Università degli Studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

Corso di Laurea in Matematica

The Four Color Theorem: from Graph Theory to proof assistants.

Relatore:

dott. Samuele Maschio

Laureanda:

Sofia Gambarin

Matricola N. 2056072

Anno Accademico 2022/2023, Sessione di Laurea: Settembre

Contents

Chapter 1

Introduction

In 1852 the English mathematician Francis Guthrie, while coloring the map of England, noticed that four colors were enough for two regions that share a boundary to do not have the same color. Thus, he conjectured that four colors were enough to color any map. This conjecture has been immediately recognised as the "Four Color Conjecture" which later became the "Four Color Theorem". Guthrie, who could not solve the conjecture by himself, asked his brother Frederick and his professor Augustus De Morgan for trying to solve the problem. Despite its simple-sounding statement, Guthrie gave rise to a problem that took more than a century and a half as well as the work of many mathematicians to be proven.

As we shall see below, the Four Color Theorem, besides being a challenge for mathematicians, represents the "marriage" between mathematics and computer science. In other words computer programming was applied in mathematics in order to prove a theorem. Initially these techniques were not easily accepted by the mathematical world since the concept of proof had always been that of absolute truth checkable by any mathematician, while the employment of computer programming appeared unreliable.

In the following we analyse the history of the Four Color Theorem, from Guthrie's conjecture to the final (assisted) proof by Georges Gonthier.

In Chapter 2 we give some notions of Graph Theory, to better frame the problem. We will use Graph Theory as an environment for proving the theorem, thus as a mathematical model for geographic maps.

In Chapter 3 we analyse the property of planarity for a graph along with a very important theorem concerning such property: Kuratowski's theorem.

In the first part of Chapter 4 we define and give an account on graph-colorability. After that we state and prove the "Six Color Theorem" and the "Five Color Theorem".

Chapter 5 represents the main section of this work since the statement of the "Four Color Theorem" is given. In the first part we look at the problem from a historical point of view [13]. Then, we focus on the correspondence between geographic maps, planar graphs and triangulations. After that, one of the most important attempts to prove the theorem is described: we highlight the idea (by Kempe) that underlies the proof [12]. Moreover, we put our attention on Appel and Haken's proof [5], [4]: we describe its main concepts, reducibility [3] and discharging [2]. Finally, two more proofs (the one by Robertson et al. [14] and those by Gonthier) are briefly discussed and compared to each other. In particular, the increasing involvement of computer programming is emphasized.

In the last chapter, we present the main features of proof assistants [11] together with the linked problem of reliability [9]. To conclude, some philosophical considerations on the concept of mathematical proof related to the employment of computers in mathematical proofs are given [15].

Chapter 2

Basics of Graph Theory

2.1 DEFINITION. A *graph* G consists of a finite non-empty set V together with a prescribed set X of unordered pairs of distinct elements of V.

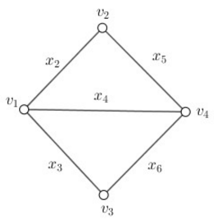
From now on, we will write $V(G)$ and $X(G)$ in place of V and X, respectively when it is not clear which graph we are taking into account: this might avoid some ambiguous situations. Moreover, if G is a graph with $|V| = p$ and $|X| = q$, then we call G a (p, q) graph.

We will call the elements of V *vertices* and the elements of X *edges*. Thus, a (p, q) graph is a graph whose set of vertices has cardinality p and whose set of edges has cardinality q.

We say that two vertices $u, v \in V$ are *adjacent* if there exists an edge $x \in X$ such that $x = \{u, v\}$. From here on we will also use the term *neighborhood* with the following meaning: if u and v are adjacent vertices, then u is said to be a neighbor of v and, conversely, v is a neighbor of u. Moreover, we call two edges *incident* if they share a common vertex. In addition, if two distinct edges $x, y \in X$ are incident, they are said to be *adjacent edges*. Analogously, we say that an edge is incident to a vertex if the vertex is an element of the edge.

Notice that there exists an easy way to graphically represent graphs: we will draw the vertices of a graph as points and the edges as arcs or lines joining the points.

Example. Let us consider the following graph:



We have the set of vertices $V = \{v_1, v_2, v_3, v_4\}$ and the one of edges

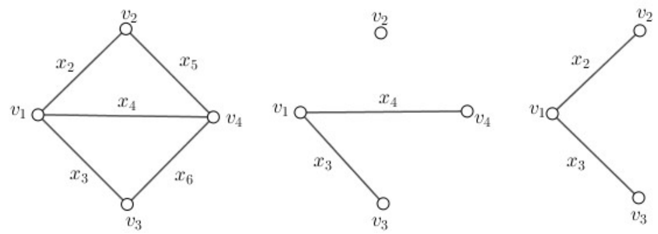
$X = \{x_2, x_3, x_4, x_5, x_6\}$.

We have that v_1 and v_2 are adjacent vertices since there exists an edge $x_2 = \{v_1, v_2\}$; however, v_2 and v_3 are not adjacent because an edge which joins the two vertices does not exist. Moreover, x_2 and x_5 are adjacent edges as they are incident to the common vertex v_2 while x_3 and x_5 are not.

Notice that the definition of graph does not allow *loops* which are edges joining a vertex to itself.

2.2 DEFINITION. A *subgraph* H of G is a graph whose vertices and edges are vertices and edges of G, that is $X(H) \subseteq X(G)$ and $V(H) \subseteq V(G)$.

For example, in the figure below three graphs are shown; let us call the first one G and the second one H. The graph H is a subgraph of G (we may also say that G is a *supergraph* of H). Moreover H is a *spanning subgraph* of G since it contains all the vertices of G i.e. $V(G) = V(H)$. Finally, let us define the subset $S = \{v_1, v_2, v_3\}$ of $V(G)$: the *induced subgraph* $\langle S \rangle$ of G is the maximal subgraph of G with vertex set S (third graph in the figure). This means that, once the set S is defined, the maximal subgraph of G is the subgraph whose edges are all the edges of G that join the elements of S.



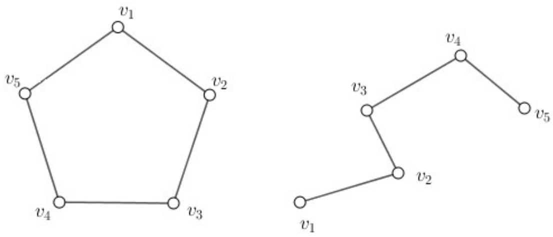
In addition, we may consider the subgraph $G - v_i$ of G with $v_i \in V$ consisting of all the vertices of G except v_i, and all the edges of G that do not contain the vertex v_i. Such operation is called the *removal* of v_i from the graph G. The induced graph $\langle S \rangle$, in the figure above, is the example of the removal of the vertex v_4 from G, i.e. it represents $G - v_4$.

Furthermore, the removal of an edge $x_i \in X$ from G yields a subgraph $G - x_i$ whose edges are those of X except x_i and whose vertices remain the same. Thus $G - v_i$ and $G - x_i$ are the maximal subgraphs of G not containing v_i and x_i, respectively. Finally, it is possible to remove more than a single vertex or edge: the definition of this operation on a graph is defined by the removal of a single element in succession. As we can see in the figure above, the subgraph H of G (second picture) is the subgraph $G - \{x_2, x_5, x_6\}$.

2.3 DEFINITION. A *walk* on a graph G is an alternating sequence of vertices and edges $v_0, x_1, v_1, ..., x_n, v_n$, beginning and ending with vertices, in which each edge is incident with the two vertices immediately preceding and following it.

We say that the walk is *closed* if $v_0 = v_n$, and it is *open* otherwise. The walk is said to be a *trail* if all the edges are distinct and that it is a *path* if all the vertices (and hence all the edges) are distinct. Finally a *cycle* is a closed walk whose vertices $v_1, \dots, v_n$ are distinct with $n \geq 3$ (it still holds $v_0 = v_n$).

We denote by C_n the graph that is an n vertex cycle and by P_n the graph consisting of a path with n vertices. In the figure below C_5 and P_5 are represented, respectively:



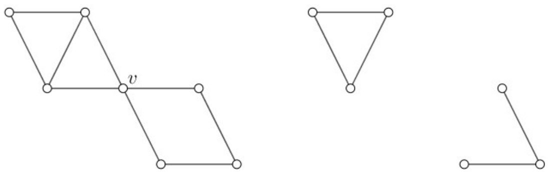
2.4 DEFINITION. A graph is *connected* if every pair of vertices can be connected by a path.

We say that a subgraph H of G is a (*connected*) *component* if it is a maximal connected subgraph of G.

Some connected graphs can be disconnected by the removal of a vertex or an edge. Let us define such elements:

2.5 DEFINITION. A *cutpoint* of a graph G is a vertex whose removal increases the number of components of G. Moreover, a *bridge* is an edge whose removal increases the number of components of G.

