

# Belyi Functions: Examples, Properties, and Applications

Alexander K. Zvonkin\*

## Abstract

Let  $X$  be a Riemann surface, and  $f : X \rightarrow \overline{\mathbb{C}}$  a non-constant meromorphic function on  $X$  (here  $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$  is the complex Riemann sphere). The function  $f$  is called a *Belyi function*, and the pair  $(X, f)$ , a *Belyi pair*, if  $f$  is unramified outside  $\{0, 1, \infty\}$ . The study of Belyi functions, otherwise called the theory of *dessins d'enfants*, provides a link between many important theories. First of all, it is related to Riemann surfaces, as follows from the definition. Then, to Galois theory since, according to the Belyi theorem, a Belyi function on  $X$  exists if and only if  $X$  is defined over the field  $\overline{\mathbb{Q}}$  of algebraic numbers. It is also related to combinatorics of maps, otherwise called embedded graphs, since  $f^{-1}([0, 1])$  is a graph drawn on the two-dimensional manifold underlying  $X$ . Therefore, certain Galois invariants can be expressed in purely combinatorial terms. More generally, many properties of functions, surfaces, fields, and groups in question may be “read from” the corresponding pictures, or sometimes constructed in a “picture form”. Group theory is related to all the above subjects and therefore plays a central role in this theory.

The theory of Belyi functions is very rich with beautiful examples. It is difficult to give a representative sample of them, but we will try to do it at least to some extent.

## 1 Maps and hypermaps

**Convention 1.1** We consider only connected graphs; loops and multiple edges are allowed. All maps in this paper will be oriented.

When we draw a graph on a piece of paper, we add to it an additional structure. Namely, we choose a particular cyclic order of the edges in the

---

\*LaBRI, Université Bordeaux I, 351 Cours de la Libération, F-33405 Talence Cedex, France; e-mail: [zvonkin@labri.fr](mailto:zvonkin@labri.fr)

vicinity of each vertex. Also, the drawing itself subdivides the plane into a number of regions. The resulting object is a map; however, the theory of maps is more beautiful when we draw graphs not on the plane (“piece of paper”) but on the sphere, the torus, or another compact two-manifold.

**Definition 1.2 (Map)** A *map* is a graph embedded in a compact oriented two-dimensional manifold in such a way that

- the edges do not intersect;
- the complement of the graph in the surface is a disjoint union of regions homeomorphic to open disks.

These regions are called *faces* of the map. The *genus* of the map is, by definition, the genus of the underlying surface.

We define the *degree of a vertex* as the number of incident edges (loops are counted twice), and the *degree of a face*, as the number of edges surrounding it (“inner edges” are counted twice).

Figures 1, 3, 2 will help to understand the above definitions.

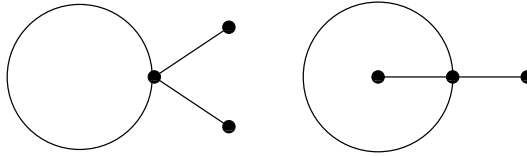


Figure 1: One graph but two different maps. The face degrees of the map on the left are 1 and 5, and for that on the right, 3 and 3. Note that a graph does not have faces at all: it has only vertices and edges.

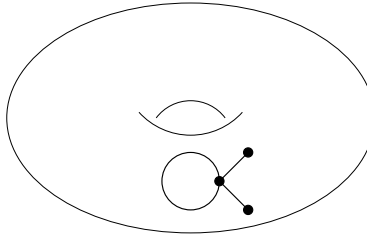


Figure 2: This is not a map: one of the “regions” obtained after cutting the surface along the edges of the graph is not homeomorphic to an open disk.

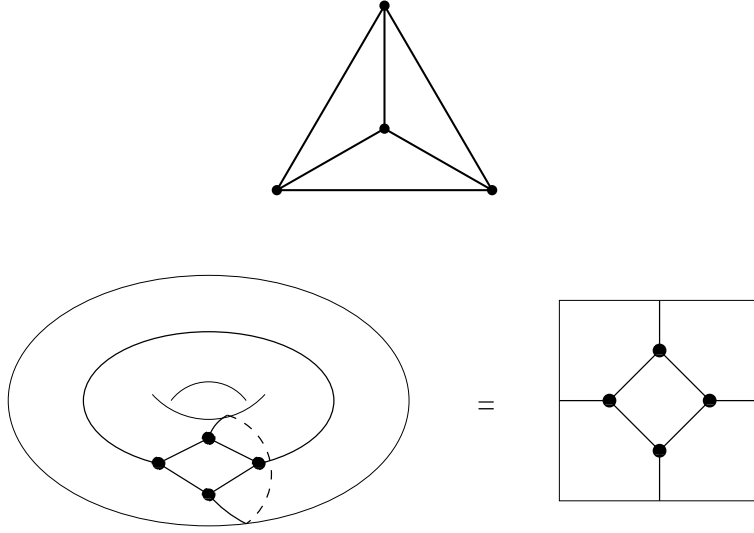


Figure 3: The graph  $K_4$  gives rise to a map of genus 0 (tetrahedron) having 4 faces of degree 3, and to a map of genus 1 having 2 faces, of degree 8 and 4 respectively.

**Definition 1.3 (Hypermap)** A *hypermap* is a bicolored map: its vertices are colored in black and white in such a way that the adjacent vertices have opposite colors.

An example of a hypermap is given in Figure 4.

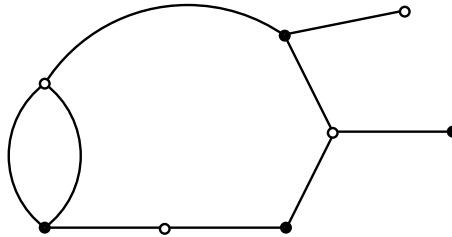


Figure 4: A hypermap.

The above definition suggests that a hypermap is a particular case of a map. However, hypermaps were first invented as generalizations of maps [5]. Indeed, taking a map, we may insert a white vertex in the middle of every edge (see Figure 5), thus obtaining a particular case of a hypermap—such that all its white vertices are of degree 2.

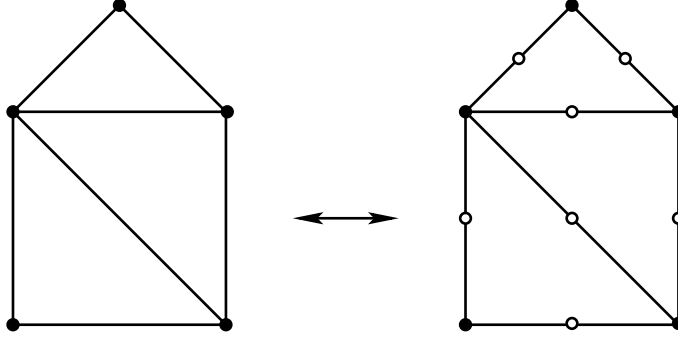


Figure 5: Maps may be considered as hypermaps whose white vertices are all of degree 2.

**Convention 1.4** Our convention will be as follows:

- we always work with hypermaps;
- if it so happens that all the white vertices of a given hypermap are of degree 2, we erase them in order to simplify the picture and draw the hypermap as a map – and quite often even call it a map;
- in order to be coherent with the previous definitions, the segments joining a black vertex with a white one are called not edges but *half-edges*; then, for example, the degree of a face is equal to the number of surrounding half-edges divided by 2; the term “edge” is reserved for maps.

For a hypermap  $H$  with  $n$  half-edges, the Euler characteristic is computed as follows:

$$\chi(H) = 2 - 2g = B + W + F - n$$

where  $B$  is the number of black vertices,  $W$  is the number of white ones, and  $F$  is the number of faces.

