2mn}$. The effect of $c_1$ is to reverse each directed edge, transposing $(v_i, w_j)$ with $(w_j, v_i)$, while $c_0$ sends $(v_i, w_j)$ to $(v_i, w_{j+1})$ and $(w_j, v_i)$ to $(w_j, v_{i+1})$, where we regard the subscripts $i$ and $j$ as elements of $\mathbb{Z}_m$ and $\mathbb{Z}_n$ respectively. It follows that $C$ is isomorphic to the wreath product $C_{[m,n]} \wr C_2$ of $C_{[m,n]}$ by $C_2$: this has a normal subgroup $C_{[m,n]} \times C_{[m,n]}$ (generated by $c_0$ and its conjugate $c_1c_0c_1$), complemented by a subgroup $C_2$ (generated by $c_1$) which transposes the two direct factors $C_{[m,n]}$ by conjugation. If $m \neq n$, then $\mathcal{K}_{m,n}$ is not regular (as an *uncoloured* map), even though it is bipartite-regular; indeed $K_{m,n}$ can have no regular embeddings for $m \neq n$, since its automorphism group is not transitive on the vertices. If $m = n$, on the other hand, then $C$ permutes the directed edges regularly, so $\mathcal{K}_{n,n}$ is a regular map with $\operatorname{Aut}\mathcal{K}_{n,n} \cong C_n \wr C_2$.

(4) The $n$-cube embedding $\mathcal{Q}_n$ is another example of a regular map. In this case, $\operatorname{Aut}\mathcal{Q}_n$ has a normal subgroup $V_0 \cong C_2^{n-1}$ preserving the two vertex-colours and the $n$ edge-labels, with quotient-group a dihedral group $D_n$ of order $2n$. The automorphisms mapping onto the subgroup $C_n \leq D_n$ preserve the colours and the cyclic ordering of the labels, while the other automorphisms transpose the colours and reverse the cyclic ordering. When $n$ is odd, this extension splits: one can take a complement $D_n$ for $V_0$ to consist of those automorphisms which either fix or transpose the antipodal pair of vertices 0 and $e_1 + \cdots + e_n$, which are respectively black and white. When $n$ is even, however, these vertices are both black, and the extension does not split.

(5) The embedding $\mathcal{Q}'_n$ of $Q_n$ is also regular, but its automorphism group differs from that of $\mathcal{Q}_n$. (This should not be surprising: a permutation which

preserves the set $\{g_0, g_1\}$ by conjugation need not preserve $\{g_0, g_1^{-1}\}$.) In fact, Aut $\mathcal{Q}'_n$ is isomorphic to $C_2 \wr C_n$, a split extension of a normal subgroup $V \cong C_2^n$, which permutes the vertices regularly, by the stabilizer $C_n$ of the vertex $0$. However, if one includes orientation-reversing automorphisms, then both $\mathcal{Q}_n$ and $\mathcal{Q}'_n$ have the same automorphism group, namely the wreath product $C_2 \wr D_n$ of order $2^n.2n$, where $D_n$ acts naturally with degree $n$ on the direct factors of the base group.

This general method of representing oriented maps by permutations is quite old: it was used by Heffter in the last century [14], [15] to study embeddings of complete graphs, and the earliest example I can find is Hamilton's use of the method in 1856 to construct what we now call Hamiltonian circuits in the icosahedral graph [13]. He was well aware that all of the regular polyhedra could be described in this way, but it is not clear whether he ever considered any other maps. This general theory was developed independently in the 1970s by Malgoire and Voisin [24] and by Jones and Singerman [19], [36]; Chapter 8 of [10] gives a detailed treatment of regular maps and their automorphism groups.

One can extend this method to deal with non-orientable maps and with orientation-reversing automorphisms of orientable maps ([17], [38], [39]). The objects permuted are now *flags* consisting of a vertex, edge and face, all mutually incident. There are three generating permutations $r_0$, $r_1$ and $r_2$, each $r_i$ changing the $i$-dimensional component of each flag (in the only possible way) while fixing the other two components; thus $r_i^2 = (r_0 r_2)^2 = 1$. In the case of an oriented map, one can identify its cartographic group $C$ with the subgroup of index 2 in $\langle r_0, r_1, r_2 \rangle$ consisting of the words of even length in the generators $r_i$. By allowing the generators to have fixed-points, one can extend the theory further to include maps on surfaces with boundary ([4]). In this paper, I am mainly interested in Riemann surfaces, so I will restrict my attention from now on to the case of oriented maps without boundary.

## 3. Maps on Riemann surfaces

A *Riemann surface* is a surface with locally-defined complex coordinates, such that the changes of coordinates between intersecting neighbourhoods are conformal. Examples include the complex plane $\mathbb{C}$, the upper half-plane $\mathcal{U} = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$, and the Riemann sphere (or complex projective line) $\Sigma = P^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$. (One identifies $\Sigma \setminus \{\infty\}$ with $\mathbb{C}$ by stereographic projection, and uses the local coordinates $1/z$ near $\infty$.) Up to isomorphism, these are the only simply-connected Riemann surfaces.

The upper half-plane is a model of hyperbolic geometry, the geodesics being the euclidean lines and semi-circles which meet the real line $\mathbb{R}$ at right-angles. The *modular group* $\Gamma = PSL_2(\mathbb{Z})$ consisting of the Möbius transformations

$$T\colon z \mapsto \frac{az+b}{cz+d} \qquad (a,b,c,d \in \mathbb{Z},\ \ ad-bc=1)$$

acts on $\mathcal{U}$ as a group of orientation-preserving hyperbolic isometries. It also acts (transitively) on the rational projective line $P^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$, and hence it acts on the extended hyperbolic plane

$$\overline{\mathcal{U}} = \mathcal{U} \cup \mathbb{Q} \cup \{\infty\}.$$

One can partition $\mathbb{Q} \cup \{\infty\}$ into three disjoint subsets, which are permuted by $\Gamma$; these are

$$[0] = \left\{\tfrac{a}{b} \in \mathbb{Q} \cup \{\infty\} \mid a \text{ is even and } b \text{ is odd}\right\},$$
$$[1] = \left\{\tfrac{a}{b} \in \mathbb{Q} \cup \{\infty\} \mid a \text{ and } b \text{ are both odd}\right\},$$
$$[\infty] = \left\{\tfrac{a}{b} \in \mathbb{Q} \cup \{\infty\} \mid a \text{ is odd and } b \text{ is even}\right\}.$$

(Here $a/b$ is always in reduced form, and $\infty = 1/0$.) In this action of $\Gamma$, the set-wise stabilizer of $[0]$ is

$$\Gamma_0(2) = \{T \in \Gamma \mid c \equiv 0 \mod (2)\},$$

a non-normal subgroup of index 3 in $\Gamma$; the kernel (stabilizing each of the three sets) is

$$\Gamma(2) = \{T \in \Gamma \mid b \equiv c \equiv 0 \mod (2)\},$$

a normal subgroup of index 6 in $\Gamma$, with

$$\Gamma/\Gamma_0(2) \cong PSL_2(2) \cong S_3.$$

(These are both examples of *congruence subgroups* of $\Gamma$, defined by finite sets of congruences between the coefficients $a$, $b$, $c$ and $d$. In particular, $\Gamma_0(2)$ is the *principal congruence subgroup of level* 2. Congruence subgroups are very important in number theory; they all have finite index in $\Gamma$, but not every subgroup of finite index is a congruence subgroup. This is the negative solution to the congruence subgroup problem for $\Gamma$; see [18] for a survey of this topic.)

The *universal bipartite map* $\hat{\mathcal{B}}$ on $\overline{\mathcal{U}}$ has $[0]$ and $[1]$ as its sets of black and white vertices, and its edges are the hyperbolic geodesics between vertices $a/b$ and $c/d$, where $ad - bc = \pm 1$; this implies that $a$ and $c$ have opposite parity, so the map (part of which is shown in Figure 11) is indeed bipartite.

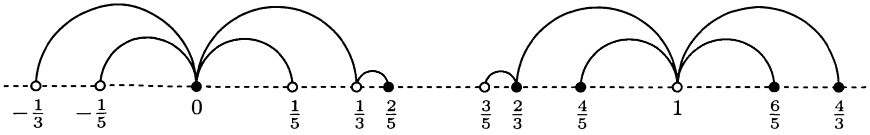

FIGURE 11. $\hat{\mathcal{B}}$.

The automorphism group of $\hat{\mathcal{B}}$ (preserving orientation and colours) is $\Gamma(2)$. This is a free group of rank 2, freely generated by

$$T_0 \colon z \mapsto \frac{z}{-2z+1} \quad \text{and} \quad T_1 \colon z \mapsto \frac{z-2}{2z-3}.$$

It follows that if $\mathcal{B}$ is any bipartite map, with monodromy group $G = \langle g_0, g_1 \rangle$, then there is an epimorphism

$$\Gamma(2) \to G, \quad T_0 \mapsto g_0, \quad T_1 \mapsto g_1,$$

giving a transitive action of $\Gamma(2)$ on the set $E$ of edges of $\mathcal{B}$. The stabilizer of an edge in this action is a subgroup $B$ of index $N = |E|$ in $\Gamma(2)$, called the *map subgroup* corresponding to $\mathcal{B}$ (different choices of an edge lead to conjugate subgroups). Since $B \leq \Gamma(2) = \operatorname{Aut} \hat{\mathcal{B}}$, one can form the quotient map $\hat{\mathcal{B}}/B$, and it is straightforward to prove that this is a bipartite map isomorphic to $\mathcal{B}$. We have $\operatorname{Aut} \mathcal{B} \cong N_{\Gamma(2)}(B)/B$, where $N_{\Gamma(2)}(B)$ is the normalizer of $B$ in $\Gamma(2)$; in particular, $\mathcal{B}$ is bipartite-regular if and only if $B$ is normal in $\Gamma(2)$, in which case $\operatorname{Aut} \mathcal{B} \cong \Gamma(2)/B \cong G$.

EXAMPLES.

(1) When $\mathcal{B} = \mathcal{K}_{m,n}$, we find that $B$ is the normal closure in $\Gamma(2)$ of $T_0^n, T_1^m$ and the commutator $[T_0, T_1] = T_0^{-1} T_1^{-1} T_0 T_1$. This is the kernel of the epimorphism $\Gamma(2) \to G = C_n \times C_m$, $T_i \mapsto g_i$, so it is a normal subgroup corresponding to the fact that $\mathcal{K}_{m,n}$ is bipartite-regular.