By this definition, if v is a cutpoint of a graph G, we have that $G - v$ is a disconnected graph. Analogously, it happens for a bridge x so that the graph $G - x$ would be disconnected. In addition, we define a *non-separable* graph as a non-trivial[1] connected graph, having no cutpoints. Finally a *block* of a graph G is a maximal non-separable subgraph of G. In the figure below the removal of the cutpoint v is shown.

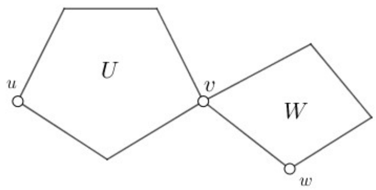


Let us state a proposition that will be useful later:

[1]We call a graph *trivial* if it is a (1,0) graph.

2.6 PROPOSITION. *Let v be a vertex of a connected graph G. We have that v is a cutpoint of G if and only if there exists a partition of the set $V(G) - \{v\}$ into subsets U and W, such that for any vertices $u \in U$ and $w \in W$, the vertex v is on every path joining u and w.*

Proof. Let us prove the first implication: let us assume that v is a cutpoint of the graph G, hence $G - v$ is a disconnected graph with at least two components. We may identify two sets (with empty intersection) of the vertices: the first one is given by a component of the disconnected graph $G - v$, while the second is composed by all the others components of $G - v$ (at least one). These two sets of $G - v$ can be seen as a partition of the set $V(G) - \{v\}$ and we call them U and W, respectively. Consider any vertices $u \in U$ and $w \in W$: they lie in different components of $G - v$ and therefore every path in G joining u and w must contain the vertex v.



On the other hand, let us prove the opposite implication. Let us assume that there exists a partition of the set $V(G) - \{v\}$ into subsets U and W, such that for any vertices $u \in U$ and $w \in W$, the vertex v is on every path joining u and w. In particular these two vertices, u and w, are in the graph G and they are distinct from v. Thus, by assumption, v is on every path in G joining u and w, which implies that there cannot exist a path in $G - v$ joining these two vertices. This means that the graph $G - v$ is disconnected and therefore v is a cutpoint in G. □

Let us state some definitions and results on connectivity.

2.7 DEFINITION. A graph G is k-connected if one can remove k vertices from G in such a way that G becomes disconnected, but there is no way to remove less than k vertices from G in order to disconnect it.

In other words, a graph G is said to be k-connected if k is the lowest number of vertices that we need to remove from the graph in order to disconnect it.

The *lenght* of a walk is the number of occurences of edges in it so that if we consider the walk $v_0, x_1, v_1, ..., x_n, v_n$ we have that its lenght is n.

We want to define the *distance* between two vertices u, v in a graph G.

2.8 DEFINITION. The *distance* $d(u, v)$ is the lenght of a shortest path joining them if any; otherwise we set $d(u, v) = \infty$.

Notice that when we consider a connected graph the distance is a metric, i.e. the following properties hold:

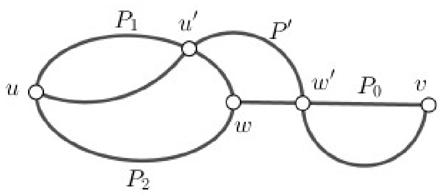
1. $d(u,v) \geq 0$ for all vertices u, v in G;

2. $d(u,v) = d(v,u)$ for all vertices u, v in G.

3. $d(u,v) + d(v,w) \geq d(u,w)$ for all vertices u, v, w in G

2.9 THEOREM. *Let G be a connected graph with at least three vertices. If G is a block then every two vertices of G lie on a common cycle.*

Proof. Let u, v be two distinct vertices of G, and let U be the set of vertices different from u which lie on a cycle containing u. By assumption G is a block with at least three vertices, thus it has no cutpoint nor bridges as for a connected graph with at least three vertices having a bridge implies having a cutpoint. Moreover, U is non-empty since every vertex adjacent to u is in U and G is connected. Suppose, by contradiction, that v is not in U and let w be a vertex in U such that the distance $d(w,v)$ is minimum. Let P_0 be a shortest path from w to v and let P_1 and P_2 be two different paths of a cycle joining u and w. Since w is not a cutpoint, there exists a path P', joining u and v, that does not contain w. Let w' be the vertex nearest to u such that it is both in P_0 and P'; moreover, let u' be the last vertex of the subpath of P' joining u and w' that belongs to P_1 or P_2 as well. Without loss of generality, we may assume that u' is in P_1.

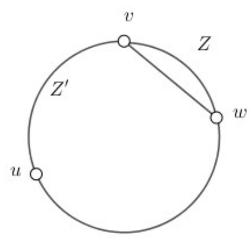


Let Q_1 be the path joining u and w' consisting of the subpath of P_1 joining u and u' and the subpath of P' joining u' and w'. Analogously, let Q_2 be the path joining u and w' consisting of P_2 and the subpath of P_0 joining w and w'. It follows that Q_1 and Q_2 are two different disjoint paths joining u and w' hence they form a cycle, so w' is in U. Since w' is on a shortest path joining w and v, it holds that $d(w',v) < d(w,v)$ and this contradicts the choice of w, so that the thesis is proved. □

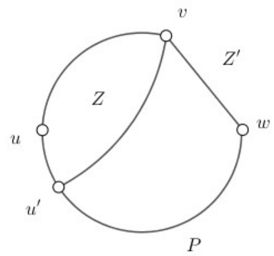
2.10 PROPOSITION. *Let G be a connected graph with at least three vertices. If every two vertices of G lie on a common cycle, then every vertex and edge of G lie on a common cycle.*

Proof. Let u be a vertex and $\{v, w\}$ an edge of G; moreover let Z be a cycle of G containing both u and v. Let us build a cycle Z' which contains both u and $\{v, w\}$.

If w is on Z, then Z' consists of $\{v, w\}$ along with the path joining v and w of Z, that contains u.



On the other hand, if w does not belong to the cycle Z, however, there must be a path, say P, which joins w and u and not v, since otherwise, by Proposition 2.6, v would be a cutpoint and that would lead to a contradiction. Let u' the first vertex in the path P that also belongs to the cycle Z. Z' consists of the edge $\{v, w\}$, together with the path joining w and u' (it exists since it is a subpath of P) and the path of Z that joins u' and v, which contains the vertex u.



□

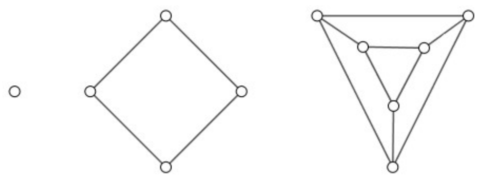
2.11 DEFINITION. Let G be a graph. The *degree* of a vertex v_i in $V(G)$ is the number of edges which contain v_i and it is denoted by d_i or *deg* v_i.

We call $\delta(G)$ the *minimum degree* among the vertices of a graph G and $\Delta(G)$ the largest of such number.

Moreover, if G is a graph with $|V| = p$ and $|X| = q$, i.e. it is a (p, q) graph we have that $0 \leq deg\ v_i \leq p - 1$ for every $v_i \in V$. We call *regular of degree* r a graph whose vertices have all the same degree r. Notice that if a graph G is regular then $\delta(G) = \Delta(G) = r$, and conversely, if $\delta(G) = \Delta(G) = r$ holds, then the graph is regular of degree r.

Moreover, a *complete graph* is a graph which has every pair of its p vertices adjacent and it is called K_p. In particular K_p is a regular graph of degree $p - 1$.

In the figure below we can see a regular graph of degree 0, a regular graph of degree 2 and a regular graph of degree 3 (also called *cubic*), respectively.



Let us state a paramount result:

2.12 THEOREM. *Let G be a (p,q) graph. The sum of the degrees of the vertices of G is twice the number of edges, i.e. it holds*

$$\sum_{v_i \in V(G)} deg\ v_i = 2q.$$

Proof. Let us call $I(v_i)$ the set of all the edges that are incident to the vertex $v_i \in V(G)$.

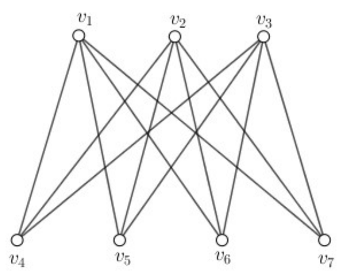
$$\sum_{v_i \in V(G)} deg\ v_i = \sum_{v_i \in V(G)} \sum_{q \in I(v_i)} 1 = \sum_{q \in X(G)} \sum_{\substack{v_i \in V(G) \\ s.t.\ q \in I(v_i)}} 1 = \sum_{q \in X(G)} 2 = 2q$$

□

2.13 DEFINITION. A *bigraph (or bipartite graph) G* is a graph whose set of vertices V can be partitioned into two subsets V_1 and V_2 such that every edge of the graph joins vertices of V_1 with vertices of V_2.