**Encoding by permutations.** Hypermaps admit an encoding by triples of permutations. In constructing this incoding, we twice use the fact that the surface on which the hypermap is drawn is oriented: first, a half-edge has a left bank and a right bank; second, the vicinity of every vertex has a positive and a negative orientation.

Let  $H$  be a hypermap with  $n$  half-edges. We label the half-edges by the labels from 1 to  $n$ , and place the label of a half-edge near its left bank when we go from its black end to the white one. Then we associate to  $H$  the following triple of permutations  $(\sigma, \alpha, \varphi)$  on the set of  $n$  labels:

- a cycle of  $\sigma$  contains the labels of the half-edges incident to a black vertex, taken in the positive (i. e., counterclockwise) direction around this vertex; thus, there are as many cycles in  $\sigma$  as there are black vertices, and the degree of a vertex is equal to the length of the corresponding cycle;
- the cycles of  $\alpha$  correspond, in the same way, to white vertices;
- a cycle of  $\varphi$  contains the labels placed inside a face; these labels are taken in the positive direction around the center of the face; thus, there are as many cycles in  $\varphi$  as there are faces, and the degree of a face is equal to the length of the corresponding cycle.

**Example 1.5** For the hypermap of Figure 6, we obtain the following permutations:

$$\begin{aligned}\sigma &= (1, 2, 3)(4, 5)(6)(7, 8, 9), \\ \alpha &= (1, 4)(2, 9, 3)(5, 6, 7)(8), \\ \varphi &= (1, 5, 9)(2)(3, 8, 7, 6, 4).\end{aligned}$$

For the outer face, the first impression is that the corresponding cycle  $(3, 8, 7, 6, 4)$  turns in the negative direction. In fact, we must look “from the opposite side of the sphere”, or, otherwise, “from the inside of the outer face”, and then the direction becomes positive.

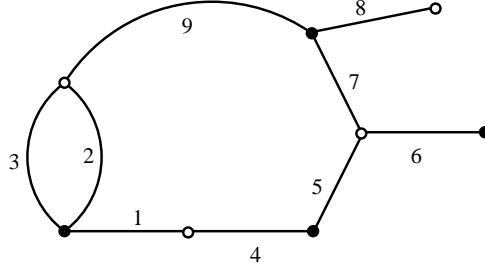


Figure 6: A labeling of the hypermap of Figure 4.

**Remark 1.6** The following very important observation is true for any hypermap (the proof is a simple exercise):

$$\sigma\alpha\varphi = 1.$$

Therefore, in order to encode a hypermap, we may take any two of the three permutations. The form we have chosen is, however, more symmetric.

The correspondence between hypermaps and triples of permutations also works in the opposite direction.

**Proposition 1.7** *To any triple of permutations  $(\sigma, \alpha, \varphi)$  such that*

- *the permutation group  $G = \langle \sigma, \alpha, \varphi \rangle$  is transitive,*
- *$\sigma\alpha\varphi = 1$ ,*

*there corresponds a hypermap.*

The condition of transitivity ensures the connectivity of the corresponding graph.

## 2 Digression: the cartographic group

A fact which rarely attracts attention is that a simple picture drawn on a piece of paper, via the triple of permutations described above, generates a permutation group. We call this group  $G = \langle \sigma, \alpha, \varphi \rangle$  the *cartographic group* corresponding to a (hyper)map. Of course, more often than not the group thus obtained is either  $S_n$  or  $A_n$ . But there exist also many other, more interesting examples. A very small sample is given in Figures 7, 8, 9.

## 3 Belyi functions: planar case

**Definition 3.1 (Belyi function)** Let  $H$  be a planar hypermap with  $n$  half-edges. A rational function  $f$  of degree  $n$  is a *Belyi function* corresponding to  $H$  if  $H$  may be embedded in the Riemann complex sphere  $\overline{\mathbb{C}}$  in such a way that:

1. All black vertices of  $H$  are roots of the equation  $f(x) = 0$ , the multiplicity of each root being equal to the degree of the corresponding vertex.

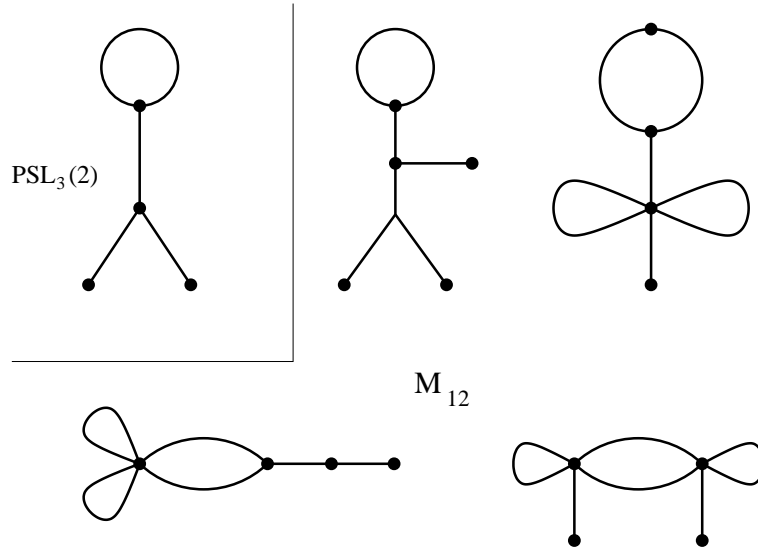


Figure 7: A map with 4 edges representing the group  $\text{PSL}_3(2)$ , and four maps with 6 edges representing the Mathieu group  $M_{12}$ . In total, there are 50 planar maps with 6 edges representing  $M_{12}$ .

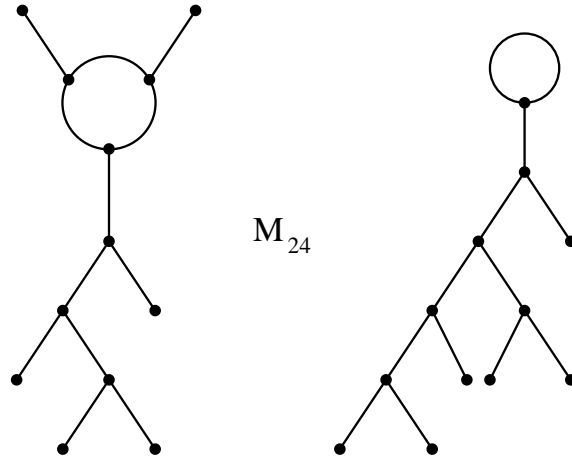


Figure 8: Two maps with 12 edges representing the Mathieu group  $M_{24}$ ; the one on the right was found by N. Adrianov (private communication).

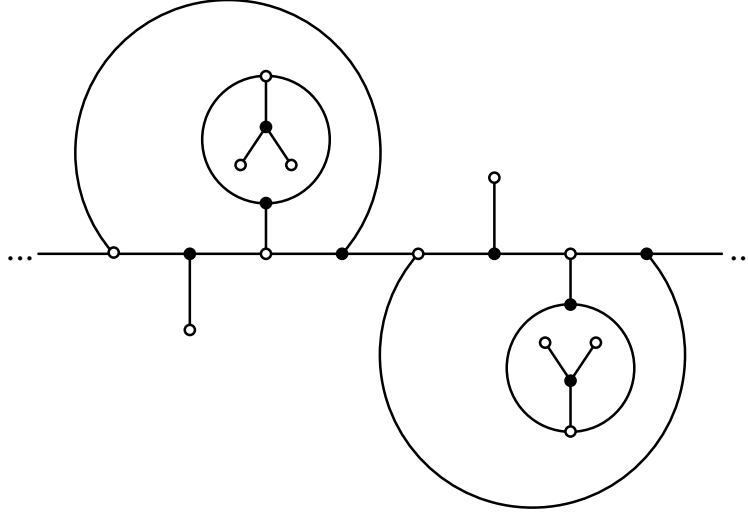


Figure 9: A hypermap with 24 half-edges representing  $M_{24}$  (this example is borrowed from [3]). The horizontal line is an equator on the sphere. All the 6 faces are of degree 4.