(2) When $\mathcal{B} = \mathcal{Q}_n$, we can construct $B$ in two steps: first we take $\tilde{B}$ to be the normal closure in $\Gamma(2)$ of $T_0 T_1$ and $T_0^n$; thus $\Gamma(2)/\tilde{B} \cong C_n$, and $\tilde{B}$ has generators

$$T_0^{-n}, \ U_1 = T_0 T_1, \ U_2 = T_1^{-1} T_0 T_1^2, \ \ldots, \ U_n = T_1^{1-n} T_0 T_1^n, \ T_1^{-n},$$

and a single defining relation

$$T_0^{-n} U_1 U_2 \ldots U_n T_1^{-n} = 1,$$

so $\tilde{B}$ is a free group of rank $n+1$. We then map $\tilde{B}$ onto $V_0 = C_2^{n-1}$ by sending the generators $T_0^{-n}$ and $T_1^{-n}$ to the identity, and making the other generators

$U_i$ commute and have order 2. The map subgroup $B$ is the kernel of this epimorphism, the normal closure in $\tilde{B}$ of $T_0^n, T_1^n$, the commutators $[U_i, U_j]$, and the elements $U_i^2$. This is normal, not just in $\tilde{B}$ but in fact in $\Gamma(2)$, confirming that $\mathcal{Q}_n$ is bipartite-regular. The subgroup $\tilde{B}/B$ of $\Gamma(2)/B$ corresponds to the elementary abelian normal subgroup $V_0$ of $\operatorname{Aut}\mathcal{Q}_n$, while the complement $C_n$ of $V_0$ is generated by the image of $T_0$.

(3) The construction of $B$ is similar for $\mathcal{B} = \mathcal{Q}'_n$, except that we now take $\tilde{B}$ to be the normal closure of $T_0T_1^{-1}$ and $T_0^n$; this is equivalent to applying the automorphism $T_0 \mapsto T_0$, $T_1 \mapsto T_1^{-1}$ of $\Gamma(2)$ to the previous example.

This process constructs an isomorphic copy of our original bipartite map $\mathcal{B}$, endowed with some extra structure. The underlying surface is now a compact Riemann surface $X = \overline{\mathcal{U}}/B$, in which $\mathcal{G}$ is very rigidly embedded: for example, the edges are all geodesics, the angles between successive edges around a vertex are all equal, and the automorphisms of $\mathcal{B}$ are all conformal automorphisms of $\overline{\mathcal{U}}/B$ (induced by the action of $N_{\Gamma(2)}(B)$ on $\hat{\mathcal{B}}$). One can regard $\hat{\mathcal{B}}/B$ as a canonical form for $\mathcal{B}$, since it is a specially chosen representative of the isomorphism class of $\mathcal{B}$.

Instead of obtaining this canonical form as a quotient of the universal bipartite map $\hat{\mathcal{B}}$, one can also obtain it as a branched covering of the trivial bipartite map. For any bipartite map $\mathcal{B}$, the inclusions $1 \leq B \leq \Gamma(2)$ induce coverings

$$\hat{\mathcal{B}} \to \hat{\mathcal{B}}/B \cong \mathcal{B} \to \hat{\mathcal{B}}/\Gamma(2) \cong \mathcal{B}_1,$$

where $\mathcal{B}_1$ is the *trivial bipartite map* on the sphere $\Sigma \cong \overline{\mathcal{U}}/\Gamma(2)$ with one black vertex (at 0), one white vertex (at 1), one edge (the closed interval $I = [0,1]$ in $\mathbb{R}$), and one face $\Sigma \setminus I$. An elegant way of obtaining $\mathcal{B}_1$ from $\hat{\mathcal{B}}$ is via the $\lambda$-function. This is an analytic function $\lambda\colon \mathcal{U} \to \mathbb{C}$ with the property that $\lambda(z) = \lambda(z')$ if and only if $z$ and $z'$ are equivalent under $\Gamma(2)$; if we extend $\lambda$ to take the values 0, 1 and $\infty$ on $[0]$, $[1]$ and $[\infty]$, we get a function $\overline{\mathcal{U}} \to \Sigma$ which maps $\hat{\mathcal{B}}$ onto $\mathcal{B}_1$. I will define the $\lambda$-function more precisely in §5; for more on this and related functions, see books on complex function theory, such as [1], [20].

The $N$-sheeted covering $\beta\colon \mathcal{B} \to \mathcal{B}_1$, unbranched outside $\{0,1,\infty\}$, is a meromorphic function $X = \overline{\mathcal{U}}/B \to \Sigma$ of degree $N$ with no critical values outside $\{0,1,\infty\}$. Such a function is called a *Belyĭ function*, and $(X,\beta)$ is called a *Belyĭ pair*. We have seen how bipartite maps give rise to Belyĭ pairs, and the converse is also true. If $(X,\beta)$ is a Belyĭ pair, then $\mathcal{B} = \beta^{-1}(\mathcal{B}_1)$ is a bipartite map on $X$: the black and white vertices of $\mathcal{B}$ are the elements of $\beta^{-1}(0)$ and $\beta^{-1}(1)$, the embedded graph $\mathcal{G}$ is $\beta^{-1}(I)$, and the faces of $\mathcal{B}$ are the connected components of $\beta^{-1}(\Sigma \setminus I)$ in one-to-one correspondence with the poles (the elements of $\beta^{-1}(\infty)$). By identifying each edge of $\mathcal{G}$ with the sheet of

$\beta$ containing it, one can identify $G$ with the monodromy group of $\beta$ (the action of $\pi_1(\Sigma \setminus \{0,1,\infty\}) \cong \Gamma(2)$ on the sheets induced by lifting closed paths via $\beta$ from $\Sigma \setminus \{0,1,\infty\}$ to $X$); then $g_0$, $g_1$ and $g_\infty$ are the permutations induced by loops in $\Sigma$ around $0$, $1$ and $\infty$.

One can apply similar techniques to maps in general (see [17], [19], [24], [36] for early versions of this theory). The *universal map* $\hat{\mathcal{M}}$, introduced by Singerman in [37], is simply $\hat{\mathcal{B}}$ without the colouring of its vertices, and its automorphism group is $\Gamma_0(2)$. (The subgroup $\Gamma(2)$ of index 2 preserves the sets $[0]$ and $[1]$, so it preserves the vertex-colours, while the other coset $\Gamma_0(2) \setminus \Gamma(2)$ transposes them.) Now $\Gamma_0(2)$ is generated by

$$U_0 \colon z \mapsto \frac{z}{-2z+1} \quad \text{and} \quad U_1 \colon z \mapsto \frac{z-1}{2z-1}$$

with a single defining relation $U_1^2 = 1$, so if $\mathcal{M}$ is a map with cartographic group $C = \langle c_0, c_1 \mid c_1^2 = 1, \dots \rangle \leq S_{2N}$, then there is an epimorphism $\Gamma_0(2) \to C$ given by $U_i \mapsto c_i$. One can then reconstruct $\mathcal{M}$ as the quotient $\hat{\mathcal{M}}/M$, where $M$ is a point-stabilizer in this action of $\Gamma_0(2)$. The details are similar to those in the bipartite case: for instance, the fact that $\mathcal{K}_{m,n}$ is bipartite-regular but not regular when $m \neq n$ corresponds to the fact that its map subgroup $B$ is normal in $\Gamma(2)$ but not in $\Gamma_0(2)$. Similarly, the fact that $\mathrm{Aut}\,\mathcal{Q}_n \cong \mathrm{Aut}_0\,\mathcal{Q}'_n$ whereas $\mathrm{Aut}\,\mathcal{Q}_n \not\cong \mathrm{Aut}\,\mathcal{Q}'_n$ corresponds to the fact that the map subgroups $B$ of these two regular maps have isomorphic quotient-groups $\Gamma(2)/B$ in $\Gamma(2)$, whereas the quotients $\Gamma_0(2)/B$ are not isomorphic. Every map $\mathcal{M}$ can be obtained as a branched covering of the *trivial map* $\mathcal{M}_1 = \hat{\mathcal{M}}/\Gamma_0(2)$ on $\Sigma$: this has a single vertex at $0$, a single half-edge along $I = [0,1]$, and a single face $\Sigma \setminus I$. (In place of $\lambda \colon \overline{\mathcal{U}} \to \Sigma$ one uses the function $4\lambda(1-\lambda)$, which is invariant under $\Gamma_0(2)$.) The covering $\mathcal{M} \to \mathcal{M}_1$ is now a *clean* Belyĭ function $X \to \Sigma$: this is a Belyĭ function with the property that the $2N$ sheets of the covering come together in pairs over the critical value $1 \in \Sigma$, so that the single half-edge of $\mathcal{M}_1$ lifts to $N$ complete edges of $\mathcal{M}$.

## 4. Plane trees and polynomials

Perhaps the simplest class of bipartite maps are the *plane trees*, the maps of genus 0 with a single face. There are several advantages to working with these: they are easy to draw, their Belyĭ functions are polynomials, which makes computation a little easier, and they all lie on the same Riemann surface. (By contrast, infinitely many complex structures can arise for any given genus $g > 0$; for instance, $\mathcal{Q}_4$ and $\mathcal{Q}'_3$ both have genus 1, but their Riemann surfaces are non-isomorphic, having the form $X = \mathbb{C}/\Lambda$ for non-similar lattices $\Lambda = \mathbb{Z} + \mathbb{Z}\tau$ with

$\tau = i$ and $\exp(2\pi i/3)$ respectively.) Many of the results about Belyĭ functions and plane trees stem from [33]; for a very readable survey, see [34].

A *Shabat polynomial* (or *generalized Chebyshev polynomial*) is a polynomial $P(z) \in \mathbb{C}[z]$ with at most two critical values in $\mathbb{C}$. By replacing $P$ with $aP+b$ for suitable constants $a, b \in \mathbb{C}$ $(a \neq 0)$, one can assume that these critical values are in $\{0,1\}$; the only other critical value in $\Sigma$ is $\infty$ (if $\deg(P) > 1$), so $P\colon \Sigma \to \Sigma$ is then a Belyĭ function. It has a unique pole (at $\infty$), so the bipartite map $\mathcal{B} = P^{-1}(\mathcal{B}_1)$ has a single face; the graph $\mathcal{G} = P^{-1}(I)$ is therefore a tree, and since it is embedded in $\mathbb{C}$, we call $\mathcal{B}$ a plane tree. (More precisely, $\mathcal{B}$ is a *bicoloured plane tree* since its vertices are coloured black or white as they project onto $0$ or $1$.)