If G contains all the edges joining V_1 with V_2 the graph is a *complete bigraph.* If we set $|V_1| = m$ and $|V_2| = n$ we write $G = K_{m,n}$.

Example. The following graph is $K_{3,4}$ with $V_1 = \{v_1, v_2, v_3\}$ and $V_2 = \{v_4, v_5, v_6, v_7\}$.



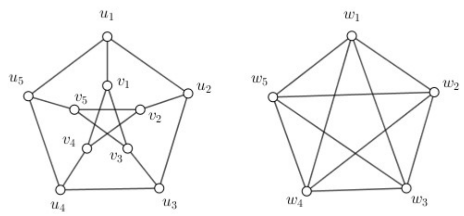
2.14 DEFINITION. Let G be a graph and $u, v \in V(G)$ be any two adjacent vertices. An *elementary contraction* of G is obtained by identifying the two adjacent vertices u and v; in other words it is the removing of the two vertices u and v together with the addition of a new vertex, say w, adjacent to those vertices to which u or v was adjacent.

Moreover we say that a graph G is *contractible* to a graph H if H can be obtain from G by a sequence of elementary contractions.

Let G be a graph and $x = \{u, v\}$ one of its edges; then we will denote with $G \cdot x$ the contracted graph given by removing x and identifying u and v.

It is important to notice that a graph G, that is contracted to a new graph G', is not necessary homeomorphic[2] to it. The following example shows that *"Petersen graph"* is contractible to K_5 but, since every vertex has degree three, it does not contain any subgraph homeomorphic to K_5.

Example. Petersen graph is contracted to K_5 by removing the vertices u_i and v_i and replacing them with w_i for $i = 1, ..., 5$. The figure below may clarify the situation.



[2]The definition of graphs homeomorphism is given in Chapter 3.

Chapter 3

Planarity

In this chapter we will introduce the notions of planar graph and planar map. Next, we will state Euler's formula and some results which follow from it. Finally, we will deal with the proof of Kuratowski's Theorem, establishing a criterion for planarity.

3.1 Plane Graphs, Plane Maps and Convex Polyhedrons

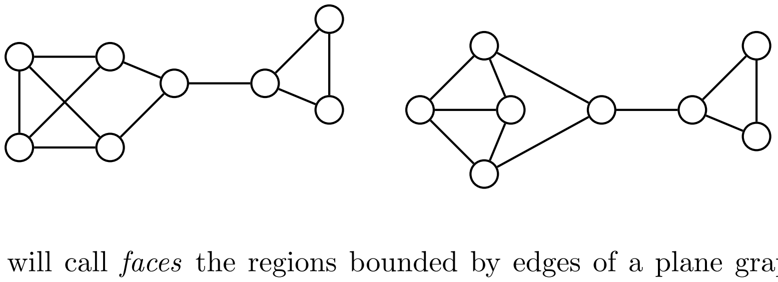
3.1 DEFINITION. A graph G is said to be embedded in a surface S if, when it is drawn on S, no pair of edges intersect.

3.2 DEFINITION. A graph G is *planar* if it can be embedded in the plane.

We call a graph that has been embedded in the plane a *plane graph*.

Remark: If a graph G can be embedded in the plane then, by a stereographic projection, it can be embedded on the sphere $\mathbb{S}^2$ as well.

The figure below clarifies the difference between planar and plane graphs. The first graph is a planar graph as it can be embedded in the plane (see the second picture) but a pair of its edges still intersects while the second graph represents one of its embedding in the plane.



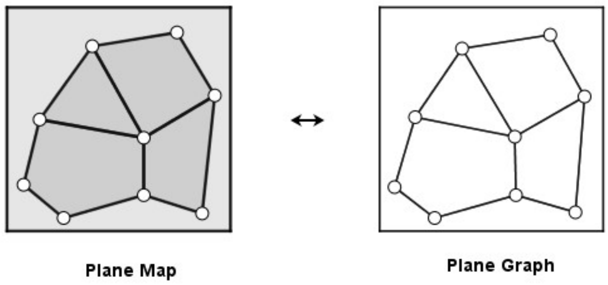
We will call *faces* the regions bounded by edges of a plane graph and *exterior face* the unbounded region.

3.3 DEFINITION**.** A maximal plane graph is a graph to which no edge can be added without losing planarity.

This implies that whenever an edge is added to a maximal plane graph it will intersect at least one other edge of the graph.

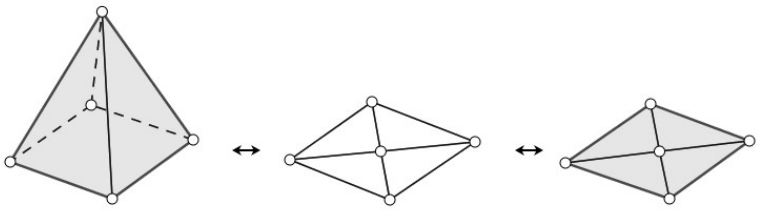
3.4 DEFINITION**.** A *plane map* is a connected plane graph together with all its faces.

This definition establishes a correspondence between plane maps and plane graphs; whenever a (p, q) graph, say G, is given, we can identify (with the same notation) the (p, q) plane map, which is the map associated to the graph G.



Let us see some results on plane maps.

It is possible to associate a plane graph to every polyhedron, in such a way that vertices and edges of the polyhedron corresponds to those of the graph. Moreover, if we consider all the faces of a convex polyhedron, a plane map can be associated to it. This means that there is a correspondence between vertices, edges and faces of convex polyhedrons and those of plane maps. The figure below may clarify such a relation.



The following theorem explains the relationship between vertices, edges and faces of a polyhedron:

3.5 THEOREM (Euler's Formula)**.** *[8] For any convex polyhedron with V vertices, E edges and F faces the following holds*

$$V - E + F = 2 \tag{3.1.1}$$

As we have already seen above, there is a correspondence between convex polyhedrons and plane maps thus this theorem can be restated in graph theoretic terms. We may indicate the number of vertices of the map with p, the number of edges with q and the number of faces (including the external one) with r. The following table highlights the correspondence between convex polyhedrons and plane maps.

convex polyhedrons	plane maps
V	p
E	q
F	r

Thanks to this relation (3.1.1) turns to

$$p - q + r = 2. \tag{3.1.2}$$

The following result follows from (3.1.2):

3.6 Corollary. *If G is a (p,q) plane map in which every face is a n-cycle then*

$$q = \frac{n(p-2)}{n-2} \tag{3.1.3}$$

Proof. By hypothesis we have that every face of the plane map is a n-cycle and that each edge of G is in two faces. Thus $nr = 2q$ i.e. $r = \frac{2q}{n}$ which replaced into (3.1.2) gives (3.1.3). □

The next result is about maximal plane graph:

3.7 Corollary. *If G is a (p,q) maximal plane graph, then every face of the associated plane map is a triangle and $q = 3p - 6$. Moreover, if every face of the plane map associated to G is a 4-cycle, then $q = 2p - 4$.*

Proof. If the G is a maximal plane graph, then every face is a triangle. In particular, it suffices to replace n with 3 or 4 into (3.1.3), respectively and the thesis is proved. □

Thanks to this result it holds that the maximal number of edges in a plane graph occurs when all its faces are triangles, thus we obtain the following:

3.8 Corollary. *If G is a plane (p, q) graph with $p \geq 3$, then*

$$q \leq 3p - 6. \tag{3.1.4}$$

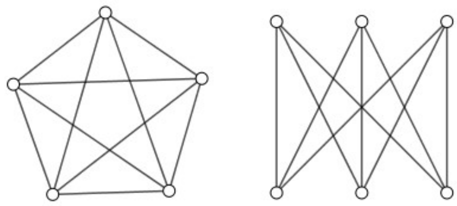
Moreover, if G has no triangle then

$$q \leq 2p - 4. \tag{3.1.5}$$

Proof. This follows from the previous Corollary since a maximal plane graph is a graph with the largest number of edges thus we get the two inequalities. □

3.9 Corollary. *The graphs K_5 and $K_{3,3}$ are non-planar.*

Proof. It is enough to notice that the graph K_5 is a (5,10) graph, therefore it does not satisfy (3.1.4). Furthermore, according to (3.1.5), $K_{3,3}$ is a non-planar graph as well. □



3.10 THEOREM. *Every planar graph G has a vertex of degree not exceeding 5.*

Proof. Let us prove this by contradiction. Suppose G is a (p, q) planar graph in which every vertex has degree at least 6, i.e. $deg\ v_i \geq 6$ for every $v_i \in V(G)$. By Theorem 2.12 we have that $\sum_{v_i \in V(G)} deg\ v_i = 2q$. Moreover, according to (3.1.4) it holds that $q \leq 3p - 6$, therefore, we get the following chain of inequalities

$$6p \leq \sum_{v_i \in V(G)} deg\ v_i = 2q \leq 6p - 12$$

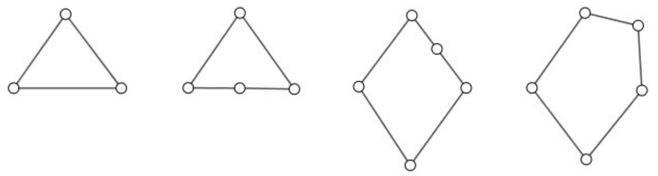
that is a contradiction. □

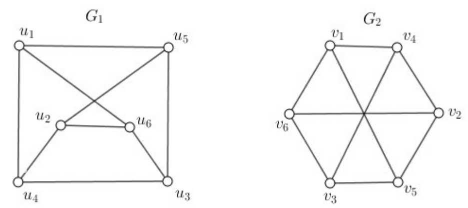
We need to keep in mind this result as it will be taken up later.