2. All white vertices of  $H$  are roots of the equation  $f(x) = 1$ , the multiplicity of each root being equal to the degree of the corresponding vertex.
3. The hypermap itself is the preimage of the segment  $[0, 1]$ , that is,  $H = f^{-1}([0, 1])$ .
4. Inside each face of  $H$  there exists a (single) pole of  $f$  (or, if you like, a root of the equation  $f(x) = \infty$ ), the multiplicity of the pole being equal to the degree of the face. We will call this pole the *center* of the face (of course, it is in no way its “geometric center”).
5. Besides 0, 1, and  $\infty$ , there are no other critical values of  $f$ .

(A *critical point* of  $f$  is a root of its derivative (with a standard change of variables when it comes to infinity); a *critical value* of  $f$  is the value of  $f$  at its critical point.)

In fact, it is not difficult to see that *any* rational function which does not have critical values outside the set  $\{0, 1, \infty\}$  is a Belyi function corresponding to a hypermap. (If there are three critical values but they are not equal to 0, 1 or  $\infty$ , we may apply a linear fractional transformation and place them



to 0, 1, and  $\infty$ .) Therefore, we may take as a definition of a Belyi function the property of not having critical values outside the set  $\{0, 1, \infty\}$ .

The correspondence also works in the opposite direction.

**Theorem 3.2** *For every hypermap  $H$ , there exists a corresponding Belyi function  $f = f(x)$ . This function is unique, up to a linear fractional transformation of the variable  $x$ .*

This theorem is a particular case of Riemann's existence theorem which we will formulate later.

An interesting and highly nontrivial question is, given a hypermap, how to compute the corresponding Belyi function. Let us illustrate some initial stages of such a computation. Suppose we have the “dessin d'enfant” shown in Figure 10.

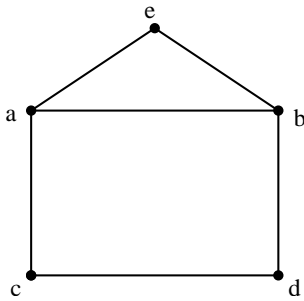


Figure 10: A “dessin d'enfant”

According to the 1st condition of Definition 3.1, the points  $a, b, c, d, e$  must be the roots of the function  $f$  we are looking for, with multiplicities corresponding to the vertex degrees. Therefore, the numerator of  $f$  must have the form

$$(x - a)^3(x - b)^3(x - c)^2(x - d)^2(x - e)^2.$$

In fact, it is much more convenient to write this expression in the form

$$(x^2 + px + q)^3(x^3 + rx^2 + sx + t)^2$$

since the parameters  $p, q, r, s, t$  belong to a smaller and simpler field than  $a, b, c, d, e$  (just imagine  $r, s, t$  being rational numbers while  $c, d, e$  being cubic irrationals).

Our map has 6 edges, which means that the hypermap in question has 6 white vertices, all of degree 2. Therefore, the numerator of the function  $f - 1$  is a product of 6 quadratic factors, or, better (as before), a square of a polynomial of degree 6.

Finally, the map has three faces, of degree 5, 4, and 3 respectively. Therefore, according to the 4th condition of Definition 3.1, the denominator of  $f$  factorizes as  $(x - A)^5(x - B)^4(x - C)^3$ .

As a result we have:

$$f(x) = K \frac{(x^2 + px + q)^3(x^3 + rx^2 + sx + t)^2}{(x - A)^5(x - B)^4(x - C)^3},$$

and

$$f(x) - 1 = K \frac{(x^6 + mx^5 + nx^4 + ux^3 + vx^2 + wx + z)^2}{(x - A)^5(x - B)^4(x - C)^3}.$$

Taking the first expression, computing  $f - 1$  and equating the result to the second expression, we obtain a system of 12 algebraic equations in 15 unknowns  $K, p, q, r, s, t, m, n, u, v, w, z, A, B, C$ .

The three remaining degrees of freedom correspond to the possibility of making a linear fractional transformation of  $x$ . It allows us to choose certain parameters to our convenience. For example, traditionally the center of the outer face is placed to  $\infty$ , thus making the denominator of  $f$  equal to  $(x - B)^4(x - C)^3$ . We may also put, for example,  $B = 0$  (any other choice not equal to  $\infty$  would also be possible), and to choose a value for one more parameter almost arbitrarily. However, one must be careful since certain choices of values for unknowns may turn out to be contradictory. For example, if a map is centrally symmetric then the sum of the positions of black vertices is equal to the sum of the positions of white vertices; hence, we cannot take one sum equal to 0 and the other one equal to 1.

**Galois action.** Now it comes as no surprise that the coefficients of Belyi functions can always be made algebraic numbers. (Of course, we could have a weird idea to take  $B = \pi$  instead of  $B = 0$  and thus to spoil everything; but we will be wise.) This property of Belyi functions leads to the most interesting aspect of the theory of dessins d'enfants, namely, the action of the *absolute Galois group*  $\text{Gal}(\overline{\mathbb{Q}} | \mathbb{Q})$  (the automorphism group of the field  $\overline{\mathbb{Q}}$  of algebraic numbers) on the hypermaps. We act on the coefficients of a Belyi function, replacing them by their Galois conjugates. In this way we obtain another Belyi function which corresponds to another hypermap.

Unfortunately, we cannot develop this subject here since it needs a much longer exposition. Therefore, we limit ourselves with few remarks concerning some combinatorial invariants of the Galois action. The simplest and most obvious such invariant is the triple of cycle structures of the permutations  $\sigma, \alpha, \varphi$ . Indeed, even from our brief discussion above it becomes clear that these cycle structures represent the only information we used while writing down the system of algebraic equations on the coefficients of the function. More advanced invariants are: the automorphism group of a hypermap; its cartographic group; the triple of conjugacy classes of  $\sigma, \alpha, \varphi$  in the cartographic group (a very subtle case is represented, for example, by the Mathieu group  $M_{22}$  in which there exist two algebraically non-conjugate classes,  $4A$  and  $4B$  in the Atlas [4] notation, with cycle structure  $4^4 2^2 1^2$ ; note that hypermaps with  $\sigma, \alpha, \varphi$  all having the cycle structure  $4^4 2^2 1^2$  are planar); and even some highly non-trivial information about the number field to which the coefficients belong can be found in the character table of the cartographic group.

A less advanced but more combinatorial in nature is such an invariant as the property of a map to be self-dual. It involves the operation of composition of which we will speak later. For other types of invariants, such as certain diophantine relations between vertex degrees, see [9].

**Plane trees and Shabat polynomials.** A particular case of a hypermap is a *bicolored plane tree*, that is, a planar hypermap with a single face, see Figure 11. If we put the single pole to infinity, the corresponding Belyi function becomes a polynomial. These polynomials are called *Shabat polynomials*.

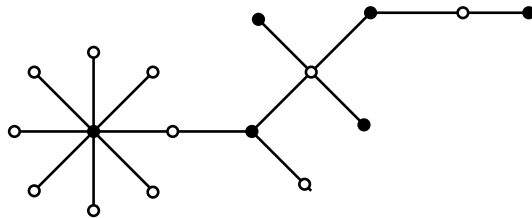


Figure 11: A plane tree.