Conversely, any plane tree in $\mathbb{C}$ can be regarded as a bipartite map $\mathcal{B}$ on $\Sigma$, so, by choosing a bicolouring of its vertices, we obtain a Belyĭ function $\beta\colon \Sigma \to \Sigma$. Being meromorphic on $\Sigma$, $\beta$ must be a rational function; since $\mathcal{B}$ has a single face, $\beta$ has only one pole, and by using a Möbius function to send this to $\infty$, we can assume that $\beta$ is a polynomial. Since $\beta$ has no finite critical values outside $\{0,1\}$, it is a Shabat polynomial. Thus plane trees and Shabat polynomials are essentially equivalent.

EXAMPLES.

(1) The polynomial $P(z) = z^n$ has $0$ as its only finite critical value, so it is a Belyĭ function $\Sigma \to \Sigma$. The plane tree $P^{-1}(\mathcal{B}_1)$ corresponding to $P$ is the $n$-*star* $\mathcal{S}_n$ shown in Figure 12; it has a black vertex of valency $n$ at $0$, joined by $n$ edges to white vertices of valency $1$ at the $n$-th roots of unity.

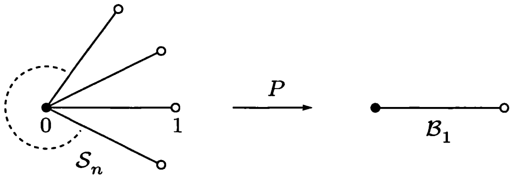

FIGURE 12.

(2) The finite critical points of the polynomial $z^m(1-z)^n$ are at $z=0$ and $z=1$ (if $m,n>1$) and at the point $c = m/(m+n)$ (where the derivative has a simple zero); the critical values are respectively $0$, $0$ and $c^m(1-c)^n = m^m n^n/(m+n)^{m+n}$, so this is a Shabat polynomial, and

$$P_{m,n}(z) = \frac{(m+n)^{m+n}}{m^m n^n} z^m (1-z)^n$$

is a Belyĭ polynomial. Topologically, the corresponding plane tree is made up of an $m$-star centred at $0$ and an $n$-star centred at $1$, with a common white vertex at $c$, as shown in Figure 13.

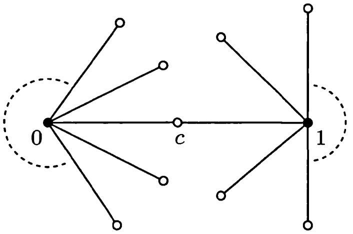

FIGURE 13.

Note that $P_{m,n}$ sends the four points $0$, $1$, $c$ and $\infty$ to the three points $0$, $1$ and $\infty$; we will need this useful property in the next section.

(3) The $n$-th degree Chebyshev polynomial $T_n(z) = \cos\bigl(n\cos^{-1} z\bigr)$ has only $\pm 1$ as its finite critical values, so it is a Shabat polynomial, and the polynomial $P_n = (T_n + 1)/2$ is a Belyĭ function of degree $n$ on $\Sigma$. By considering the graph of the function $P_n$ (restricted to $\mathbb{R}$), one easily sees that the associated plane tree $\mathcal{P}_n = P_n^{-1}(\mathcal{B}_1)$ is an embedding of a path with $n$ edges along the real axis of $\mathbb{C}$; its $n+1$ vertices, alternately coloured white and black, are at $c_r = \cos(r\pi/n)$ for $r = 0, 1, \ldots, n$ (the points where $T_n(z) = \pm 1$). For instance, Figure 14 shows the construction of $\mathcal{P}_5$ (on the $z$-axis) as a 5-sheeted branched covering of $\mathcal{B}_1$ (on the $w$-axis).

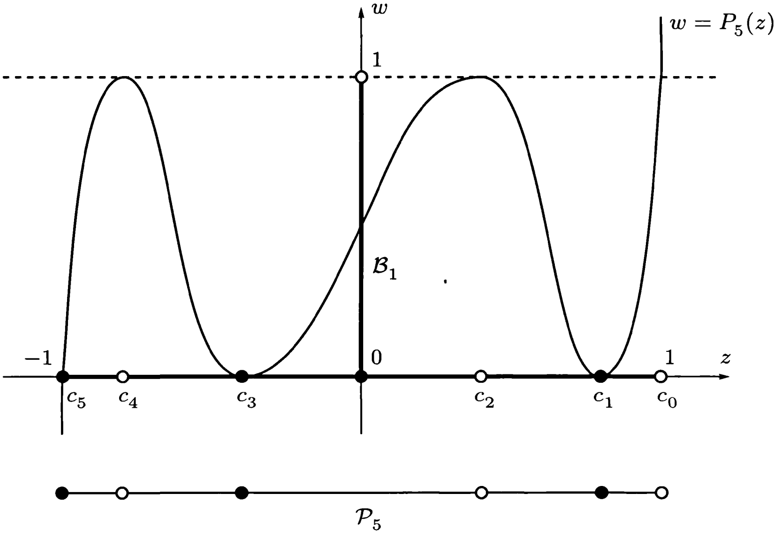

FIGURE 14. $\mathcal{P}_5$.

## 5. Riemann surfaces and algebraic curves

The most familiar examples of compact Riemann surfaces are those obtained from algebraic functions. If $A(x, y) \in \mathbb{C}[x, y]$ (that is, $A(x, y)$ is a polynomial in $x$ and $y$ with complex coefficients), then the equation $A(x, y) = 0$ defines the complex variable $y$ as an $N$-valued function of the complex variable $x$, where $N$ is the degree of $A$ in $y$. One of the basic techniques one learns in complex function theory is the construction of the Riemann surface $X_A$ of this equation by taking $N$ copies of the Riemann sphere $\Sigma$ (one for each branch of the function), cutting them between the branch-points, and then rejoining the sheets across these cuts to show how the function changes from one branch to another by analytic continuation. Let us call a Riemann surface *algebraic* if it is isomorphic to $X_A$ for such a polynomial $A$. The following major result is essentially due to Riemann:

**THEOREM.** *A Riemann surface is compact if and only if it is algebraic.*

This gives rise to a rich correspondence between the theories of compact Riemann surfaces $X$ (which are essentially analytic and topological objects), and of complex algebraic curves (which are geometric and algebraic in nature). For instance, the group of conformal automorphisms of $X$ can be identified with the Galois group of the field of meromorphic functions on $X$. For further details, see [11], [28].

EXAMPLE.

(1) An *elliptic curve* is a compact Riemann surface of genus 1, that is, a complex structure on the torus (see [20], [28], for example). Every such surface $X$ can be represented in Legendre normal form

$$y^2 = x(x-1)(x-\lambda)$$

for some $\lambda \in \mathbb{C} \setminus \{0, 1\}$, and conversely, every such equation defines an elliptic curve, which I will denote by $E_\lambda$. (The value of $\lambda$ is not uniquely determined by $X$: there may be up to six possible values, permuted transitively by the group generated by the transformations $\lambda \mapsto 1 - \lambda$ and $\lambda \mapsto 1/\lambda$.) Incidentally, this allows us to define the $\lambda$-function used in §3: if $\tau \in \mathcal{U}$ and if $\Lambda = \{m + n\tau \mid m, n \in \mathbb{Z}\}$ is the lattice (discrete additive subgroup) generated by 1 and $\tau$ in $\mathbb{C}$, then the torus $X = \mathbb{C}/\Lambda$ is an elliptic curve, and $\lambda(\tau)$ is one of the values of $\lambda$ such that $X \cong E_\lambda$, chosen to vary continuously with respect to $\tau$; for our purposes (though this is not traditional), it is convenient to choose the branch of the function satisfying $\lambda(\tau) \to 0, 1, \infty$ as $\tau \to 0, 1, \infty$ respectively along the hyperbolic geodesics joining these three points.

If $K$ is a subfield of $\mathbb{C}$, then we say that a compact Riemann surface $X$ is *defined over* $K$ if $X \cong X_A$ for some polynomial $A(x, y) \in K[x, y]$. We will be particularly interested in the case where $K$ is the field $\overline{\mathbb{Q}}$ of algebraic numbers. Recall that an element $a \in \mathbb{C}$ is an *algebraic number* if $p(a) = 0$ for some non-zero polynomial $p(t) \in \mathbb{Q}[t]$, or equivalently, if $a$ lies in some finite extension of the rational field $\mathbb{Q}$. The algebraic numbers form a field $\overline{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$; this is a countable, infinite, algebraic extension of $\mathbb{Q}$, equal to the union of all the algebraic number fields (the finite extensions of $\mathbb{Q}$ in $\mathbb{C}$). The following powerful result is due to Belyĭ [2]:

**THEOREM.** *A compact Riemann surface $X$ is defined over $\overline{\mathbb{Q}}$ if and only if there is a Belyĭ function $\beta\colon X \to \Sigma$.*

(Recall that a Belyĭ function is a meromorphic function with no critical values outside $\{0, 1, \infty\}$.) This condition was already known to be sufficient as a direct consequence of Weil's Rigidity Theorem ([41]); Belyĭ's contribution was a simple but ingenious proof of the converse, which I will now outline.

If $X$ is defined over $\overline{\mathbb{Q}}$ by an algebraic equation $A(x, y) = 0$, then the projection $\pi$ onto the $x$-coordinate is a meromorphic function $X \to \Sigma$ with finitely many critical values, all in $\overline{\mathbb{Q}} \cup \{\infty\}$. (If $X$ is an elliptic curve $E_\lambda$ with $\lambda \in \overline{\mathbb{Q}}$, for instance, then there are four critical values $0$, $1$, $\infty$ and $\lambda$, so $\pi$ is "nearly" a Belyĭ function.) The finite critical values of $\pi$ are all roots of some polynomial $p_1(t) \in \mathbb{Q}[t]$, so one could try $p_1 \circ \pi$ as a second approximation to a Belyĭ function on $X$. These critical values are now conveniently sent to $0$, but unfortunately, we may have introduced some new critical values, namely those of $p_1$. However, these are algebraic numbers, so they are all annihilated by a polynomial $p_2(t) \in \mathbb{Q}[t]$; one can show that $\deg(p_2) < \deg(p_1)$, so by iterating this process, one eventually obtains a meromorphic function $f = p_k \circ \ldots \circ p_1 \circ \pi\colon X \to \Sigma$ with a finite number of critical values, all contained in $\mathbb{Q} \cup \{\infty\}$. Any such critical value $c \neq 0, 1, \infty$ can be eliminated by writing $c = m/(m+n)$ and composing $f$ with the function $P_{m,n}$ in §4, which sends $c$ to $1$ and which introduces no new critical values. (One may have to allow $m$ or $n$ to be negative integers, in which case $P_{m,n}$ is a rational function rather than a polynomial, but the principle still applies.) By doing this finitely many times, one can eliminate all unwanted critical values, thus giving a Belyĭ function on $X$.