3.11 DEFINITION. Two graphs are *homeomorphic* if both can be obtained from the same graph by a sequence of subdivisions of lines.

Moreover, two graphs are *isomorphic* if there exists a bijective function between their vertices sets, which preserves adjacency, and whose inverse preserves adjacency as well. In particular, as we can see in the figure below, any two cycles are homeomorphic but cycles of different length are not isomorphic.

Furthermore, the graphs G_1 and G_2 in the figure below are isomorphic under the correspondence $u_i \longleftrightarrow v_i$.





3.2 Kuratowski's Theorem

In this section we give a proof of *Kuratowsi's Theorem* [10], whose statement is as follows:

3.12 THEOREM. *A graph is planar if and only if it has no subgraph homeomorphic to K_5 or $K_{3,3}$*

Proof. The first part of the theorem follows from Corollary 3.9; K_5 and $K_{3,3}$ are non-planar graphs, thus if a graph contains a subgraph homeomorphic to either of these it is also non-planar.

The converse requests a bit more effort: in order to complete the proof we will state and prove some results.

Before doing this, we assume that there exists a graph that is a non-planar graph with no subgraph homeomorphic to either K_5 or $K_{3,3}$. Let $\overline{G}$ be such a graph which has the minimum number of edges.

Let us state our first proposition:

3.13 PROPOSITION. *Any minimally non-planar graph is at least 2-connected.*

Proof. Let us proceed by contradiction; assume that there exists a minimally non-planar graph, say G, that is strictly less then 2-connected (i.e. it is $\leq$ 1-connected), hence there is at least a vertex, say v that we can delete from G, which disconnects G. Let us call such resulting components $G'_1, \ldots, G'_n$ and let $G_1 = \langle V(G'_1) \cup \{v\} \rangle, \ldots, G_n = \langle V(G'_n) \cup \{v\} \rangle$.

Notice that $G_1, \ldots, G_n$ are all planar graphs since all these components are subgraphs of G. We have that for any planar graph G_i and any face f_i of G_i, it is possible to draw G_i in the plane in such a way that f_i is the outside face of G_i. To see this, it is enough to consider a planar embedding of G_i on the unit sphere such that the face f_i contains the "north pole" (the point (0,0,1)) of the sphere. By a stereographic projection, we obtain a planar embedding in $\mathbb{R}^2$, such that the face f is the outside face.

Let f_i be the face that contains v on each G_i. We may place the vertex v at the origin of the plane in which we are embedding each G_i and by resizing all G_i's and placing all of them together, each adjacent to the other, in such a way that each of them is bounded in a less then $\frac{2\pi}{n}$ degrees region. The resulting graph is a planar embedding of G, despite the fact that G was non-planar. Thus, we get a contradiction that provides that our assumption that G was less then 1-connected was false.

□

We now want to show that $\overline{G}$ must be a block:

3.14 PROPOSITION. *If G is a non-planar graph which contains no subgraphs homeomorphic to K_5 or $K_{3,3}$ and it has the minimum number of edges among all of these graphs, then G is at least 3-connected.*

Proof. Let us prove this proposition by contradiction.

Consider any such graph G; by Proposition 3.13 it must be at least 2-connected.

The goal now is to show that if we suppose that G is 2-connected we get a contradiction, so that it must be at least 3-connected. If G is 2-connected, then there must be two vertices, say v, w, such that, if they are removed the graph will be disconnected into different components. We call such components $G'_1, \dots, G'_n$ and we define $G_1 = \langle V(G'_1) \cup \{v, w, x\} \rangle, \dots, G_n = \langle V(G'_n) \cup \{v, w, x\} \rangle$, where x is the edge joining v and w, even if it is not in $X(G)$. We now aim to show that there exists an index i such that G_i is non-planar. Suppose not, i.e. each G_i is a planar graph, therefore it can be drawn in the plane in a way such that the face containing x is the outer face of the graph (in particular both v and w are on the outer face as well). We now want to create a new graph in the following fashion: starting from G_1, we draw it in the plane. The following step is to draw the graph G_2 in such a way that the edge x of G_2 is overlapped to that of G_1 (the edge x belongs to both G_1 and G_2 by construction) and draw the rest of G_2 in such a way that it entirely "surrounds" G_1. We may do the same for every G_i to create a sequence of nested planar graphs. The result is a planar graph again, but this contradicts the fact that G is non-planar and therefore there is at least one of the G_i's that is non-planar. Let $G_{\tilde{i}}$ be such a graph: by assumption G is a non-planar graph which does not contain any subgraph isomorphic to K_5 or $K_{3,3}$ with the minimum number of edges, thus $G_{\tilde{i}}$ must contain a subgraph homeomorphic to either K_5 or $K_{3,3}$.

We now have these four cases:

1. The edge x is not an edge of the subgraph of $G_{\tilde{i}}$ homeomorphic to either K_5 or $K_{3,3}$ but x is in G. This is not possible as G would contain such subgraph homeomorphic to either K_5 or $K_{3,3}$ since removing x from G would not affect the rest of the subgraph $G_{\tilde{i}}$.

2. The edge x is neither an edge of the subgraph of $G_{\tilde{i}}$ homeomorphic to either K_5 or $K_{3,3}$, nor in G. This case is similar to the previous one.

3. The edge x is an edge of both the subgraph homeomorphic to either K_5 or $K_{3,3}$ and G. In this case we get a contradiction since G would contain the subgraph of $G_{\tilde{i}}$ homeomorphic to either K_5 or $K_{3,3}$.

4. The edge x belongs to the subgraph homeomorphic to K_5 or $K_{3,3}$ and it is not an edge of G. We may consider any other component G_j and a vertex u of G'_j. Since G is 2-connected, there must exist a path joining the vertex u with both v and w in G_j, as otherwise we may disconnect u from $G_{\tilde{i}}$ simply by removing one between v or w. Thus, there must be a path joining v and w in G_j so that if we replace x with such path in G_j we still have a graph homeomorphic to either K_5 or $K_{3,3}$ (we only split an edge). This leads to a contradiction as G would contain a subgraph homeomorphic to either K_5 or $K_{3,3}$.

Since all the possible cases lead to a contradiction the proof is completed. □

Let us now have a look at a proposition that provides that 3-connectedness still holds if a graph is contracted in some way.

3.15 PROPOSITION. *If G is a graph with at least 5 vertices that is at least 3-connected, then there is some edge $x \in X(G)$ such that $G \cdot x$ is at least 3-connected.*

Proof. Let us assume, by contradiction, that for every edge $x = \{u, v\} \in X(G)$ the graph $G \cdot x$ is less than 3-connected. Notice that one of these 2-disconnecting vertices of $G \cdot x$ must be the new vertex, say v_x, created by the contraction of the graph G, since otherwise the whole graph G may be disconnected by the removing of just two vertices despite the fact that G is at least 3-connected. Before starting, we may convey that we will call a vertex w a *"mate"* of $x = \{u, v\}$ whenever $\{v_x, w\}$ is a 2-disconnecting set in $G \cdot x$. The assumption we just made provides that every edge $\{u, v\}$ has a mate w, therefore the removing the two vertices of any edge $x = \{u, v\}$ together with its mate w will disconnect the graph G, since removing v_x together with w disconnects the contracted graph $G \cdot x$.

Let us call $G_1, \ldots, G_n$ the connected components of G obtained by removing three vertices u, v, w.

Among all the possible edges $\{u, v\}$ together with their mate w whose removing disconnects the graph G, take the one that give rise to the "largest" component of G. Without loss of generality let us call such edge $\{\overline{u}, \overline{v}\}$, the mate $\overline{w}$ and the largest component G_1.

Moreover, there must exist an edge from $\overline{w}$ to another vertex, say z, that belongs to a connected component G_i with $i \neq 1$, as otherwise, the graph G would be disconnected by the removing of the two vertices $\overline{u}, \overline{v}$ only, despite the fact that G is at least 3-connected. Without loss of generality we call G_2 the component which contains z.