**Example 3.3** The two simplest trees are shown in Figure 12; their Shabat polynomials are:

- for the star-tree,  $f(x) = x^n$ ;
- for the chain-tree,  $f(x) = T_n(x)$ , the Chebyshev polynomial (critical values, instead of being 0 and 1, are  $\pm 1$ ).

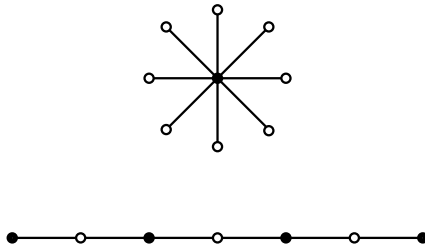


Figure 12: The star-tree and the chain-tree.

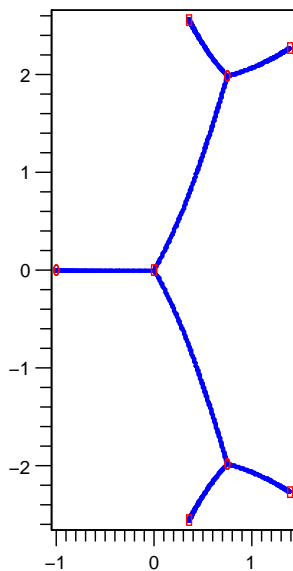


Figure 13: A plane tree obtained as a preimage of a segment via its Shabat polynomial (courtesy of J. B  tr  ma).

One more example is given in Figure 13. The corresponding Shabat polynomial is

$$f(x) = \frac{1}{729} \cdot (2x^2 - 3x + 9)^3(x + 1),$$

while

$$f(x) - 1 = \frac{1}{729} \cdot x^3(8x^4 - 28x^3 + 126x^2 - 189x + 378)$$

(the vertex of degree 3 at the origin must be considered as white).

The whole construction leads to a very unusual property of plane trees:

**Proposition 3.4** *Every plane tree has a canonical geometric form.*

Indeed, linear fractional transformations that don't move infinity are affine transformations. Therefore, the tree obtained via its Shabat polynomial can be rotated, translated, and undergo a homothetic transformation, but all these do not change its geometric form.

## 4 Davenport–Stothers–Zannier bound: an application of Belyi functions

Let  $P, Q \in \mathbb{C}[x]$  be two coprime complex polynomials. The question we would like to discuss is, what is the smallest possible degree of  $P^3 - Q^2$ ?

Denote  $\deg P = 2n$ ,  $\deg Q = 3n$ , and let

$$P^3 - Q^2 = R.$$

In the paper [2] (1965), the authors have formulated two conjectures:

- $\deg R \geq n + 1$ ;
- this bound is sharp; that is, it is attained for infinitely many values of  $n$  (as we will see shortly, it is, in fact, attained for every  $n$ ).

The first conjecture was proved by Davenport in the same year [6]. The second one remained open for 16 years and was finally proved by Stothers [15]; later on, it was reproved once again and generalized by Zannier [18]. It is this second, and apparently more difficult conjecture (the sharpness of the bound) that we will prove now.

Let us denote

$$f = \frac{P^3}{R}$$

and remark that

$$f - 1 = \frac{P^3 - R}{R} = \frac{Q^2}{R}.$$

The computational part of the proof is finished. Now we make the following

**Assumption 4.1** Assume that  $f$  is a Belyi function.

What follows is just a translation of the initial problem data into combinatorial properties of the corresponding hypermap.

- The numerator of  $f$  is  $P^3$ ,  $\deg P = 2n$ ; this means that the hypermap in question has  $2n$  vertices, all of them of degree 3.
- The numerator of  $f - 1$  is  $Q^2$ ,  $\deg Q = 3n$ ; this means that the hypermap has  $3n$  white vertices, all of them of degree 2. It is a map! Forget white vertices and half-edges, let us speak of edges.
- For the number of faces  $F$ , the Euler formula gives

$$2n - 3n + F = 2 \quad \Rightarrow \quad F = n + 2.$$

Put the center of one of the faces to  $\infty$ ; then the  $\deg R$  becomes equal to the sum of the degrees of the remaining  $n + 1$  faces.

Summing up: in order to prove the statement, we must construct a planar map having  $3n$  edges,  $2n$  vertices of degree 3, and all its  $n + 1$  “finite” faces of degree 1.

We would like to underline the fact that up to now we did nothing: we have only translated the initial problem into a combinatorial language. But now the problem becomes trivial: its solution is shown in Figure 14.

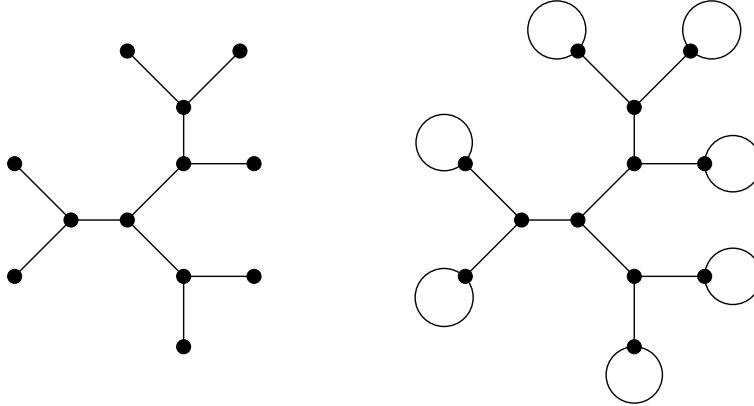


Figure 14: Draw a tree with all its internal vertices of degree 3 and attach a loop to each leaf.

## 5 Composition

Many operations with maps and hypermaps can be represented using a composition of functions:

$$h : X \xrightarrow{f} Y \xrightarrow{g} Z.$$

Here, at least in this section,  $X$ ,  $Y$ , and  $Z$  all mean the complex sphere  $\overline{\mathbb{C}}$  but we use different letters in order to clearly distinguish on which level our operations take place.

**Dual map.** Let  $f$  be a Belyi function for a map. How to find a Belyi function to its dual map? It is very simple: we need to exchange vertices with face centers while leaving the (invisible) white vertices at the same places. In the language of Belyi functions this just means exchanging critical values 0 and  $\infty$  while leaving 1 unchanged. The function that makes this operation is  $g_1(y) = 1/y$ . Therefore, the Belyi function corresponding to the dual map is  $h(x) = 1/f(x)$ . The operation, of course, remains meaningful also for hypermaps.

**Doubling edges.** For a map, we would like to replace every edge by two parallel edges surrounding a face of degree 2. For a hypermap, a new face must surround each white vertex, and the degree of this face must be equal to the degree of the corresponding white vertex. The degrees of all black vertices become twice bigger than before; the degrees of all “old” faces are preserved. Note that even for a hypermap the result of this operation will be a map.

Using the same reasoning as before, we understand that we need a function  $g(y)$  which would send 0 to 0 (doubling the multiplicity),  $\infty$  to  $\infty$ , and 1 to  $\infty$ . And we easily find  $g(y) = y^2/(y-1)$ . However, there is a mistake here. The function  $g$  creates a new critical value:  $g'(y) = y(2-y)/(y-1)^2$  has two roots 0 and 2, and  $g(0) = 0$ , which is OK, but  $g(2) = 4$ . The function  $h = g \circ f$  thus constructed will have four critical values:  $z = 0, 1, 4, \infty$ . The mistake is very easy to correct: we must just take  $g_2(y) = y^2/4(y-1)$ .