It is straightforward to reinterpret the existence of a Belyĭ function in terms of uniformisation. First it is useful to introduce the idea of a *hyperbolic triangle group* (see [20; §5.6]). This is a subgroup $\Delta = \Delta(l, m, n)$ of $\mathrm{Aut}(\mathcal{U}) = PSL_2(\mathbb{R})$ generated by rotations through angles $2\pi/l$, $2\pi/m$ and $2\pi/n$ about the vertices of a hyperbolic triangle with internal angles $\pi/l$, $\pi/m$ and $\pi/n$, where $l$, $m$ and $n$ are integers greater than 1. (Such triangles exist in $\mathcal{U}$ if and only if $l^{-1} + m^{-1} + n^{-1} < 1$.) Belyĭ's Theorem can now be restated as follows:

**THEOREM.** *If $X$ is a compact Riemann surface, then the following are equivalent:*

a) *$X$ is defined over $\overline{\mathbb{Q}}$;*
b) *$X \cong \overline{\mathcal{U}}/M$ for some subgroup $M$ of finite index in the modular group $\Gamma$;*
c) *$X \cong \overline{\mathcal{U}}/B$ for some subgroup $B$ of finite index in $\Gamma(2)$;*
d) *$X \cong \mathcal{U}/H$ for some subgroup $H$ of finite index in a hyperbolic triangle group $\Delta$.*

The advantage of using condition (d) is that one can work in $\mathcal{U}$, which is a surface, rather than in $\overline{\mathcal{U}}$, which is not. The disadvantage is that different Riemann surfaces $X$ will, in general, correspond to subgroups of different triangle groups $\Delta$, whereas in (b) and (c) one can work with subgroups of a single group.

From our point of view, the significance of Belyĭ's Theorem is that it shows that the Riemann surfaces defined over $\overline{\mathbb{Q}}$ are precisely those which can be obtained from maps (or bipartite maps) by the methods described earlier: one can regard these combinatorial structures as pictures of Belyĭ pairs. For example, if $X$ is defined over $\overline{\mathbb{Q}}$, then there is a Belyĭ function $\beta\colon X \to \Sigma$, and by using $\beta$ to lift the trivial map or bipartite map from $\Sigma$ to $X$, we obtain a similar structure on $X$. Conversely, if $X$ is obtained in this way, then it is uniformised by a subgroup of finite index in $\Gamma$, and hence, it is defined over $\overline{\mathbb{Q}}$. This means that simple, purely topological objects (such as the childish drawings in Figure 9) can carry a whole wealth of mathematical meaning and structure. For this reason, oriented maps are sometime referred to as *dessins d'enfants.*

EXAMPLES.

(2) We have seen that the bipartite map $\mathcal{B} = \mathcal{K}_{m,n}$ corresponds to a subgroup $B = B_{m,n}$ of index $N = mn$ in $\Gamma(2)$, so this *dessin* can be drawn on the Riemann surface $X = \overline{\mathcal{U}}/B$ as $\beta^{-1}(\mathcal{B}_1)$, where $\beta$ is the projection $\overline{\mathcal{U}}/B \to \overline{\mathcal{U}}/\Gamma(2) = \Sigma$. By Belyĭ's Theorem, this surface $X$ is defined over $\overline{\mathbb{Q}}$. Now in general, it is very difficult to compute the algebraic equation defining the compact Riemann surface uniformised by a given group, but in this particular case, it is quite straightforward. One can verify that $x = \lambda^{1/n}$ and $y = (1-\lambda)^{1/m}$ are single-valued meromorphic functions on $X$. For instance, if we rotate a point $z \in X$ through an angle $2\pi/n$ around a black vertex (of valency $n$), then the point $w = \beta(z) \in \Sigma$ rotates once around the critical value $0$, and hence analytic continuation multiplies $\lambda^{1/n}$ by $\exp(2\pi i/n)$; it follows that a complete rotation of $z$ around the vertex leaves $x$ unchanged. Similar arguments apply to $x$ and $y$ at the other critical points, so they are both single-valued. One can regard the $mn$ sheets of the covering $\beta\colon X \to \Sigma$ as being copies of a fundamental region for $\Gamma(2)$, all carrying the same values of $\lambda$ but carrying the $mn$ distinct branches of the pair $(x, y)$: incomplete rotations around black or white vertices enable one to change the branches of $x$ or $y$ by multiplying them by the appropriate roots of unity. It follows that $x$ and $y$ distinguish points on $X$ in the sense that any two distinct points have neighbourhoods on which $x$ or $y$ (or both) differ. The

polynomial equation of least degree satisfied by $x$ and $y$ is

$$x^n + y^m = 1\,,$$

so $X$ is the Riemann surface of this equation. We call $X$ a *generalized Fermat curve*, since the case $m = n$ gives the Fermat curve $x^n + y^n = 1$ (see §6). The Belyĭ function $\beta$ on $X$ sends each point $(x, y)$ to $\lambda = x^n$. The automorphisms of $\mathcal{K}_{m,n}$ are clearly visible in its equation: this is invariant under simultaneous multiplication of $x$ and $y$ by $n$th and $m$th roots of unity, and these operations form a group $C_n \times C_m$ which is the group of covering transformations of $\beta$ and hence the automorphism group $\mathrm{Aut}_0 \mathcal{B}$ of the bipartite map $\mathcal{B} = \mathcal{K}_{m,n}$.

(3) It is a little more complicated to construct a Belyĭ pair corresponding to the $n$-cube embedding $\mathcal{Q}_n$. Recall that $\mathrm{Aut}\,\mathcal{Q}_n$ contains an elementary abelian normal subgroup $V_0$ of order $2^{n-1}$, which acts regularly on the black vertices and on the white vertices. This is generated by half-turns $h_j$ $(1 \leq j \leq n)$ which induce the transformations $v \mapsto v + e_{j-1} + e_j$ of the vertices $v \in V$ (as usual, we use subscripts $j \in \mathbb{Z}_n$ so that $e_0 = e_n$). The quotient $\tilde{\mathcal{Q}}_n = \mathcal{Q}_n / V_0$ is a bipartite map of genus 0 with one black vertex, one white vertex, $n$ edges and $n$ faces (all 2-gons). One can realise this map on $\Sigma$ as $\tilde{\beta}^{-1}(\mathcal{B}_1)$, where $\tilde{\beta} \colon \Sigma \to \Sigma$ is given by $x \mapsto \big(x/(x-1)\big)^n$, so that the black and white vertices are at $0$ and $\infty$, and the edges are the lines $\arg(x) = 2\pi j/n$ for $j = 1, \ldots, n$ (this parameter $j$ gives the edge-labelling of $\tilde{\mathcal{Q}}_n$ induced from that of $\mathcal{Q}_n$). We now reconstruct $\mathcal{Q}_n$ as a $2^{n-1}$-sheeted regular covering of $\tilde{\mathcal{Q}}_n$, branched at the face-centres. These are the points $a_j = \exp\big((2j-1)\pi i/n\big)$, the $n$th roots of $-1$, so let us consider the Riemann surface $R$ of the algebraic equation

$$y = \pm\sqrt{(x-a_1)(x-a_2)} \pm \sqrt{(x-a_2)(x-a_3)} \pm \cdots \pm \sqrt{(x-a_n)(x-a_1)}\,. \quad (1)$$

The ± signs indicate that one can choose either of the two values for each of the $n$ square roots, so this defines $y$ as a $2^n$-valued function of $x$, branched only at the points $x = a_j$ in $\Sigma$. Analytic continuation around a small loop enclosing $a_j$ has the effect of multiplying the $(j-1)$th and $j$th square roots by $-1$, while leaving the rest unchanged. Since the fundamental group of $\Sigma \setminus \{a_1, \ldots, a_n\}$ is generated by these loops, it follows that analytic continuation in this region allows us to change any even number of the ± signs, but never an odd number. This implies that equation (1) defines two distinct algebraic functions $y$, differing from each other by an odd number of sign-changes, and so $R$ has two connected components, one carrying each function. If $X$ is either of these components, then the projection onto the $x$-coordinate is a $2^{n-1}$-sheeted covering $\pi \colon X \to \Sigma$, branched at the points $a_j$, and $\pi^{-1}(\tilde{\mathcal{Q}}_n)$ is a bipartite map on $X$ with $2^{n-1}$ black and white vertices (covering $0$ and $\infty$), and $n2^{n-1}$ edges, $2^{n-1}$ over each of the $n$ edges of $\tilde{\mathcal{Q}}_n$. Above each point $a_j$ the $2^{n-1}$ sheets of $X$ come together in $2^{n-2}$ pairs, so each 2-gonal face of $\tilde{\mathcal{Q}}_n$ lifts to $2^{n-2}$ quadrilateral faces of $\pi^{-1}(\tilde{\mathcal{Q}}_n)$. It is straightforward to check that the underlying graph of $\pi^{-1}(\tilde{\mathcal{Q}}_n)$ is isomorphic to $Q_n$: for instance, each black vertex $v \in \pi^{-1}(0)$ can be identified with a vector $(v_j) \in V$, where $v_j = 0$ or $1$ as $v$ supports the branch of $\sqrt{(x-a_j)(x-a_{j+1})}$ taking the value $\exp(2\pi ij/n)$ or $-\exp(2\pi ij/n)$ at $x = 0$; if we take $X$ to be the component of $R$ on which $\sum v_j$ is even, then the set of black vertices is identified with $V_0$; by continuing analytically along the edge $\arg(x) = 2\pi j/n$ of $\tilde{\mathcal{Q}}_n$ from $x = 0$ to $x = \infty$, we pass from $v \in V_0$ to a white vertex $w \in \pi^{-1}(\infty)$, which we identify with the vector $v + e_j \in V \setminus V_0$. If we use $\pi$ to lift the edge-labels from $\tilde{\mathcal{Q}}_n$, we obtain the cyclic rotations $(1, 2, \ldots, n)$ and its inverse around the black and white vertices. This shows that $\pi^{-1}(\tilde{\mathcal{Q}}_n) \cong \mathcal{Q}_n$, so we have obtained $\mathcal{Q}_n$ as $\beta^{-1}(\mathcal{B}_1)$, where $\beta\colon X \to \Sigma$ is the Belyĭ function

$$\beta = \tilde{\beta} \circ \pi\colon (x, y) \mapsto \left(\frac{x}{x-1}\right)^n .$$

In theory, one could find the polynomial equation defining $X$ by eliminating all the square roots in (1) and then decomposing the resulting polynomial into two irreducible factors, but I will leave this to the reader's imagination.