Let r be a mate of the edge $\{\overline{w}, z\}$ so that the removing of the vertices $\{\overline{w}, z, r\}$ disconnects the graph $G - \{\overline{w}, z, r\}$ into components. Let us consider the following cases.

1. If $r = \overline{u}$ then the subgraph $\langle V(G_1) \cup \{\overline{v}\}\rangle$ of G is a connected subgraph of the graph G which is larger than G_1. This is a contradiction since we assumed that G_1 was the largest.

2. If $r = \overline{v}$ then the subgraph $\langle V(G_1) \cup \{\overline{u}\}\rangle$ of G is a connected subgraph of the graph G which is larger than G_1. This is a contradiction since we assumed that G_1 was the largest.

3. If r belongs to any of the connected components G_i with $i \neq 1$, then G would be disconnected into different components, one of which would contain G_1 together with the two vertices $\overline{u}, \overline{v}$, however that is not possible as G_1 is the largest component.

4. If r belongs to G_1, then the removing of $\{z, r, \overline{w}\}$ does not disconnect the graph G. Let us see why: Let a be any vertex in any component G_j, thus there must exist two vertex-disjoint paths joining a with both $\overline{u}$ and $\overline{v}$, as otherwise, it would be enough deleting r along with at most another vertex on this path in order to disconnect a from other vertices in our graph. This implies that, once that a vertex has been removed from any G_j, there still must be a path joining any vertex $a \in V(G_j)$ to either $\overline{u}$ or $\overline{v}$. However, this implies that the graph $G - \{z, r, \overline{w}\}$ is still connected since it is possible to find a path that joins any vertex a (which belongs to any component G_j) to either $\overline{u}$ or $\overline{v}$ and any other vertex, say b, in any other component G_k.

Since both the analysed cases lead to a contradiction, there must be some edge $x \in X(G)$ such that $G \cdot x$ is still 3-connected.

$\square$

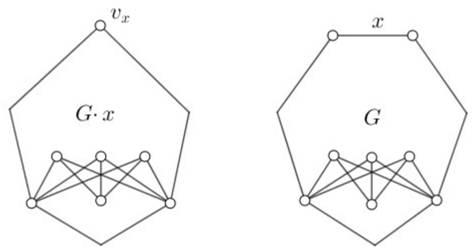
The following proposition provides that whenever a graph G, which does not contain any subgraph homeomorphic to K_5 or $K_{3,3}$, is contracted then neither $G \cdot x$ does.

3.16 PROPOSITION. *If G is a graph which does not have any subgraph homeomorphic to either K_5 or $K_{3,3}$, then neither its contraction $G \cdot x$ does for every edges $x \in X(G)$.*

Proof. Let us prove this claim by its contrapositive statement, i.e. if $G \cdot x$ has a subgraph homeomorphic to either K_5 or $K_{3,3}$, then so G does.

Consider any graph G and any edge $x \in X(G)$, such that $G \cdot x$ has a subgraph homeomorphic to either K_5 or $K_{3,3}$: let us call it K. Let us denote with v_x the vertex that results after the contraction of x; we have the following cases:

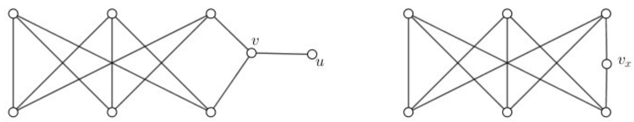
1. if v_x is not a vertex of the subgraph of $G \cdot x$ that is homeomorphic to to either K_5 or $K_{3,3}$, then it easily follows that G contains the same subgraph.



2. Let us suppose that v_x is a vertex of the subgraph of $G \cdot x$ that is homeomorphic to either K_5 or $K_{3,3}$.

   First, let us notice that each vertex of either K_5 or $K_{3,3}$ has degree 4 or 3 respectively. This means that a graph that is homeomorphic to one of those must contain vertices that have degree at least 2 and at most 4. Let us examine each of these three cases:

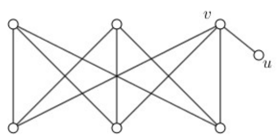
   - Consider the case in which the vertex v_x has degree 2: there are two possibilities up to switching the role of the two vertices u and v.
     The first one occurs if both edges of K have the vertex v as their endpoint. The graph, when expandend, would still be homeomorphic to either K_5 or $K_{3,3}$.

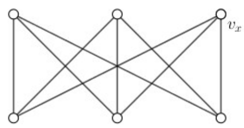


     The second (and last) case occurs when a edge of K has v as its endpoint, and the other has u instead. In this case also, when we expand back K into the original graph G, we would still have a subgraph that is still homeomorphic to either K_5 or $K_{3,3}$.

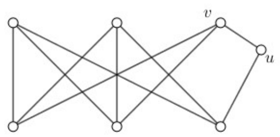
- Let us consider the case in which v_x has degree 3: there are two possibilities.

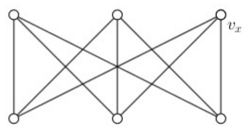
  On the one hand suppose that all the three edges of K have the same endpoint v. In this case, expanding back K to the original graph, it still is a subgraph that is homeomorphic to either K_5 or $K_{3,3}$.





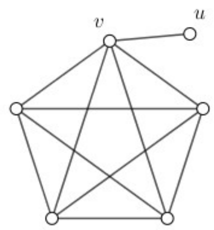
  On the other hand, suppose that two edges of K have v as endpoint and the other has u. In this case also we still would have a subgraph of G that is homeomorphic to either K_5 or $K_{3,3}$ once K is expanded back.

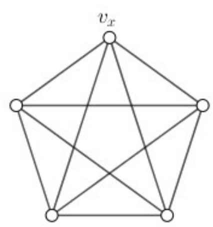




- Let us finally consider the case in which $deg\ v_x = 4$. Let us face all the (three) possibilities which may occur.

  In the first one we may have all the four edges of K have the same endpoint vertex v.

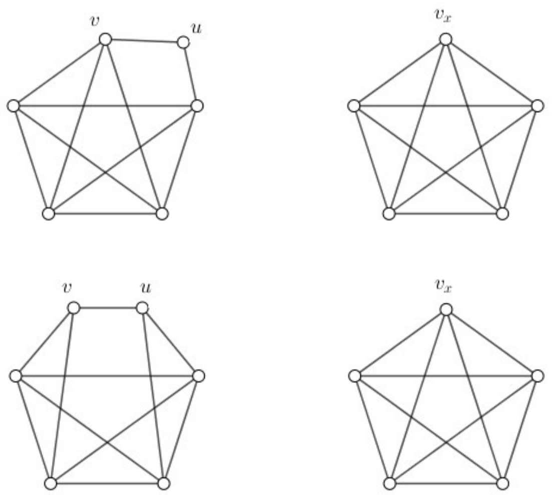




  The second case occurs when we suppose that three edges of K have v as their endpoint, and only one has u.

  In either case, the graph subgraph K is still a subgraph that is homeomorphic to either K_5 or $K_{3,3}$ when expanded back out to G.

  Last, we may consider the case in which two edges have v as their endpoint and the two have u.



This configuration is not as simple as the above ones. Since we are considering that *deg* $v_x = 4$, K has to be a graph that is homeomorphic to K_5. Let w_i with $i = 1, \dots, 4$ be the four vertices that along with v_x form the subgraph K_5 of K. In our assumptions, without loss of generality, we may consider that, when we expand back K out to the original graph there are two edges, say x_1, x_2, such that $x_1 = \{w_1, v\}$ and $x_2 = \{w_2, v\}$, and another pair of edges, x_3, x_4 such that $x_3 = \{w_3, u\}$ and $x_4 = \{w_4, u\}$. If we consider the two sets of points $S_1 = \{v, w_3, w_4\}$ and $S_2 = \{u, w_1, w_2\}$, we have that they form a subgraph that is homeomorphic to $K_{3,3}$[1].

This proves our claim: if G does not have any subgraph homeomorphic to either K_5 or $K_{3,3}$, then neither does $G \cdot x$ $\square$

So far, by the propositions above[2], we proved that given any at least 3-connected graph with no subgraph homeomorphic to either to K_5 or $K_{3,3}$, there exists a contraction whereby these two properties are preserved.

Let us prove another important result.

3.17 PROPOSITION. *If G is a graph at least 3-connected and it does not contain any subgraph homeomorphic to K_5 or $K_{3,3}$, then G is planar.*

Proof. Let us prove this by induction on the number of vertices of G.

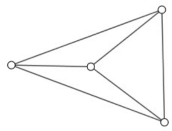
Before starting, let us notice that if G admits a convex embedding in the plane in which no three vertices lie on a line i.e. they are not aligned, then G is planar. A convex embedding in the plane is a planar embedding in

[1] To see it we may toss the edges $\{w_1, w_2\}$ and $\{w_3, w_4\}$ of G.