**A map together with its dual.** We have a map, and we would like to draw it together with its dual in the same picture. Note that now the white vertices become “visible” since they all acquire degree 4. The function  $g$  we are looking for must send 1 to 1 doubling the multiplicity, and send both 0 and  $\infty$  to 0. Such a function is  $g_3(y) = 4y/(y+1)^2$ .

**Medial map.** Taking a map, we transform all its (invisible) white vertices of degree 2 into black vertices of degree 4 by joining them successively by new edges inside each face. The face degrees remain the same while all black vertices are replaced by faces whose degrees are equal to degrees of the former vertices. The corresponding transformation is realized by  $g_4(y) = -(y - 1)^2/4y$ .

**Truncation.** When applied to a polyhedron, this operation means “cutting vertices”. For an arbitrary map, it means replacing each black vertex by a face of the same degree, and inserting a new black vertex of degree 3 inside each edge (thus, each edge gets two new vertices which are placed near both its ends).

It is important to understand that the functions  $g_1, g_2, g_3, g_4$  constructed up to now are themselves Belyi functions. We leave it to the reader to draw the corresponding hypermaps. All the above operations can be understood in the following way: instead of drawing the segment  $[0, 1]$  on  $Y$ -sphere, we draw there a simple hypermap  $H$ , which is sent to the segment  $[0, 1]$  on the  $Z$ -sphere by the function  $g$ . Then, the hypermap on the  $X$ -sphere is obtained as  $f^{-1}(H)$ . A necessary condition for this operation to be successful is the following one: the critical values of  $f$ , that is,  $0, 1, \infty \in Y$ , must find themselves among the  $g$ -preimages of  $0, 1, \infty \in Z$ , that is, among vertices and face centers of  $H$ .

The case of the truncation is not an exception. In this case, the hypermap  $H$  on the  $Y$ -sphere looks as is shown in Figure 15. The corresponding Belyi function is  $g_5(y) = (4y - 1)^3/27y$ . (Thus, the position of the black vertex is  $y = 1/4$ , and the position of the white vertex of degree 2 may be easily computed: it is the double root of  $g_5(y) - 1$ , which is  $y = -1/8$ .)

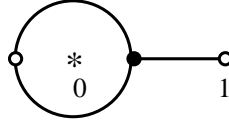


Figure 15: If we replace the segment  $[0, 1]$  by this hypermap, its preimage will give us a truncated map.

**Remark 5.1** One can verify that  $g_5 \circ g_4 = g_5 \circ g_2$ . At first sight, this seems to be not very surprising since  $g_4$  and  $g_2$  are Belyi functions corresponding to the same map which is only placed in two different ways on the sphere.



Hence, this fact represents an invariance of a function under certain linear fractional transformations. Yes, but the invariance of *which* function? It is, in fact, the composition which is invariant, and the map corresponding to it has a symmetry of order 3, but neither  $g_5$  itself nor  $g_4$  or  $g_2$  are symmetric.

This is one of the examples of rational functions represented as a composition in a non-unique way. The problem was studied by Ritt in [13], [14] but completely solved (in [13]) only for polynomials. Essentially, only the polynomials  $x^n$  and  $T_n(x)$  of Example 3.3 with  $n$  non prime allow a non-uniqueness. For rational functions in general, as Ritt wrote, “there is a much greater variety of possibilities”.

**Edge subdivision.** In order to subdivide every edge of a map in  $n$  parts (that is, insert in it  $n - 1$  new vertices) we must apply the Chebyshev polynomial  $T_n(y)$  (see Example 3.3) normalized in such a way as to have critical values 0 and 1 instead of  $\pm 1$ , namely,  $g_6(y) = (T_n(y) + 1)/2$ . Don’t think that the parts in which an edge is subdivided are equal.

**Rotational symmetry.** The other “simplest” polynomial of Example 3.3 is commonly used in another way. Suppose we have a hypermap on the  $X$ -sphere which is invariant under rotation through the angles  $k(2\pi/n)$ ,  $k = 0, \dots, n - 1$  around one of its elements (i.e., a vertex or a face center). Note that this rotation center cannot be alone since a rotation of the sphere is always around an axis that passes through two antipodal points. Then, putting one of these points to 0 and the other one to  $\infty$ , and applying  $f(x) = x^n$  we obtain a *reduced*, or a *quotiented* hypermap on the  $Y$ -sphere. After this operation, the degrees of the elements placed at 0 and  $\infty$  become divided by  $n$  while all other elements of the initial hypermap are subdivided in blocs of size  $n$  (orbits of the rotation) and each bloc becomes represented by a single element of the reduced hypermap. Now we can apply a Belyi function  $g(y)$  corresponding to this reduced hypermap.

**A Belyi function of a cube.** We are now able to attack a more complicated task and compute a Belyi function for the map of the cube. Let us place the center of a face at 0, and the center of the opposite face, at  $\infty$ . Then, applying the function  $x^4$  we get the map shown in Figure 16. We recognize in it a truncation of a very simple map consisting of a unique edge  $[0, \infty]$ . Incidentally, the Belyi function of this latter map is  $g_3$ . Summing up, we obtain a Belyi function of the cube as a composition  $f_{\text{cube}}(x) = g_5 \circ g_3 \circ x^4$ .

The result of the computation gives

$$f_{\text{cube}}(x) = -\frac{1}{108} \cdot \frac{(x^8 - 14x^4 + 1)^3}{x^4(x^4 + 1)^4}.$$

This function, as well as Belyi functions for the other Platonic maps were found by Felix Klein in 1875, see [8]. In [10], we have computed Belyi functions for all the Archimedean maps. The paper is a long series of exercises in composition, with two relatively more difficult computations for chiral maps, and with a very difficult one for the only non vertex-transitive Archimedean map.

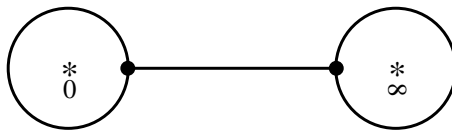


Figure 16: A cube quotiented by a rotational symmetry of order 4.

**Not everything is that simple.** Let us consider a family of maps with 6 edges, with vertex degrees 6, 3, 2, 1, and with face degrees also 6, 3, 2, 1. The first, and rather challenging question is to find them all and to say how many of them there are. I never met a single person (including myself) who would find a correct answer within 3 days.

In a paper format, I cannot keep the answer secret for 3 days, so here it is: there are 18 maps with this set of vertex and face degrees. (Try to find them now!) For 6 of them the cartographic group is  $A_{12}$ ; for other 8, it is  $M_{12}$ ; and for remaining 4 maps we get an imprimitive group. This is why the family of 18 splits into three Galois orbits, of size 6, 8, and 4, respectively. One of the maps with an imprimitive cartographic group is shown in Figure 17 on the left.

According to a famous theorem due to Ritt, a ramified covering with an imprimitive monodromy group is a composition of two (or more) coverings. (Ritt [13] formulated this theorem only for polynomials but it remains true also in a more general setting.) In our case, this means that the Belyi function  $h$  of the map on the left in Figure 17 is a composition:  $h = g \circ f$ . Does it jump to the eye when you look at the picture? I bet it doesn't.

But watch out! In this case the function  $f$  itself is not a Belyi function. Indeed, it has not 3 but 4 critical values. However, it turns out that these

4 critical values are two (black) vertices and two face centers of a simpler map which is obtained via a Belyi function  $g$ . (In Figure 17, critical values of  $f$  are marked by little squares.) Hence,  $g$  sends all the four critical values of  $f$  to 0 and to  $\infty$ , and also creates a new critical value 1. This is why the composition  $h = g \circ f$  is a Belyi function.