At this point, one might object that, whereas Belyĭ's Theorem treats the three critical values $0$, $1$ and $\infty$ with complete symmetry, the maps and bipartite maps I have introduced seem to depend more heavily on $0$ and $1$ than on $\infty$. This is true, and it was done purely for conceptual and diagrammatic simplicity. To avoid the appearance of bias against $\infty$, I will briefly mention that perhaps a more appropriate combinatorial object to associate with a Belyĭ pair is a triangulation (see [21], [33], [43], for example). The *trivial triangulation* $\mathcal{T}_1$ of $\Sigma$ has three vertices (at $0$, $1$ and $\infty$), three edges (the real line-segments joining these vertices) and two faces (the upper and lower half-planes). Given any Belyĭ pair $(X, \beta)$, we obtain a triangulation $\mathcal{T} = \beta^{-1}(\mathcal{T}_1)$ of $X$, together with a 3-colouring of its vertices (as they cover $0$, $1$ or $\infty$). One can also obtain $\mathcal{T}$ as a quotient of the *universal triangulation* $\hat{\mathcal{T}}$ on $\overline{\mathcal{U}}$: this has vertex-set $\mathbb{Q} \cup \{\infty\}$, with $a/b$ joined by a hyperbolic geodesic to $c/d$ if and only if $ad - bc = \pm 1$. If one deletes the vertices of $\mathcal{T}$ labelled $\infty$, together with their incident edges, one obtains the bipartite map $\mathcal{B}$ associated with $(X, \beta)$; conversely, one can reconstruct $\mathcal{T}$ from $\mathcal{B}$ by inserting a new vertex, labelled $\infty$, in each face of $\mathcal{B}$, and joining it by an edge to each vertex incident with that face. Thus bipartite maps and triangulations are essentially equivalent; the former are more economical, while the latter treat the critical values more symmetrically.

EXAMPLE.

(4) If $X$ is the Fermat curve $x^n + y^n = 1$, and $\beta$ is the Belyĭ function $(x, y) \mapsto x^n$ of degree $n^2$, then $\mathcal{T}$ has three sets of $n$ vertices, labelled $0$, $1$ and

$\infty$. Each pair of vertices with different labels are joined by a single edge, so the embedded graph is the complete tripartite graph $K_{n,n,n}$. Since the faces are all triangles, this is a minimum-genus embedding of $K_{n,n,n}$ (see [42]). It is regular as a 3-coloured triangulation, since the map subgroup $B = B_{n,n}$ is normal in $\Gamma(2)$. In fact, since $B$ is the subgroup generated by the $n$th powers and the commutators, it is a characteristic subgroup of $\Gamma(2)$, that is, it is invariant under all automorphisms of $\Gamma(2)$; now $\Gamma(2)$ is a normal subgroup of $\Gamma$, so conjugation by $\Gamma$ induces automorphisms of $\Gamma(2)$ which preserve $B$, and hence, $B$ is a normal subgroup of $\Gamma$. This implies that $\mathcal{T}$ is also regular as an uncoloured triangulation: as such, its automorphism group is an extension of $\Gamma(2)/B \cong C_n \times C_n$ (preserving the vertex-colours) by $\Gamma/\Gamma(2) \cong S_3$ (permuting the colours transitively).

A Belyĭ pair $(X, \beta)$ can equally well be described by the dual of $\mathcal{T}$. This is a *hypermap* $\mathcal{H}$, that is, a trivalent map on $X$ with a 3-colouring of its faces: the faces labelled 0, 1 and $\infty$ are called the hypervertices, hyperedges and hyperfaces of $\mathcal{H}$. The bipartite map $\mathcal{B}$ associated with $(X, \beta)$ then coincides with the Walsh map $W(\mathcal{H})$ of $\mathcal{H}$ introduced by Walsh in [40]: it has black and white vertices corresponding to the hypervertices and hyperedges of $\mathcal{H}$, with edges representing incidences between them, and faces corresponding to the hyperfaces of $\mathcal{H}$. Hypermaps were first studied by Cori [6] in 1975 as a generalization of maps (they embed hypergraphs, rather than graphs); the theory was developed mainly by Cori and Machì, who have given an excellent survey in [7]; for connections between hypermaps and Belyĭ pairs, see [21].

## 6. Fermat's Last Theorem

Having stated Belyĭ's Theorem, I now find it hard to resist making a detour into an area which is not strictly essential for this survey, but which is nevertheless of great topical interest and which exhibits many features in common with the present subject-matter. The impatient reader can skip this section.

A compact Riemann surface (or algebraic curve) $X$ is *modular* if $X \cong \overline{\mathcal{U}}/M$ for some congruence subgroup $M$ of the modular group $\Gamma$. One of the main open problems in the theory of elliptic curves is the following:

**Shimura-Taniyama-Weil Conjecture.** (STW) *If an elliptic curve is defined over $\mathbb{Q}$, then it is modular.*

Notice the similarities and the differences between this and Belyĭ's Theorem: it concerns Riemann surfaces of genus 1, rather than of arbitrary genus; the field of definition is $\mathbb{Q}$, rather than $\overline{\mathbb{Q}}$; it refers to *congruence* subgroups of $\Gamma$, rather than all subgroups of finite index; the implication is in only one direction (the converse is known to be false); finally, it is a conjecture rather than a theorem.

Fermat's Last Theorem (FLT) is the assertion (apparently still unproved) that for integers $p > 2$ there are no positive integer solutions $a$, $b$, $c$ of the equation

$$a^p + b^p = c^p\,,$$

or equivalently, the Fermat curve $x^p + y^p = 1$ has no non-trivial rational points. It is well-known that it is sufficient to prove this when $p$ is prime. It is also well-known (but less easy to prove) that if $a$, $b$, $c$ and $p$ are a counter-example to FLT, then the *Frey curve*

$$y^2 = x(x - a^p)(x + b^p)$$

(which is an elliptic curve defined over $\mathbb{Q}$) cannot be modular. This means that a proof of STW (at least, for the Frey curves) would immediately imply FLT (see [9], [27], [29]). In June 1993, Wiles announced a proof of STW for a wide class of elliptic curves, but at the time of writing it is not clear whether his methods will apply to the Frey curves.

## 7. Galois theory

Returning to our main theme, let us consider the *absolute Galois group* (of the rational numbers); this is the Galois group

$$\boldsymbol{G} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$$

of the field extension $\overline{\mathbb{Q}} \geq \mathbb{Q}$, or simply, the group of field-automorphisms of $\overline{\mathbb{Q}}$. The rest of this section (which is not essential to what follows) is intended to show that it is both important and difficult to understand the structure of $\boldsymbol{G}$. In the next section, I shall show that there is some hope of doing this by considering the action of $\boldsymbol{G}$ on *dessins d'enfants*. For background on Galois theory, see [16].

A *Galois extension* of $\mathbb{Q}$ is a finite normal extension of $\mathbb{Q}$, that is, a finite extension $K \geq \mathbb{Q}$ such that the Galois group $G_K = \mathrm{Gal}(K/\mathbb{Q})$ has fixed-field $\mathbb{Q}$ (in general, it could be a field properly containing $\mathbb{Q}$); for such an extension, the order $|G_K|$ is equal to the degree $|K : \mathbb{Q}|$. Every algebraic number is contained in a Galois extension of $\mathbb{Q}$, namely the splitting-field of its minimal polynomial over $\mathbb{Q}$, so

$$\overline{\mathbb{Q}} = \bigcup_{K \in \mathcal{K}} K\,, \tag{1}$$

where $\mathcal{K}$ is the set of Galois extensions of $\mathbb{Q}$ in $\mathbb{C}$. If $K, L \in \mathcal{K}$ and $L \leq K$, then every element of $G_K$ leaves $L$ invariant, so the restriction mapping gives a group-homomorphism $\rho_{K,L}\colon G_K \to G_L$. (This is, in fact, an epimorphism since every automorphism of $L$ extends to an automorphism of $K$.) These groups $G_K$ $(K \in \mathcal{K})$ and homomorphisms $\rho_{K,L}$ form a projective (or inverse) system, meaning that each pair $L_1, L_2 \in \mathcal{K}$ are contained in some $K \in \mathcal{K}$, and $\rho_{K,L} \circ \rho_{L,M} = \rho_{K,M}$ whenever $K \geq L \geq M$ in $\mathcal{K}$.

EXAMPLE. The *cyclotomic field* of the $n$th roots of unity is the field $K = \mathbb{Q}(\varepsilon_n)$, where $\varepsilon_n = \exp(2\pi i/n)$. This is a normal extension of $\mathbb{Q}$ of degree $\phi(n)$, where $\phi$ is Euler's function. Its Galois group $G_K = G_n$ consists of the $\phi(n)$ automorphisms $\varepsilon_n \mapsto \varepsilon_n^u$, where $(u,n) = 1$, so $G_K$ is isomorphic to the group $U_n$ of units in $\mathbb{Z}_n$. We have $\mathbb{Q}(\varepsilon_m) \leq \mathbb{Q}(\varepsilon_n)$ if and only if $m$ divides $n$, in which case the restriction mapping corresponds to the natural epimorphism $U_n \to U_m$.

In view of (1), an element $g \in \boldsymbol{G}$ is simply a choice of automorphisms $g_K \in G_K$ of $K$, one for each $K \in \mathcal{K}$, which is compatible with the restriction mappings, that is, such that $\rho_{K,L}$ sends $g_K$ to $g_L$ whenever $K \geq L$. Thus $\boldsymbol{G}$ is the subgroup of the cartesian product $\prod_{K \in \mathcal{K}} G_K$ defined by these compatibility restrictions, so $\boldsymbol{G}$ is the projective (or inverse) limit

$$\boldsymbol{G} = \varprojlim G_K$$

of the groups $G_K$. (This shows that $\boldsymbol{G}$ is a *profinite* group, meaning a *pro*jective limit of *finite* groups.) It follows that $\boldsymbol{G}$ is uncountable, so one cannot expect $\boldsymbol{G}$ to have a particularly simple structure. If we impose the discrete topology on each $G_K$, then $\boldsymbol{G}$ becomes a topological group, which is compact by Tychonoff's Theorem. This topology on $\boldsymbol{G}$ (called the *Krull topology*) is important, since under the Galois connection between subfields and subgroups, the subfields of $\overline{\mathbb{Q}}$ correspond to the closed subgroups of $\boldsymbol{G}$. In particular, the algebraic number fields correspond to the subgroups of finite index in $\boldsymbol{G}$ (which are both open and closed), and the Galois extensions $K \in \mathcal{K}$ correspond to the normal subgroups of finite index (with quotient-group isomorphic to $G_K$).