[2] Proposition 3.14, Proposition 3.15, Proposition 3.16.

which all the edges are straight lines and all of the internal faces are convex polygons. In order to prove the claim, we will prove that G admits a convex embedding, in fact.

Base step. If the graph G has strictly less than five vertices, since it is at least 3-connected by hypothesis then G must be the complete graph K_4[3], therefore it admits a convex embedding in the plane with no three vertices on a line.

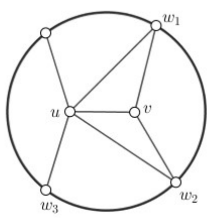


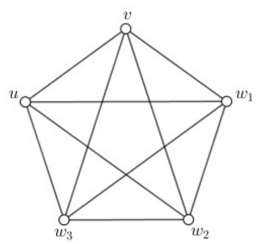
Inductive step. By assumption, we have that G is any at least 3-connected graph with no subgraph homeomorphic to either K_5 or $K_{3,3}$. Thanks to the propositions above (3.14, 3.15, 3.16), there exists an edge $x = \{v, u\}$ of G, such that it holds that the contracted graph $G \cdot x$ is still at least 3-connected and it has no subgraph homemomorphic to either K_5 or $K_{3,3}$, therefore by the inductive hypothesis [4], the graph $G \cdot x$ admits a convex embedding in the plane with no three vertices on a line.

Consider such $G \cdot x$'s embedding: let v_x be the vertex that takes the place of the edge x in $G \cdot x$ and let us remove it. It holds that the face that contains where v_x used to be is a cycle as otherwise it would be enough to remove at most one other vertex to disconnect G, which implies that $G \cdot x$ would be at most 2-connected. Since $G \cdot x$ admits a planar embedding, all the neighbors (adjacent vertices) of v_x must be on this cycle, therefore the very same holds for u and v's neighbors.

Let v_i for $i = 1, \ldots, n$ and u_i for $i = 1, \ldots, m$ be the neighbors of v and u respectively, labeled in cyclic order. Let us examine u's neighbors setting.

1. Consider the case in which u has at least three neighbors in common with v, say w_1, w_2, w_3. That leads to a contradiction, since there would be a subgraph of G homeomorphic to K_5.

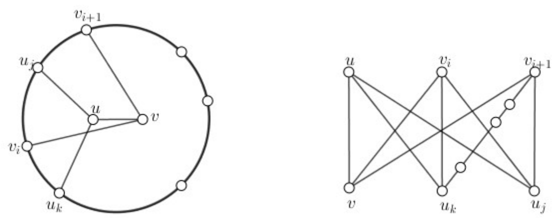




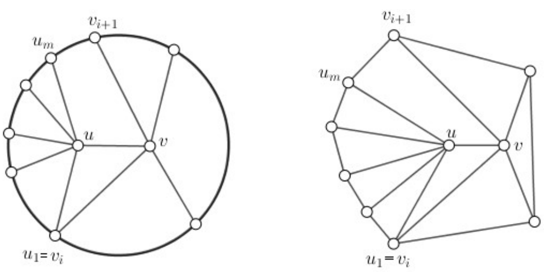
[3] K_4 is the only 3-connected graph with less than 5 vertices.

[4] since the graph $G \cdot x$ has a vertex less that G

2. Consider the case in which at least two neighbors of u alternate with two of v, i.e. there is the sequence u_k, v_i, u_j, v_{i+1} for some $k, j \in \{1, \dots, m\}$ and $j \neq k$ and $i = \{1, \dots, n-1\}$. That leads to a contradiction, since there would be a subgraph of G homeomorphic to $K_{3,3}$.



3. Consider the case in which all the neighbors of u are between two consecutive neighbors of v (say v_i, v_{i+1})[5]. In this case it is possible to create a convex embedding of G by moving u into the triangle formed by the vertex v and its two consecutive neighbors v_i, v_{i+1} which contain all the neighbors of u.



This argument proves the following claim "if G is at least 3-connected graph and it does not contain any subgraph homeomorphic to K_5 or $K_{3,3}$, then G admits a convex embedding with no three points that lie on a line" that implies that G is planar.

□

In order to conclude the proof it is enough to notice that if a non-planar graph G does not contain any subgraph homemomorphic to either K_5 or $K_{3,3}$ and, without loss of generality, it is minimal with respect to the number of edges, then, by Proposition 3.14, it must be at least 3-connected. At the same time, by Proposition 3.17, it holds that if a graph is at least 3-connected and it does not contain any subgraph homeomorphic to K_5 or $K_{3,3}$, then it is planar. Contradiction.

In particular, if we assume that $\overline{G}$ exists, then we get a contradiction, therefore the theorem is proved. □

[5] Notice that u and v may have at most two common neighbors that are the two consecutive neighbors of v that "bound" all the neighbors of u.

Chapter 4

The Six and the Five Colour Theorems

4.1 Colorability

4.1 DEFINITION. A *coloring* of a graph is an assignment of colors to its vertices so that no two adjacent vertices have the same color.

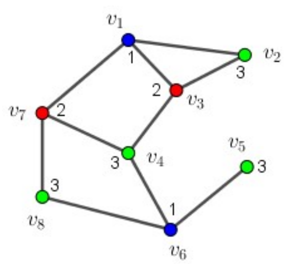
In other words, let us consider the graph G: an n-coloring of G is a mapping

$$\phi \ : \ V(G) \longrightarrow \{1, ..., n\}$$

such that $\phi(u) \neq \phi(v)$ for every $u, v \in V(G)$ that are adjacent.

It is clear the correspondence between the assigned color of a vertex and a number between 1 and n. This convention will be used in the following as it simplifies notations.

Let us consider a coloring of a graph G. A color is assigned to each vertex, therefore, it is possible to consider the *color class*, which is the set of all vertices that have the same color, hence defining a partition of the set $V(G)$. We say that an *n-coloring* of a graph G uses n colors.



In the figure above the following partition is represented

$$V_1 = \{v_1, v_6\}, \quad V_2 = \{v_3, v_7\} \ \ V_3 = \{v_2, v_4, v_5, v_8\}$$

Each color, blue, red and green, is associated to a number 1, 2 and 3, respectively.

4.2 DEFINITION. The *chromatic number* $\chi(G)$ is the minimum n for which a graph G has an n-coloring.

We say that G is n-*colorable* if $\chi(G) \leq n$ and that it is n-*chromatic* if $\chi(G) = n$.

We may notice that if G is a (p, q) graph, then it has both a p-coloring and a $\chi(G)$-coloring therefore G must have a n-coloring for $\chi(G) \leq n \leq p$.

Example. In the figure below two different colorings of the same graph are represented.

There is no known method for determining the chromatic number of an arbitrary graph. However, in some cases it is immediate to determine it, for example the complete graph K_p must have chromatic number $\chi(K_p) = p$, while for the complete bipartite graph $K_{m,n}$ it holds that $\chi(K_{m,n}) = 2$. For all the other cases there is not a general rule, but many results on colorability have been found.

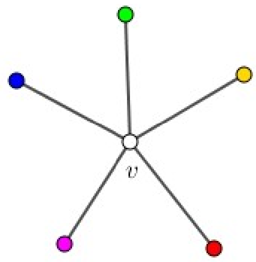
4.2 The Six Color Theorem

In the following section the proof of the Six Color Theorem is given. the Theorem states that every planar graph is 6-colorable i.e. six colors are sufficient to color the vertices of a planar graph so that two adjacent vertices do not have the same color.

4.3 THEOREM. *Let G be a planar graph, then there exists a* 6*-coloring of G.*

Proof. Let G be a planar graph which does not admit a 6-coloring and let G be the minimal with respect to the number of vertices. We may suppose that the number of vertices of G is greater than 6, otherwise the graph is trivially 6-colorable. By Theorem 3.10 there exists a vertex, say v, of G which has degree that does not exceed 5.

By assumption, we have that the planar graph $G - v$ admits a 6-coloring, hence we shall color all the vertices of $G - v$ with six colors. Since v has at most five adjacent vertices in G, these are colored with less then six different colors. Therefore, by assigning to v a color that it has not been used to color its neighbors, we end up with a 6-coloring of G, that contradicts our assumptions.



□

It is immediate to see that the hypothesis of planarity for the graph is necessary. Just think of the nonplanar graph K_8: each vertex is connected to all the other seven vertices, so six colors would not be enough.

4.3 The Five Color Theorem

In this section we present the Five Color Theorem. The focus should be on the idea that underlies the proof of this Theorem, together with the whole technique it is used to prove it.

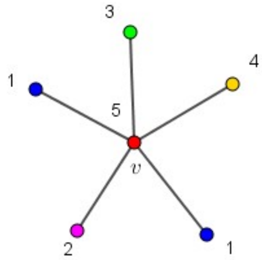
4.4 THEOREM. *Every planar graph is 5-colorable.*

Proof. Let us proceed by induction on the number p of vertices of the graph.