We do not present the results of the computations. Don't forget that this Galois orbit contains 4 elements; therefore, the coefficients of our rational function are themselves expressions involving roots of an equation of degree 4.

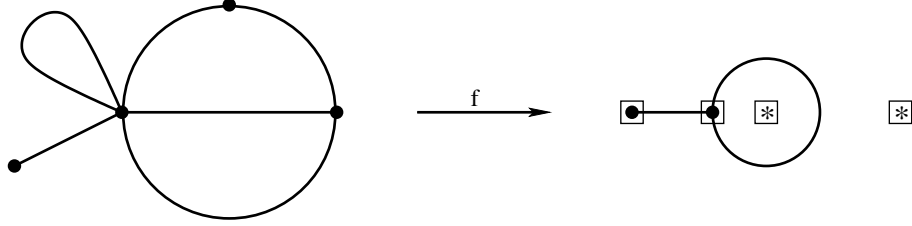


Figure 17: A decomposable map, and how it covers another map. Little squares show critical values of  $f$ .

## 6 Belyi functions and Julia sets

**Definition 6.1 (Dynamical Belyi function)** A Belyi function  $f : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$  is called *dynamical* if it sends  $\{0, 1, \infty\}$  to  $\{0, 1, \infty\}$  (any of the 27 possible combinations is allowed).

**Remark 6.2** We know that

$$f^{-1}(\{0, 1, \infty\}) = \{\text{black vertices, white vertices, and face centers}\}.$$

Hence, a Belyi function  $f$  is dynamical if and only if the corresponding hypermap is positioned on the complex sphere in such a way that the three points 0, 1, and  $\infty$  are “occupied” by black or white vertices or face centers (any combination is possible). If this is not the case then we can achieve this by using a linear fractional transformation which permits to put any three points of our choice to any three positions.

The reason of imposing this conditions is that a composition  $h = g \circ f$  of two dynamical Belyi functions  $f$  and  $g$  is once again a dynamical Belyi

function. Therefore, it is possible to iterate a dynamical Belyi functions

$$f^n = f \circ f \circ \dots \circ f \quad (n \text{ times})$$

and get as a result a Belyi function.

**Definition 6.3 (Fatou and Julia sets)** For a sequence of iterations  $f^n$  the *Fatou set* is an open set  $F \subseteq \overline{\mathbb{C}}$  such that for any  $x \in F$  there exists a neighborhood  $U$  of  $x$  on which the iterations  $f^n$  form a “normal family”: for any sequence  $n_1, n_2, \dots$  and for any compact  $K \subset U$  there exists a subsequence for which the iterations converge uniformly on  $K$ . The *Julia set* is the complement to the Fatou set:  $J = \overline{\mathbb{C}} \setminus F$ .

Informally speaking, the Fatou set is the set of the regular behavior of the iterations, and the Julia set is the set of their irregular behavior.

We all saw beautiful pictures of fractal Julia sets. Less attractive visually but interesting from the theoretical point of view is the case of the so called “complete chaos”.

**Definition 6.4 (Complete chaos)** The dynamical system  $f^n$ ,  $n \geq 1$ , is completely chaotic if its Julia set coincides with the whole sphere:  $J = \overline{\mathbb{C}}$ .

The following famous theorem due to D. Sullivan (1985), see [16], gives a criterion for a complex dynamical system to be completely chaotic.

**Theorem 6.5** *If all critical points of  $f$  eventually (i. e., after some number of iterations) become periodic but are not themselves periodic then  $J = \overline{\mathbb{C}}$ .*

For Belyi functions, critical points are under control; namely, they are black and white vertices and face centers, though not all of them but only those of degree  $> 1$ . Sullivan himself gave the following example:

$$f(x) = \left( \frac{x-2}{x} \right)^2.$$

Incidentally, this is a Belyi function corresponding to the hypermap shown in Figure 18.

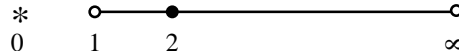


Figure 18: The hypermap corresponding to Sullivan’s example.

Let us read the trajectories of critical points without using the formula but only looking at the picture. There are only two critical points: the point 2 (a vertex of degree 2), and the point 0 (the center of the face of degree 2). The point 2, being a black vertex, by definition goes to 0. The point 0, being a face center, goes – also by definition – to  $\infty$ . The point  $\infty$ , being a white vertex, goes to 1. Finally, the point 1, being a white vertex, also goes to 1. Summing up, we obtain

$$2 \mapsto 0 \mapsto \infty \mapsto 1 \mapsto 1.$$

Thus, 1 is a periodic point, and both critical points 2 and 0 go to 1 after some number of iterations. What is important is the fact that the point 1 itself is not critical: while being a vertex it is a vertex of degree 1. Hence, the requirement of the theorem is satisfied: no critical points are themselves periodic.

We would like to propose here one more example to Sullivan’s theorem, see Figure 19. Our example is somewhat less explicit since we are unable to give an explicit expression of the rational function in question. Nevertheless, we may affirm that the conditions of the theorem are satisfied and that therefore this is a case of complete chaos.

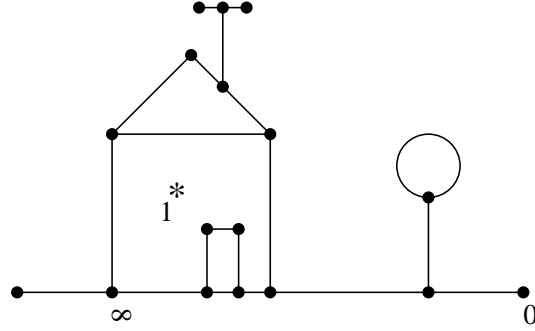


Figure 19: Iterations of the Belyi function corresponding to this “dessin d’enfant” create a dynamical system with a complete chaos.

Indeed, after the first application of the Belyi function all critical points go to 0, 1, and  $\infty$ . As before, we can easily follow the subsequent applications of this function to these three points; they give

$$1 \mapsto \infty \mapsto 0 \mapsto 0.$$

Thus, all critical points of this function after several iterations go to 0 which

is a periodic point. However, this point, being a vertex of degree 1, is not itself critical. QED.

We see that any hypermap having an element (a vertex or a face) of degree 1 permits to construct a similar example, quite often even several examples. In principle, it should be possible to study the behavior of the corresponding dynamical systems in a purely combinatorial and group-theoretic way, though there remain many obstacles to overcome. And, sure enough, such a study must be extended from Belyi functions to *postcritically finite* complex dynamical systems (see [11]) which are also *rigid* and therefore defined over  $\overline{\mathbb{Q}}$ .

## 7 Non-planar case

We remind to the reader that for every genus  $g \geq 1$  there are infinitely many non-isomorphic Riemann surfaces (all of them are homeomorphic to each other as topological surfaces). Therefore, for a bigger genus we must speak not about a Belyi function but about a *Belyi pair*.

**Definition 7.1 (Belyi pair)** A pair  $(X, f)$ , where  $X$  is a Riemann surface and  $f : X \rightarrow \overline{\mathbb{C}}$  is a meromorphic function on  $X$ , is called a *Belyi pair* if  $f$  is unramified outside the set  $\{0, 1, \infty\} \subset \overline{\mathbb{C}}$ .

In the same way as in the planar case, the preimage  $f^{-1}([0, 1])$  is a hypermap embedded (in a very specific way) in the surface  $X$ . The preimages of 0 are its black vertices, the preimages of 1, white vertices, and the preimages of  $\infty$  are the “face centers” (exactly one of them inside each face). Also as before, the multiplicities of these preimages correspond to the degrees of vertices or faces.