In a sense, the whole of classical Galois theory is contained in this single group $\boldsymbol{G}$. For instance, one of the most difficult open problems in Galois theory is the Inverse Galois Problem, which is, in its simplest form, Hilbert's conjecture that every finite group $F$ is isomorphic to $G_K$ for some Galois extension $K$ of $\mathbb{Q}$. By the preceding remarks, this is equivalent to the conjecture that every such $F$ is an epimorphic image of $\boldsymbol{G}$. This is easy to prove for the cyclic, dihedral, symmetric and alternating groups, and, with considerably more difficulty, Shafarevich has proved it for finite solvable groups (the error in [35] concerning the prime 2 is corrected in his collected papers). It has been verified for a few classes of finite simple groups, but progress is very slow, and a complete solution seems to be a long way off. See [26], [32] for detailed treatments of this subject.

## 8. Actions of $G$

In his *Esquisse d'un Programme* [12], Grothendieck proposed that one should study this highly algebraic object $G$ through its actions on topological, geometric, and even combinatorial structures. There are many levels at which this can be done (the whole edifice is known as the *Teichmüller tower*), but at the simplest and most explicit level, it is maps (or *dessins d'enfants*) which provide the context for representing $G$.

By definition, $G$ acts on $\overline{\mathbb{Q}}$, so it has natural induced actions on polynomials and rational functions over $\overline{\mathbb{Q}}$, and hence on Belyĭ pairs: one simply applies $G$ to the coefficients of these functions and notes that since the critical values $0, 1$ and $\infty$ are fixed by $G$, the set of Belyĭ functions is preserved. As we have seen, there are correspondences between Belyĭ pairs and various combinatorial structures such as maps, bipartite maps, triangulations and hypermaps on Riemann surfaces, so we obtain further induced actions of $G$ on all of these combinatorial categories. The remarkable fact, pointed out by Grothendieck, is that these actions are all faithful, that is, each non-identity element of $G$ sends some object to a non-isomorphic object in the same category, so one loses none of the structure of $G$ by representing it in these ways.

For a very simple example of this action of $G$, due to Malle, consider the third map $\mathcal{M}$ in Figure 9, with the Mathieu group $M_{12}$ as its cartographic group. The associated Belyĭ function is a rational function $\beta\colon X = \Sigma \to \Sigma$ defined over the algebraic number field $K = \mathbb{Q}(\sqrt{-11}) < \overline{\mathbb{Q}}$. Complex conjugation is an element $\sigma \in G$ which induces the unique non-trivial automorphism of $K$, and by replacing $\beta$ with the conjugate Belyĭ function $\beta^{\sigma}$ on $\Sigma$, we obtain the mirror-image $\mathcal{M}^{\sigma}$ of $\mathcal{M}$. (This is equivalent to applying the outer automorphism of order 2 of $M_{12}$, which transposes its two transitive representations of degree 12.)

It is pointless looking at our examples based on the Fermat curves or the Chebyshev polynomials for non-trivial actions of $G$: in all these cases, the Belyĭ pairs are defined over $\mathbb{Q}$, so they and their associated *dessins* are fixed by $G$. In the case of the $n$-cube embedding $\mathcal{Q}_n$, we see from Example 3 of §5 that the corresponding Belyĭ pair $(X, \beta)$ is defined over the cyclotomic field $\mathbb{Q}(\varepsilon_{2n})$ of the $2n$th roots of unity; it follows that $G$ acts as the Galois group $G_{2n} \cong U_{2n}$ of this field, permuting the branch-points $a_j$ in equation (1) and hence redefining the cyclic order of edge-labels around the vertices. However, this is not a very good example since the resulting maps are all isomorphic to $\mathcal{Q}_n$ under the obvious isomorphisms. Instead, one can find non-trivial actions of $G$, again based on cyclotomic fields, by taking double coverings of some of our earlier examples.

EXAMPLES.

(1) The plane tree $\mathcal{P}_n$, obtained in §4 from the Belyĭ polynomial $P_n(z) = (T_n(z) + 1)/2$, is a path of $n$ edges with vertices at $c_r = \cos(r\pi/n)$ for

$r = 0, 1, \ldots, n$ coloured alternately white and black. If $k = 0, 1, \ldots, n$, then $P_n(z + c_k)$ is also a Belyĭ polynomial, and its plane tree is obtained by translating $\mathcal{P}_n$ along the real axis so that it has $k$ positive vertices and edges, $n - k$ negative vertices and edges, and one vertex at the origin. We now take a double covering of this plane tree, branched at 0 and $\infty$, by forming the Belyĭ polynomial $P_n(z^2 + c_k)$. The associated plane tree $\mathcal{Q}_{n,k}$ has a vertex at the origin, coloured black or white as $k$ is odd or even; there are two paths of $k$ edges emanating from 0 along the positive and negative parts of the real axis, both ending in white vertices, and two paths of $n - k$ edges along the imaginary axis, ending in black or white vertices as $n$ is odd or even. Figure 15 illustrates $\mathcal{Q}_{n,k}$ for odd $n$, with $k$ odd or even, as a covering of $\mathcal{P}_n$. Note that if $n$ is odd, then $\mathcal{Q}_{n,0}, \ldots, \mathcal{Q}_{n,n}$ are mutually non-isomorphic (as bicoloured plane trees), whereas $\mathcal{Q}_{n,k} \cong \mathcal{Q}_{n,n-k}$ when $n$ is even.

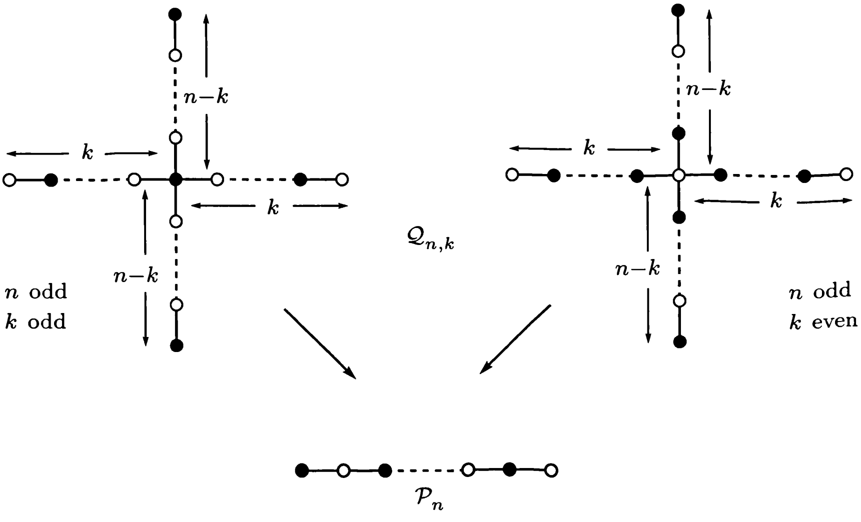

FIGURE 15. $\mathcal{Q}_{n,k}$.

Now $c_k = (\varepsilon_{2n}^k + \varepsilon_{2n}^{-k})/2$, so $\mathcal{Q}_{n,k}$ is defined over the cyclotomic field $\mathbb{Q}(\varepsilon_{2n})$. (Of course, it is actually defined over the proper subfield $\mathbb{Q}(c_k) \leq \mathbb{Q}(\varepsilon_{2n}) \cap \mathbb{R}$, but it is more convenient to work with $\mathbb{Q}(\varepsilon_{2n})$.) The group $G$ acts on $\mathbb{Q}(\varepsilon_{2n})$ as the Galois group $G_{2n} \cong U_{2n}$ of this field, so it permutes the vertices $c_k$ of $\mathcal{P}_n$ in the same way as it permutes the mutually inverse pairs $\{\varepsilon_{2n}^k, \varepsilon_{2n}^{-k}\}$ of $2n$th roots of unity: two vertices $c_k$ and $c_l$ are in the same orbit of $G_{2n}$ if and only if $\varepsilon_{2n}^k$ and $\varepsilon_{2n}^l$ have the same multiplicative order, that is, $(k, 2n) = (l, 2n)$, and so this is the condition for the plane trees $\mathcal{Q}_{n,k}$ and $\mathcal{Q}_{n,l}$ to be conjugate under $\boldsymbol{G}$. Thus the number of orbits of $\boldsymbol{G}$ on these trees is $\tau(2n)$, the number of positive divisors of $2n$. If $n$ is odd, then since $\tau$ is multiplicative, we have $\tau(2n) = 2\tau(n)$; both $\mathcal{Q}_{n,0}$ and $\mathcal{Q}_{n,n}$ (which are embeddings of paths) form orbits of length 1, while the other trees $\mathcal{Q}_{n,k}$ lie in orbits of length

$$\frac{1}{2}\phi\left(\frac{2n}{(k,2n)}\right) = \frac{1}{2}\phi\left(\frac{n}{(k,n)}\right).$$

When $n$ is an odd prime, for instance, there are four orbits: apart from the two orbits of length 1, there are also two orbits of length $(n-1)/2$ consisting of the remaining trees $\mathcal{Q}_{n,k}$ for $k$ odd and $k$ even. Thus the orbits of $\boldsymbol{G}$ can be arbitrarily large. The situation is similar when $n$ is even, though in this case, the extra isomorphisms make the counting slightly more complicated.

(2) For a similar class of examples, but now of genus 1, let $X$ be the elliptic curve $E_\lambda$, given by $y^2 = x(x-1)(x-\lambda)$, where $\lambda^n = 1 \neq \lambda$. As we saw in §5, the projection $\pi\colon X \to \Sigma$, $(x,y) \mapsto x$, is a 2-sheeted covering, branched over $x = 0, 1, \infty$ and $\lambda$. The polynomial $P(x) = x^n$ sends these critical values of $\pi$ to $0, 1, \infty$ and 1, and has critical values 0 and $\infty$, so the composition $\beta = P \circ \pi\colon (x,y) \mapsto x^n$ is a Bely̆ı function $X \to \Sigma$, of degree $N = 2n$. Now $P^{-1}(\mathcal{B}_1)$ is the $n$-star $\mathcal{S}_n$ constructed in Example 1 of §4, so the bipartite map $\mathcal{B} = \beta^{-1}(\mathcal{B}_1)$ associated with $(X, \beta)$ is a double covering $\pi^{-1}(\mathcal{S}_n)$ of $\mathcal{S}_n$, branched over its black vertex 0, its white vertices 1 and $\lambda$, and also over its face-centre $\infty$. Let us denote this torus map by $\mathcal{T}_{n,k}$, where $\lambda = \varepsilon_n^k$; it is shown in Figure 16, with opposite sides identified.