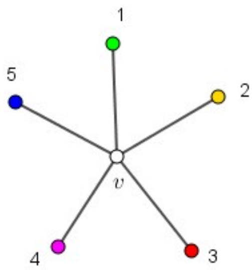
If we consider a graph with $p \leq 5$ vertices the result follows trivially since the graph is p-colorable.

Let us now assume that all planar graphs with $p > 5$ vertices are 5-colorable and let G be a planar graph with $p+1$ vertices. By Theorem 3.10, G contains a vertex, say v, of degree 5 or less therefore, by assumption, the graph $G - v$ is 5-colorable. Consider now a 5-color assignment for the vertices of the graph $G - v$ and let us denote these colors by a number i with $i = 1, ..., 5$. We now face two possibilities:

- If at least one of the 5 colors, say $j \in \{1, \ldots, 5\}$, is not used in the coloring of the vertices adjacent with v, then it suffices to assign the color j to v and so that the graph G is 5-colored. The figure below shows an example of a graph G, which as a vertex v whose neighbors are colored with colors 1, 2, 3 and 4, but none of them is colored with the color 5. This allows to assign color 5 to the vertex v and hence 5-coloring the graph G.

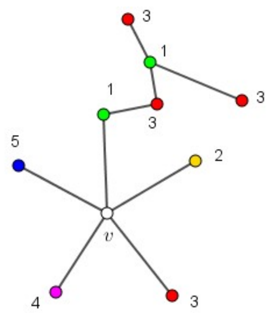


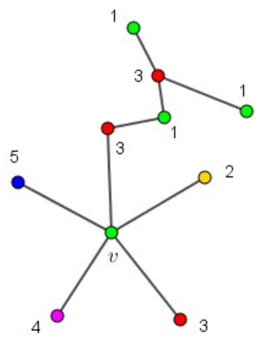
- If v has degree 5 and all the colors are used for its adjacent vertices, we may assume, without loss of generality[1], that the five colors are assigned cyclically about v as shown in the figure below (we call v_i the vertex adjacent with v that is colored with the color i for $i = 1, ..., 5$).



Let $G_{1,3}$ be the subgraph of $G - v$ induced by those vertices colored 1 or 3.[2]

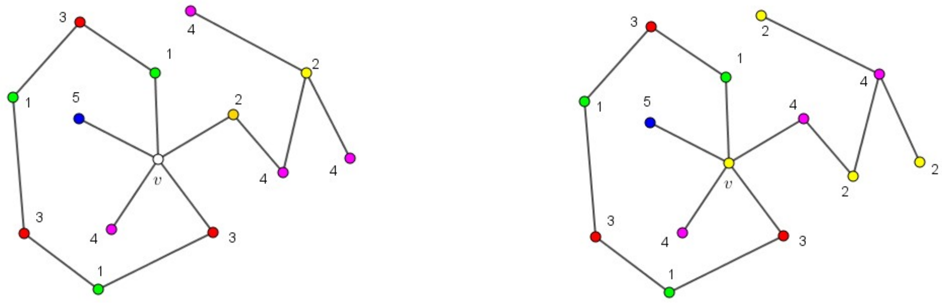
If v_1 and v_3 belong to different components of $G_{1,3}$, we may switch the colors of the vertices of the component of $G_{1,3}$ which contains v_1 such that the color 1 is not involved in the coloration of any vertex adjacent with v. Therefore, we may assign the color 1 to v and the 5-coloring of G results.





[1]It suffices to permute the vertices.
[2]In the following we will refer to such subgraphs as *1-3 tree-subgraph* of G.

On the other hand, if v_1 and v_3 belong to the same component of $G_{1,3}$, then there exists a path, say P, in G joining v_1 and v_3, whose vertices are colored with colors 1 or 3 only (we are considering $G_{1,3}$). We may consider the cycle composed by P together with the path v_1vv_3: the cycle must encloses the vertex v_2 or both the vertices v_4 and v_5. In any case, there cannot exists a path joining v_2 and v_4 whose vertices are colored with colors 2 or 4 only, otherwise planarity would be violated. Thus, it is enough to consider $G_{2,4}$, so that the unique possibility is the one that provides that v_2 and v_4 lie in different components of $G_{2,4}$; therefore, as above, we can swap the colors of the vertices of the component of $G_{2,4}$ which contains v_2, in such a way that no vertex adjacent to v has color 2 and by assigning to v the color 2, a 5-coloring of G results.



□

It is important to note that in this case, as well as in the previous one, the planarity assumption plays a key role. A counterexample is given by the non-planar graph K_7, which turns out not to be 5-colorable.

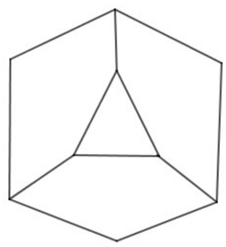
Chapter 5

The Four Color Theorem

We have all either asked ourselves how many colors it suffices to paint any geographic map at least once in our lives or just noticed how beautifully these maps are colored. We may think that it would be an easy problem to solve, but it is not, in fact.

5.1 A Historical Introduction to The Four Color Problem

The first that wondered himself about the problem of coloring a geographic map had been Francis Guthrie, who conjectured that the countries of every map could always be colored with only four colors in such a way that no two adjacent countries[1] have the same color. In 1852 he wrote his brother Frederick about this tricky issue, but, since he was not able to solve it, he asked his professor Augustus De Morgan for some help. They were able to state that the three color conjecture is false since a map with four countries such that each country is adjacent to the other three can be drawn and it requires four colors. The following figure represents an example of planar map that cannot be colored with three colors since each region is adjacent to the others.



[1]By adjacent countries we mean the neighbors of a country i.e. all the countries that share a border with that country.

Moreover, DeMorgan proved that it is not possible for five countries to be in a position such that each of them is adjacent to the other four[2]. This statement looks like the solution of the "Four Color Conjecture" but it is not in fact and many mathematicians have been tricked by that.

This simple-sounding statement started to become a very difficult and complicated problem, that took many mathematicians' effort and more than a century to be solved. The first attempt was done by Arthur Bray Kempe, who published a paper that claimed to prove that the conjecture was true. Kempe's argument turned out to be very clever, however an error was found by Percy John Heawood eleven years later. Despite this flaw, Kempe's proof has been the guide for all the succeeding proofs.

In 1913, George David Birkhoff analysed Kempe's error and developed new techniques to approach the problem starting from Kempe's ideas. In the following years these new methods supplied very powerful tools to all later mathematicians in order to verify the conjecture. In 1950, Philip Franklin showed that every planar map with less than 36 countries admits a 4-coloring.

After these achievements, two more main concepts were introduced and developed by Heinrich Heesch: *reducibility* and *discharging*. In 1977, the final proof has been published by Kenneth Appel and Wolfgang Haken. What is surprising and bothering at the same time it is the involving of computers' effort and programming work, something that has been concerning the mathematical community for a long time.

It is now clear how the "Four Color Theorem" started to become a worldwide problem, despite its "good-looking" statement:

5.1 THEOREM. *The regions of any planar map*[3] *can be colored with four colors, in such a way that any two adjacent regions have different colors.*

5.2 Geographic Maps

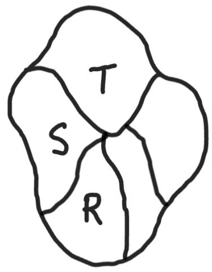
In this section we analyse the correspondence between geographic maps and plane graphs. In addition, without loss of generality, we will make some explicit restrictions on the maps we will consider.

A *geographic map* can always be seen as a plane divided into different regions separated by borderlines. We say that two regions are adjacent if they share a borderline. In particular, if two countries share only a point they are not adjacent. If two regions, R and S, share a borderline we will call S a neighbor of R and viceversa. Moreover we will call "neighborhood" of a region the set of all the countries that are adjacent to such region.

[2]This result can be easily obtained from Theorem 3.12, since such a map is associated to the graph K_5, which is non-planar.

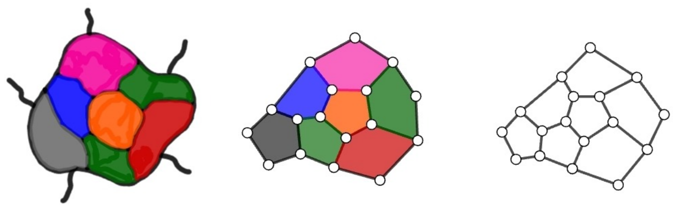
[3]See Section 5.2

The figure below represents an example of geographic map, where R and S are neighbors, but T and R are not.



Any geographic map can be associated with a plane map in the sense described above. Furthermore, any region can be approximated to a polygon of n, with n the number of neighbors it has, such that each side is the boundary with another region. It now becomes clear how a geographic map can be associated with a plane graph, since the correspondence between a plane map and a plane graph has already been dealt with (see Chapter 3).