In the opposite direction, for any hypermap (of any genus) there exists a corresponding Belyi pair. Please, note: the hypermap specifies not only a Belyi function; it also specifies a Riemann surface on which this function is defined. But it is better to formulate a more general theorem.

There are two ways of representing Riemann surfaces. We may represent them as complex algebraic curves, i.e., sets of solutions of algebraic equations (one equation with two unknowns, or two equations with three unknowns, etc.). Or, otherwise, we may represent them as ramified coverings of the complex Riemann sphere  $\overline{\mathbb{C}}$ . In order to specify a ramified covering of degree  $n$  we must specify two sequences:

- a sequence of  $k$  ramification points  $y_1, y_2, \dots, y_k \in \overline{\mathbb{C}}$  (without any constraint imposed on them);
- a sequence of  $k$  permutations  $g_1, g_2, \dots, g_k \in S_n$  which act transitively on  $n$  points and such that  $g_1 g_2 \dots g_k = 1$ .

The permutations represent the *monodromy*:  $g_i$  shows how the preimage  $x$  goes from a sheet of the covering to another one when its image  $y = f(x)$  makes a turn around  $y_i$ .

**Theorem 7.2 (Riemann's existence theorem)** *For any two sequences  $y_1, y_2, \dots, y_k \in \overline{\mathbb{C}}$  and  $g_1, g_2, \dots, g_k \in S_n$  satisfying the above conditions, there exists a Riemann surface  $X$  and a meromorphic function  $f : X \rightarrow \overline{\mathbb{C}}$  such that  $y_1, y_2, \dots, y_k$  are the ramification points (or, in another terminology, the critical values) of  $f$ , and  $g_1, g_2, \dots, g_k$  are the corresponding monodromy permutations. Such a Riemann surface is unique up to an automorphism.*

The surface  $X$  also does not change if we apply an automorphism of  $\overline{\mathbb{C}}$ , that is, a linear fractional transformation. Doing that, we can put once and for all the three last critical values  $y_{k-2}, y_{k-1}, y_k$  to the fixed positions 0, 1, and  $\infty$ . The covering now depends discretely on  $k$  permutations  $g_1, g_2, \dots, g_k$  (in fact, on  $k-1$  permutations since  $g_k$  can be computed out of  $g_1, g_2, \dots, g_{k-1}$ ), and it depends continuously on  $k-3$  complex parameters  $y_1, y_2, \dots, y_{k-3}$ . Thus, the following question comes naturally:

**Question 7.3** What happens when  $k = 3$ ?

First, the case  $k = 3$  corresponds to hypermaps, where  $g_1 = \sigma$ ,  $g_2 = \alpha$ ,  $g_3 = \varphi$ . Second, the covering becomes *rigid*: it does not have any continuous parameters and can change only discretely.

Is it possible to represent an arbitrary Riemann surface as a covering with only three ramification points? The answer is no, but the class of representable surfaces is the most interesting one. Namely, are representable in this way the surfaces *defined over*  $\overline{\mathbb{Q}}$ . The following theorem was proved by G. Belyi in 1979 [1]. Grothendieck [7] wrote about it: “I do not believe that a mathematical fact has ever struck me quite so strongly as this one, nor had a comparable psychological impact”.

**Theorem 7.4 (Belyi theorem)** *A meromorphic function  $f : X \rightarrow \overline{\mathbb{C}}$  unramified outside  $\{0, 1, \infty\}$  exist if and only if the Riemann surface  $X$  is*

defined over the field  $\overline{\mathbb{Q}}$  of algebraic numbers. This means that  $X$  can be realized as an algebraic curve with all the equations having algebraic numbers as their coefficients. In this case, the function  $f$  is also defined over  $\overline{\mathbb{Q}}$ . This means that it can be realized as a rational function in variables defining  $X$ , with its coefficients being algebraic numbers.

Thus, the absolute Galois group also acts on hypermaps of higher genera via its simultaneous action on  $X$  and  $f$ .

As for the computational aspect, the situation is hopeless. Generally, not only are we unable to solve the corresponding algebraic equations, most often we don't know how to write them down explicitly. We can only guarantee that such equations exist.

However, some very special cases can be solved. A beautiful example is given in the next section.

## 8 Fermat curve

Consider the curve

$$F = \{(x : y : z) \mid x^n + y^n = z^n\} \subset \mathbb{CP}^2$$

and its projection on the first coordinate given by

$$f : \mathbb{CP}^2 \rightarrow \overline{\mathbb{C}} = \mathbb{CP} : (x : y : z) \mapsto (x : z).$$

The function  $f$  is not defined for the points  $x = z = 0$ ,  $y \neq 0$ , but, fortunately, these points do not belong to  $F$ . The point  $\infty \in \overline{\mathbb{C}}$  is represented by  $(x : 0) \in \mathbb{CP}$ . Above this point, the equation  $x^n + y^n = 0$  has  $n$  distinct solutions  $(x : \zeta x)$  where  $\zeta$  takes  $n$  values of  $\sqrt[n]{-1}$ .

Having stated that, we may now consider only the affine part of  $F$ , that is,

$$F^\circ = \{(x, y) \mid x^n + y^n = 1\} \subset \mathbb{C}^2$$

and its projection on the first coordinate given by

$$f : \mathbb{C}^2 \rightarrow \mathbb{C} : (x, y) \mapsto x.$$

For a fixed value of  $x$ , the equation  $y^n = 1 - x^n$  considered as an equation on  $y$ , usually has  $n$  distinct solutions. The exceptions are the  $n$ th roots of unity: when  $x = \sqrt[n]{1}$  the number of solutions for  $y$  is reduced to one. Therefore,  $f$  has  $n$  critical values, namely, the  $n$ th roots of unity. (As we



have said before,  $\infty$  is not a critical value of  $f$  since there are  $n$  distinct points of the curve  $F$  over  $\infty$ .)

Now, applying the function  $g(x) = x^n$  we push all these critical values to 1, and create two new critical values: 0 and  $\infty$ . Therefore, the composition

$$h : F \xrightarrow{f} \overline{\mathbb{C}} \xrightarrow{g} \overline{\mathbb{C}} : (x, y) \mapsto x \mapsto x^n$$

is a Belyi function on  $F$ . Since  $\deg f = n$  and  $\deg g = n$ , the degree of their composition  $h = g \circ f$  is  $\deg h = n^2$ .

In this case, it is more interesting to consider not the preimage of the segment  $[0, 1]$  but the preimage of the whole real line. More exactly, we mark the points 0, 1, and  $\infty$  on  $\overline{\mathbb{C}}$  by  $\bullet$ ,  $\circ$ , and  $*$ , and join them by the segments  $[0, 1]$ ,  $[1, \infty]$ , and  $[\infty, 0]$ . In this way we get a triangulation of the sphere, two triangular faces being the upper and the lower half-planes. Now, simple reasoning leads to the following results:

- The graph embedded in  $F$  has  $3n$  vertices:  $n$  vertices of the type  $\bullet$ ,  $n$  vertices of the type  $\circ$ , and  $n$  vertices of the type  $*$ .
- The degree of each vertex is  $2n$ ; therefore, it is joined to every vertex of two other types. Hence, the graph on  $F$  is  $K_{n,n,n}$ , the complete tripartite graph.
- Except 0, 1, and  $\infty$  (which are critical values of  $h$ ), all other points of  $\overline{\mathbb{C}}$  are “repeated”  $n^2$  times on  $F$ . Therefore, the map on  $F$  has  $3n^2$  edges and  $2n^2$  faces.
- The genus of  $F$  can now be easily computed:

$$2 - 2g = 3n - 3n^2 + 2n^2 \quad \Rightarrow \quad g = \frac{(n-1)(n-2)}{2}.$$

- All faces of the map on  $F$  are triangles; therefore, this is an embedding of the least genus of the graph  $K_{n,n,n}$ .