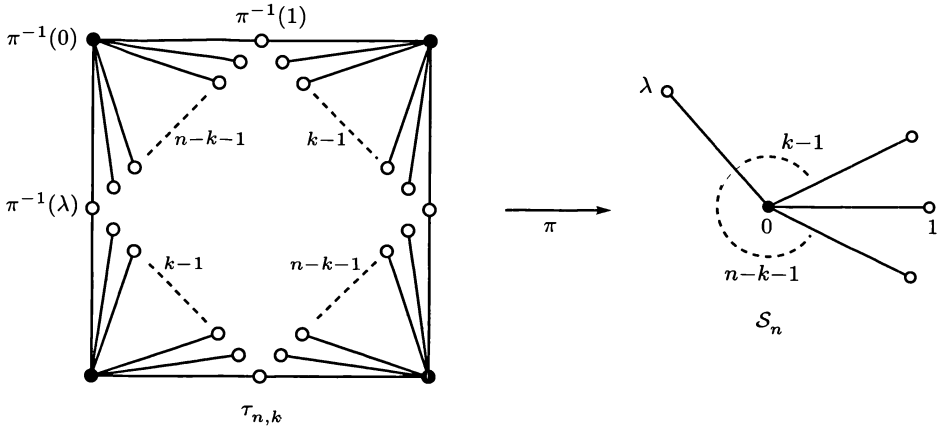

FIGURE 16.

There is one black vertex, of valency $2n$, there are $2n-2$ white vertices, two of valency 2 and the rest of valency 1, and there is a single $4n$-gonal face. In the cyclic rotation around the black vertex, the 2-valent white vertices are separated by blocks of $k-1$ and $n-k-1$ vertices of valency 1, so $\mathcal{T}_{n,k} \cong \mathcal{T}_{n,l}$ if and only if $k \equiv \pm l \mod (n)$. As in Example 1, these bipartite maps are conjugate under $\boldsymbol{G}$ if and only if their corresponding values of $\lambda$ are conjugate, that is, $(k,n) = (l,n)$. (Notice that in this case, $\boldsymbol{G}$ preserves the formula for $\beta$, and acts non-trivially on the underlying algebraic curves $X$, whereas in Example 1, it fixes $X$ and acts non-trivially on the Belyĭ functions $\beta$.) Wolfart gives several similar examples of conjugate *dessins* on the torus in [43].

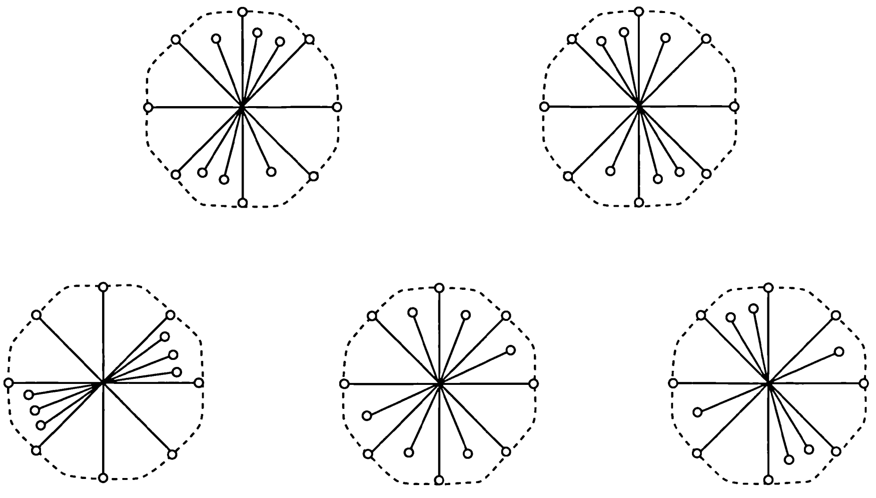

FIGURE 17.

(3) One can find similar actions of $\boldsymbol{G}$ on *dessins* of higher genus. For instance, a *hyperelliptic surface* $X$ is a 2-sheeted covering of $\Sigma$; it has equation

$$y^2 = (x-a_1)\dots(x-a_m),$$

where $a_1, \dots, a_m$ are distinct complex numbers, and conversely, every such equation defines a hyperelliptic surface. The double covering is given by the projection $\pi\colon (x,y) \mapsto x$, which has critical values $a_1, \dots, a_m$ (and $\infty$ if $m$ is odd), and the genus is $g = \lfloor (m-1)/2 \rfloor$. If we choose each $a_j$ to be 0 or an $n$th root of unity for some fixed $n$, then $\beta\colon (x,y) \mapsto x^n$ is a Belyĭ function of degree $N = 2n$ on $X$; the corresponding bipartite map $\mathcal{B}$ is a branched double covering $\pi^{-1}(\mathcal{S}_n)$ of $\mathcal{S}_n$, with one or two faces as $m$ is odd or even. For instance, if $a_1 = 0$ and if $a_j^n = 1$ for $j = 2, \dots, m$, then $\mathcal{B}$ has a single black vertex of valency $2n$, together with $m-1$ white vertices of valency 2, and $2(n-m+1)$ white vertices of valency 1. The action of $\boldsymbol{G}$ on these *dessins* corresponds to that of the cyclotomic Galois group $G_n \cong U_n$ on the sets $A = \{a_2, \dots, a_m\}$ of $m-1$ distinct

$n$th roots of unity. Two such *dessins* are isomorphic if and only if their corresponding sets $A$ are equivalent under a rotation around $0$. In Figure 17, the two rows show the two orbits of $\boldsymbol{G}$ on the isomorphism classes of these *dessins* in the case $m = 5$ and $n = 7$: opposite sides of each octagon are identified to produce a surface of genus $g = 2$. Alternatively, if $a_j^n = 1$ for each $j = 1, \ldots, m$, then $\mathcal{B}$ has two black vertices of valency $n$, together with $m$ white vertices of valency $2$, and $2(n - m)$ of valency $1$. The action of $\boldsymbol{G}$ now corresponds to that of $G_n$ on the sets of $m$ distinct $n$th roots of unity. In either case, by choosing $m$ first, one can specify the genus, and then by choosing $n$ appropriately, one can make the number of $\boldsymbol{G}$-orbits and their lengths arbitrarily large.

These examples illustrate an important general principle that in its action on *dessins*, $\boldsymbol{G}$ preserves most of their obvious numerical, topological and algebraic features; these include the genus, the numbers of vertices, edges and faces, their colours, valencies and face-sizes, the automorphism group and the monodromy group. This is because these can all be defined in terms of field-theoretic properties which are invariant under the Galois group $\boldsymbol{G}$, as shown in [23] for example. (However, orientation-*reversing* automorphisms are not generally preserved by $\boldsymbol{G}$.) Since the numbers of vertices, edges and faces are invariant, an orbit of $\boldsymbol{G}$ can contain only a finite number of *dessins* (up to isomorphism). Indeed, the number of conjugates of a *dessin* is bounded above by the degree over $\mathbb{Q}$ of its field $K$ of definition, since this is the number of embeddings of $K$ in $\mathbb{C}$; if we take $n$ odd and $(n, k) = 1$ in Example 1 so that $K = \mathbb{Q}(\varepsilon_{2n}) \cap \mathbb{R}$ of degree $\phi(2n)/2 = \phi(n)/2$, we see that this upper bound is attained.

Because the genus and the number of faces are preserved, $\boldsymbol{G}$ leaves invariant the class of plane trees. What is surprising is that, even on such apparently simple objects, the action of $\boldsymbol{G}$ is faithful as was recently proved by Schneps [30], using an argument of Lenstra. This result, which was the justification for concentrating on plane trees in §4, has motivated a systematic study of the relationships between plane trees, polynomials and Galois groups, often using computer algebra systems for the calculations [3], [8], [23], [30], [34]. Unfortunately, the various proofs of the faithful action of $\boldsymbol{G}$ do not lead to particularly simple explicit classes of *dessins*, such as those considered above. Example 1 demonstrates that the kernel of the action on plane trees has infinite index in $\boldsymbol{G}$, and Examples 2 and 3 do the same for *dessins* of higher genus, but they do not show that the kernel is trivial: since these *dessins* are all defined over cyclotomic fields, which have abelian Galois groups, they are fixed by the commutator subgroup $\boldsymbol{G}'$ of $\boldsymbol{G}$, which contains much of its most interesting structure. By contrast, the following example illustrates a non-abelian action of $\boldsymbol{G}$.

(4) Let $X$ be the elliptic curve

$$y^2 = x(x-1)(x-a)(x-b),$$

where $a$ and $b$ are distinct roots of the polynomial $f(t) = 2t^p - 1$ for some odd prime $p$. As in Example 3, the projection $\pi\colon (x, y) \mapsto x$ is a 2-sheeted covering $X \to \Sigma$ with critical values $0$, $1$, $a$ and $b$. The polynomial $x \mapsto x^p$ (which has critical values 0 and $\infty$) sends these to 0, 1 and 1/2, and these three points are then sent to 0 and 1 by the polynomial $w \mapsto 4w(1 - w)$ (this is the polynomial $P_{1,1}$ of §4, Example 2).

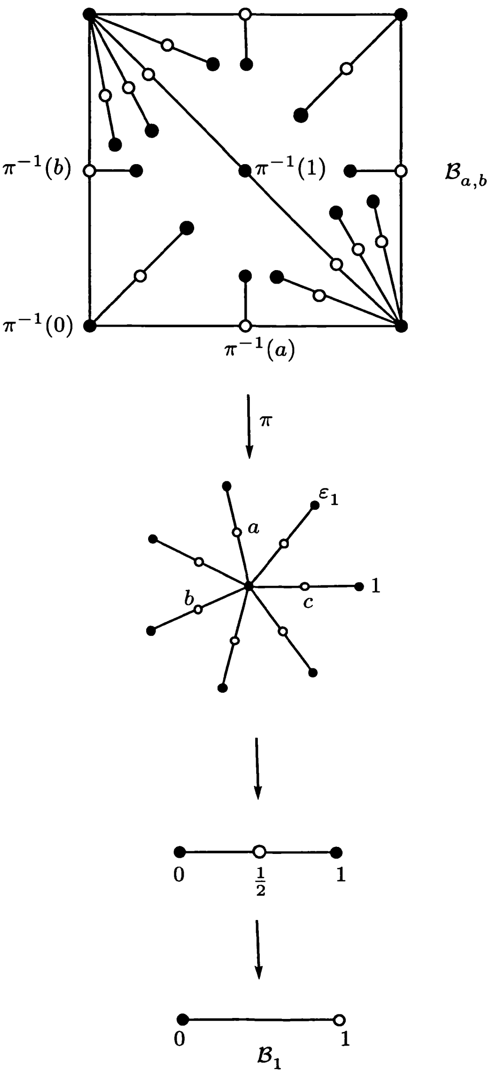

FIGURE 18.