The fact that the least possible genus of an embedding of the graph  $K_{n,n,n}$  is equal to  $(n-1)(n-2)/2$ , was established by White in 1969 [17] and by Ringel and Youngs in 1970 [12]. They would have been very much amazed had they known that their work was related to the Fermat equation.

If, according to previous conventions, we consider the preimage of the segment  $[0, 1]$ , we obtain a hypermap which is a regular embedding of the complete bipartite graph  $K_{n,n}$ . The embeddings of complete bipartite graphs are thoroughly studied by Gareth Jones. The construction using the Fermat curve also belongs to him.

**A more general construction.** The above example is a very particular case of the following more general scheme. Let  $f$  and  $g$  be two planar Belyi functions. Here the word “planar” means that they are rational functions of one complex variable, and the hypermaps corresponding to them are planar. Let  $\deg f = m$ ,  $\deg g = n$ . Consider the algebraic curve

$$X = \{(x, y) \mid f(x) = g(y)\}.$$

Then the function on  $X$  given by  $h(x, y) = f(x)$  (or, equivalently,  $h(x, y) = g(y)$  since  $f(x) = g(y)$  on  $X$ ) is a Belyi function.

Indeed, if  $z \neq 0, 1, \infty$  then the equation  $f(x) = z$  has  $m$  distinct solutions and the equation  $g(y) = z$  has  $n$  distinct solutions; therefore, the curve  $X$  contains  $mn$  distinct points  $(x, y)$  for which  $f(x) = g(y) = z$ . Only for  $z = 0, 1, \infty$  the number of solutions may become smaller. For the Fermat curve,  $f(x) = x^n$  and  $g(y) = 1 - y^n$ .

What is even more interesting, the hypermap on  $X$  corresponding to the Belyi function  $h$  can be constructed combinatorially, using only the information about two planar hypermaps corresponding to  $f$  and  $g$ .

However, this direction is not yet properly explored.

**Conclusion.** We must finish this, already long but inevitably too short, exposition of Belyi functions, dessins d’enfants, and related topics. We would like to conclude it by the following remark: such a visibly simple object as a triple of permutations  $(\sigma, \alpha, \varphi)$  acting transitively and satisfying the condition  $\sigma\alpha\varphi = 1$  (or, if you prefer, just a pair of permutations  $(\sigma, \alpha)$  acting transitively) leads to a great variety of interesting mathematical structures, namely:

- maps or hypermaps – with all their innumerable combinatorial properties extensively studied by many researchers;
- cartographic groups – with a possibility of a pictorial representation of groups, and with groups themselves being Galois invariants and character tables being a useful tool for enumeration and other things;
- Riemann surfaces – with the underlying notion of complex structure, with their representation as algebraic curves, etc.;
- number fields (finite extensions of  $\mathbb{Q}$ ) – with the corresponding Galois groups and other properties.

Sure enough, all these structures are interrelated, and this is the most attractive quality of the subject. I often recall the phrase once told by Littlewood about Ramanujan, that every positive integer was one of his personal friends. I would say that working with dessins d'enfants leads to a sort of an intimate acquaintance with many maps, while other branches of map theory often tend to consider maps in huge herds whose members lack any individuality.

**Acknowledgments.** I was happy to plunge once again into the stimulating atmosphere of Com<sup>2</sup>MaC during the conference “Applications of Group Theory to Combinatorics” (July 2007). I am grateful to the center director Professor Jin Ho Kwak and to all the conference organizers who gave me this opportunity to meet old friends and colleagues and to acquire new ones.

## References

- [1] **Belyi G. V.** On Galois extensions of a maximal cyclotomic field. – *Mathematics USSR Izvestija*, vol. 14 (1980), no. 2, 247–256. (Original in Russian: *Izvestiya Akademii Nauk SSSR*, vol. 14 (1979), no. 2, 269–276.)
- [2] **Birch B. J., Chowla S., Hall M., Jr., Schinzel A.** On the difference  $x^3 - y^2$ . – *Norske Vid. Selsk. Forh. (Trondheim)*, 1965, vol. **38**, 65–69.
- [3] **Conder M.** The symmetric genus of the Mathieu groups. – *Bull. of the London Math. Soc.*, 1991, vol. **23**, no. 5, 445–453.
- [4] **Conway J. H., Curtis R. T., Norton S. P., Parker R. A., Wilson R. A.** (With computational assistance from **J. G. Thackray**.) Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups. – Clarendon Press, Oxford, 1985, xxxiv+252 pp. (Reprinted in 2005 with corrections.)
- [5] **Cori R.** Un code pour les graphes planaires et ses applications. – *Astérisque*, vol. **27**, 1975.
- [6] **Davenport H.** On  $f^3(t) - g^2(t)$ . – *Norske Vid. Selsk. Forh. (Trondheim)*, 1965, vol. **38**, 86–87.
- [7] **Grothendieck A.** Esquisse d’un programme (1984). – In: L. Schneps, P. Lochak, eds. “Geometric Galois Action. Vol. 1: Around Grothendieck’s *Esquisse d’un Programme*”, London Math. Soc. Lecture Notes

Series, vol. 242, Cambridge Univ. Press, 1997, 5–48. (English translation: “Sketch of a programme”, the same volume, p. 243–284.)

- [8] **Klein F.** Vorlesungen über das Ikosaeder und die Auflösung der Gleichungen vom fünften Grade. – Leipzig, 1884. Reprinted by Dover Publ.: **Klein F.** The Icosahedron and the Solution of Equations of the Fifth Degree, 1956.
- [9] **Lando S. K., Zvonkin A. K.** Graphs on Surfaces and Their Applications. – Springer-Verlag, 2004.
- [10] **Magot N., Zvonkin A.** Belyi functions for Archimedean solids. – *Discrete Math.*, 2000, vol. **217**, 249–271
- [11] **Pilgrim K. M.** Combinations of Complex Dynamical Systems. – Springer-Verlag, 2003 (Lecture Notes in Math., vol. **1827**).
- [12] **Ringel G., Youngs J. W. T.** Das Geschlecht des symmetrischen vollständigen dreifärbbaren Graphen. – *Comm. Math. Helv.*, 1970, vol. **45**, 152–158.
- [13] **Ritt J. F.** Prime and composite polynomials. – *Trans. Amer. Math. Soc.*, 1922, vol. **23**, no. 1, 51–66. (Errata: 1922, vol. **23**, no. 4, p. 431.)
- [14] **Ritt J. F.** Permutable rational functions. – *Trans. Amer. Math. Soc.*, 1923, vol. **25**, no. 3, 399–448. (Errata: 1924, vol. **26**, no. 4, p. 494.)
- [15] **Stothers W. W.** Polynomial identities and Hauptmoduln. – *Quart. J. Math. Oxford, ser. 2*, 1981, vol. **32**, no. 127, 349–370.
- [16] **Sullivan D.** Quasiconformal homeomorphisms and dynamics. I: Solution of the Fatou–Julia problem on wandering domains. – *Ann. Math.*, 1985, vol. **122**, 401–418.
- [17] **White A. T.** The genus of the complete tripartite graph  $K_{mn,n,n}$ . – *J. Combinat. Theory*, 1969, vol. **7**, 283–285.
- [18] **Zannier U.** On Davenport’s bound for the degree of  $f^3 - g^2$  and Riemann’s Existence Theorem. – *Acta Arithmetica*, 1995, vol. **71**, no. 2, 107–137.