It follows that the composition $\beta\colon (x,y) \mapsto 4x^p(1-x^p)$ is a Belyĭ function of degree $N = 4p$ on $X$. The construction of the corresponding bipartite map $\mathcal{B}_{a,b} = \beta^{-1}(\mathcal{B}_1)$ is illustrated in Figure 18 in the case $p = 7$ (in general, one has to distinguish between the cases where one or none of $a$ and $b$ coincides with the unique real root $c$ of $f$). It is easily seen that $\mathcal{B}_{a,b} \cong \mathcal{B}_{a',b'}$ if and only if $\{a,b\} = \{a',b'\}$. These maps $\mathcal{B}_{a,b}$ are all defined over the splitting field $K = \mathbb{Q}(c, \varepsilon_p)$ of $f$. The Galois group of $K$ permutes the roots $a = c\varepsilon_p^j$ $(j \in \mathbb{Z}_p)$ of $f$ by inducing the transformations $j \mapsto uj + t$ of the exponents $j$, where $t, u \in \mathbb{Z}_p$ and $u \neq 0$. These transformations form the 1-dimensional affine group $AGL_1(p)$ over $\mathbb{Z}_p$, a split extension of a cyclic normal subgroup of order $p$ (the transformations $j \mapsto j+t$) by a cyclic group of order $p-1$ (the transformations $j \mapsto uj$). Its action on the roots of $f$ is doubly transitive (transitive on distinct ordered pairs), so for a given prime $p$ the $p(p-1)/2$ bipartite maps $\mathcal{B}_{a,b}$ are all conjugate under $\boldsymbol{G}$. For each $p$ the group $AGL_1(p)$ is metabelian (an extension of one abelian group by another), so the second commutator subgroup $\boldsymbol{G}''$ of $\boldsymbol{G}$ is in the kernel of this action.

It would be interesting to produce further classes of *dessins* on which $\boldsymbol{G}$ induces more complicated groups, such as non-solvable groups or solvable groups of unbounded derived length.

## REFERENCES

[1] AHLFORS, L. V.: *Complex Analysis*, McGraw-Hill, Tokyo, 1966.

[2] BELYĬ, G. V.: *On Galois extensions of a maximal cyclotomic field*, Izv. Akad. Nauk SSSR **43** (1979), 269–276 (Russian) [Math. USSR Izvestiya **14** (1980), 247–256 (English translation)].

[3] BÉTRÉMA, J.—PÉRÉ, D.—ZVONKIN, A.: *Plane trees and their Shabat polynomials.* Preprint, Bordeaux, 1992.

[4] BRYANT, R. P.—SINGERMAN, D.: *Foundations of the theory of maps on surfaces with boundary*, Quart. J. Math. Oxford Ser. (2) **36** (1985), 17–41.

[5] BIGGS, N. L.—WHITE, A. T.: *Permutation Groups and Combinatorial Structures.* London Math. Soc. Lecture Note Ser. 33, Cambridge Univ. Press, Cambridge, 1979.

[6] CORI, R.: *Un code pour les graphes planaires et ses applications*, Astérisque **27** (1975).

[7] CORI, R.—MACHÌ, A.: *Maps, hypermaps and their automorphisms: a survey I, II, III*, Exposition. Math. **10** (1992), 403–427, 429–447, 449–467.

[8] COUVEIGNES, J.-M.: *Calcul et rationalité de fonctions de Belyi en genre* 0, Ann. Inst. Fourier (Grenoble) **44** (1994).

[9] COX, D.: *Introduction to Fermat's Last Theorem*, Amer. Math. Monthly **101** (1994), 3–14.

[10] COXETER, H. S. M.—MOSER, W. O. J.: *Generators and Relations for Discrete Groups*, Springer-Verlag, Berlin-Göttingen-Heidelberg-New York, 1965.

[11] DOUADY, R. and A.: *Algebre et Théories Galoisiennes II*, CEDIC/Fernand Nathan, Paris, 1979.

[12] GROTHENDIECK, A.: *Esquisse d'un programme.* Preprint, Montpellier, 1984.

[13] HAMILTON, W. R.: *Letter to John T. Graves 'On the Icosian' (17th October 1856).* In: Mathematical Papers, Vol. III, Algebra (H. Halberstam, R. E. Ingram, eds.), Cambridge University Press, Cambridge, 1967, pp. 612–625.

[14] HEFFTER, L.: *Über das Problem der Nachbargebiete*, Math. Ann. **38** (1891), 477–508.

[15] HEFFTER, L.: *Über metacyklische Gruppen und Nachbarconfigurationen*, Math. Ann. **50** (1898), 261–268.

[16] JACOBSON, N.: *Basic Algebra I, II*, Freeman, San Fransisco, 1974, 1980.

[17] JONES, G. A.: *Graph imbeddings, groups, and Riemann surfaces.* In: Algebraic Methods in Graph Theory, Szeged 1978 (L. Lovász, V. T. Sós, eds.), North-Holland, Amsterdam, 1981, pp. 397–311.

[18] JONES, G. A.: *Congruence and non-congruence subgroups of the modular group.* In: Groups, St. Andrews 1985 (E. F. Robertson, C. M. Campbell, eds.), London Math. Soc. Lecture Note Ser. 121, Cambridge Univ. Press, Cambridge, 1986, pp. 223–234.

[19] JONES, G. A.—SINGERMAN, D.: *Theory of maps on orientable surfaces*, Proc. London Math. Soc. **37** (1978), 273–307.

[20] JONES, G. A.—SINGERMAN, D.: *Complex Functions, an Algebraic and Geometric Viewpoint*, Cambridge University Press, Cambridge, 1987.

[21] JONES, G. A.—SINGERMAN, D.: *Maps, hypermaps and triangle groups.* In: The Grothendieck Theory of Dessins d'Enfants (L. Schneps, ed.), London Math. Soc. Lecture Note Ser. 200, Cambridge Univ. Press, Cambridge, 1994, pp. 115–145.

[22] JONES, G. A.—SINGERMAN, D.: *Belyĭ functions, hypermaps and Galois groups*, Bull. London Math. Soc. (To appear).

[23] JONES, G. A.—STREIT, M.—ZVONKIN, A.: *Plane trees and Galois groups.* Preprint, Southampton, 1994.

[24] MALGOIRE, J.—VOISIN, C.: *Cartes cellulaires*, Cahiers Math. Montpellier **12** (1977).

[25] MALLE, G.—SAXL, J.—WEIGEL, T.: *Generators of classical groups*, Geom. Dedicata **49** (1994), 85–116.

[26] MATZAT, B. H.: *Konstruktive Galoistheorie.* Lecture Notes in Math. 1284, Springer-Verlag, Berlin, 1987.

[27] MAZUR, B.: *Number theory as gadfly*, Amer. Math. Monthly **98** (1991), 593–610.

[28] REYSSAT, E.: *Quelques Aspects des Surfaces de Riemann*, Birkhäuser, Boston-Basel-Berlin, 1989.

[29] RIBET, K. A.: *Wiles proves Taniyama's conjecture; Fermat's Last Theorem follows*, Notices Amer. Math. Soc. **40** (1993), 575–576.

[30] SCHNEPS, L.: *Dessins d'enfants on the Riemann sphere.* In: The Grothendieck Theory of Dessins d'Enfants (L. Schneps, ed.), London Math. Soc. Lecture Note Ser. 200, Cambridge Univ. Press, Cambridge, 1994, pp. 47–77.

[31] *The Grothendieck Theory of Dessins d'Enfants* (L. Schneps, ed.), London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge, 1994.

[32] SERRE, J.-P.: *Topics in Galois Theory*, Jones and Bartlett, Boston-London, 1992.

[33] SHABAT, G. B.—VOEVODSKY, V. A.: *Drawing curves over number fields.* In: Grothendieck Festschrift III (P. Cartier et al., eds.), Progr. Math. 88, Birkhäuser, Basel, 1990, pp. 199–227.

[34] SHABAT, G.—ZVONKIN, A.: *Plane trees and algebraic numbers.* Preprint, Moscow and Bordeaux, 1993.

[35] SHAFAREVICH, I. R.: *Construction of fields of algebraic numbers with given solvable Galois group*, Izv. Akad. Nauk. SSSR **18** (1954), 525–578 (Russian) [Amer. Math. Transl. **4** (1956), 185–237 (English translation)].

[36] SINGERMAN, D.: *Automorphisms of maps, permutation groups and Riemann surfaces*, Bull. London Math. Soc. **8** (1976), 65–68.

[37] SINGERMAN, D.: *Universal tessellations*, Revista Mat. Univ. Complut. Madrid **1** (1988), 111–123.

[38] TUTTE, W. T.: *What is a map?* In: New Directions in Graph Theory (F. Harary, ed.), Academic Press, New York, 1973.

[39] VINCE, A.: *Combinatorial maps*, J. Combin. Theory Ser. B **34** (1983), 1–21.

[40] WALSH, T. R. S.: *Hypermaps versus bipartite maps*, J. Combin. Theory Ser. B **18** (1975), 155–163.

[41] WEIL, A.: *The field of definition of a variety*, Amer. J. Math. **78** (1956), 509–524.

[42] WHITE, A. T.: *Graphs, Groups and Surfaces*, North-Holland, Amsterdam, 1973.

[43] WOLFART, J.: *Mirror-invariant triangulations of Riemann surfaces, triangle groups and Grothendieck dessins: variations on a theme of Belyi.* Preprint, Frankfurt, 1992.

Received March 1, 1995
Revised August 11, 1995

*Department of Mathematics*
*University of Southampton*
*Southampton SO17 1BJ*
*UNITED KINGDOM*

*E-mail*: gaj@maths.soton.ac.uk