The following image represents a geographic map together with the associated plane map and plane graph respectively.



In the following, we will see how it is possible to make this correspondence even more useful to the proof of the Four Color Theorem.

While discussing the Four Color Theorem, we will restrict ourselves to maps that do not have disconnected countries (both enclaves and exclaves are not allowed), otherwise four colors could be not enough to color maps having many states with either inclaves or exclaves. For that reason coloring the planisphere requires more than four colors (oceans should be colored as well) as an ocean borders with each state on the coast as well as all the islands. Moreover, we will consider maps where no country completely surrounds another country or group of countries, and no more than three countries meet at any point. This last restriction will be better explained in the next section, and it will be clear how it will not lead to a loss of generality.

It is important to keep in mind that from now on, when we appeal to plane maps we are referring to maps of the type described above.

Let us now consider any geographic map as described above. We just saw how it is always possible to restrict ourselves to a plane map.

In order to prove Theorem 5.1, we may rephrase its statement as follows:

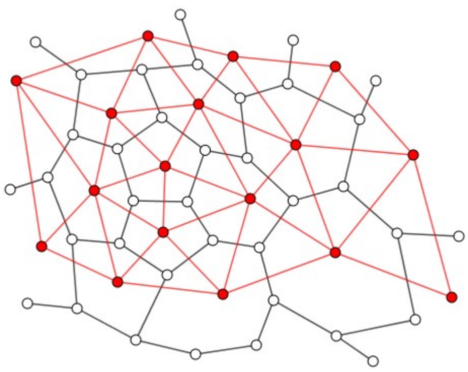
5.2 THEOREM. *Every plane graph admits a vertex-coloring with at most 4 different colors such that all the vertices have different colors pairwise.*

Let us show why this two sentences are equivalent. We start by showing that Theorem 5.1 implies Theorem 5.2 and then conversely.

Proof. Consider a plane map: the *dual graph* is the graph constructed by choosing a point (it could be seen as the capital city) on each country of the map, and joining all the points that lie in two adjacent countries with an edge. In this way a plane graph is produced, whose vertices are all the "cities" chosen to represent each country, and the edges are the lines that connect two cities of two adjacent countries. The degree of each vertex in the dual graph is the number of neighbors of the country (of the original map) represented by that vertex.

On the other hand, if we consider a plane graph whose vertices can be four colored such that any two adjacent vertices do not have the same color, it is always possible to surround each vertex by a n-polygon such that n is the number of adjacent vertices the vertex we are taking in account has. In this way, each polygon is a region of the plan map formed by collecting all the polygons together. Notice that this operation preserves adjacency so that it suffices to assign to the region surrounding each vertex the same color as the vertex. □

In the figure below we can see an example of a plane map (the one with black boundaries) and its dual graph (the red one):



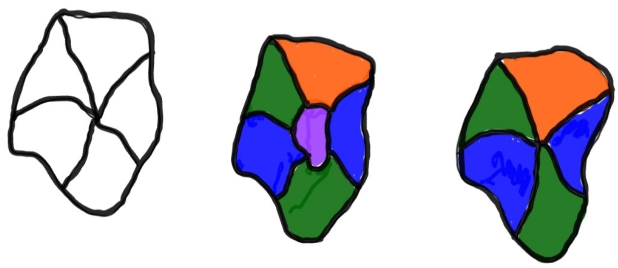
Now, it is clear how a four coloring of a plane map corresponds to a vertex four coloring of a plane graph.

5.2.1 Planar Graphs and Triangulations

Below, we make explicit some assumptions we took in the previous section.

5.3 DEFINITION. A *cubic* map is a map where there are strictly three borderlines at every meeting point of countries.

Any map can be turned into a cubic map. This is done by replacing a point where more than three countries meet by a new region. This region will have a borderline with all the other regions. This operation will not affect whether the map can be four colorable or not. Indeed, if this new map can be colored with four colors, then so can the original map since the original regions are still not adjacent in the new cubic map. Removing the new region will not alter the coloring of the map. So proving that the theorem is true with cubic maps proves that it is true for all maps.



Thanks to the relationship between planar maps and graphs that we discussed earlier, we want to define the notion of *triangulation* and make explicit the relationship with cubic maps. With an abuse of notation, we will refer to the faces of a planar graph as the regions bounded by vertices and edges of the graph associated with any planar map.

5.4 DEFINITION. Let f be a face of a planar graph G. f is called *triangular* if it is bounded by exactly three edges of G.

A plane graph is called a *triangulation* Δ if all its faces are triangular.

Notice that the graph associated to a cubic map is always a triangulation. Henceforth, we will indifferently call both the map and its associated graph triangulation, thus a "full" correspondence between planar maps and planar graphs is achieved.

5.5 DEFINITION. A planar graph G is called *maximal* if by adding any edge, it would not be planar anymore.

Thanks to this definition and Proposition 3.7 we have that all the faces of a maximal graph G must be bounded by exactly three lines, i.e. the graph G is a triangulation.

We now want to show that there exists always a triangulation Δ such that contains a graph G as a subgraph.

First, consider the case where G has $4-i$ vertices with $i = 1, 2, 3, 4$, then we may construct the graph G_1 arose from G by adding i vertices in a face of G (if any). Otherwise, if G has exactly four vertices or more, let $G_1 = G$.

We now face two possibilities: either G_1 is maximal or not. If G_1 is maximal, than it must be a triangulation because of what we said earlier. Otherwise, there are two distinct vertices, say v, w, that lie in the boundary of a face F of G_1 but there is no edge of G joining them. Thus, we construct the graph G_2 by adding to G_1 an edge joining v and w that crosses the interior of the face F. We repeat this process until the graph G_j results maximal and hence a triangulation Δ which contains the planar graph G.

It is important to underline that a coloring on the triangulation Δ leads to a coloring on the graph G, thus it is not restrictive to work with triangulation instead of general planar graphs.

5.3 Kempe's Argument

In the following section we present the first attempt to prove the *Four Color Theorem* made by Arthur Bray Kempe.

Kempe attacked the conjecture by assuming it is false (that there exists at least a map which requires five colors) in order to show that this assumption leads to a contradiction.

Let us consider a plane map such that no more than three countries meet at a single point and none of its countries encloses other countries (Kempe called such maps *normal maps*). This restriction does not affect any generality of the problem as explained in Section 5.2. From now on, we will write map in the place of normal map.

Kempe's idea is to work with a *minimal 5-chromatic map*[4] in order to show that every such map cannot contain any region with exactly two, three, four nor five borders[5]. The English mathematician correctly proved the first three cases (two, three, four borders), however he made a mistake in the "five borders" case. The most important thing is the method he used to prove his claim: despite the error he made, Kempe supplied the argument which led the ones that came after him to the correct proof a century later.

Consider Euler's formula 3.1.2

$$p - q + r = 2$$

[4]A minimal 5-chromatic map is a plane map which admits a five coloring. Moreover, whenever a vertex or an edge is removed the map is four colorable i.e. it does not exist a 5-chromatic map smaller than a minimal 5-chromatic map.

[5]We are assuming that the countries are polygons which share each side with at most a country. This kind of maps is always possible to obtain without loss of generality (see Section 5.2)

with p the number of vertices, q the number of border lines and r the number of countries in the map. Since, without loss of generality, every map is a triangulation we have that $p = \frac{2q}{3}$, therefore the 3.1.2 turns to

$$3r - q = 6 \tag{5.3.1}$$

Let f_n be the number of countries that have exactly n adjacent countries. It follows that $r = f_2 + f_3 + \cdots + f_N$ and $2q = \sum_{i=2}^{N} i \cdot f_i = 2f_2 + 3f_3 + \cdots + Nf_N$ with N the largest number of neighbors any country has. [6]

If we replace the relation we just presented into (5.3.1) we get

$$3 \cdot (f_2 + f_3 + \cdots + f_N) - \frac{1}{2} \cdot (2f_2 + 3f_3 + \cdots + Nf_N) = 6 \tag{5.3.2}$$

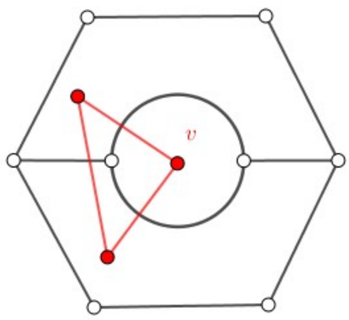
that is

$$4f_2 + 3f_3 + 2f_4 + f_5 - f_7 - \cdots - (N-6)f_N = 12 \tag{5.3.3}$$

This result has a paramount consequence: since f_i for $i = 2, \ldots, N$ is a non negative number and the number on the right of the equation is positive, there must be some countries with either 2, 3, 4 or 5 border lines. The next step is to show that a minimal 5-chromatic map that contains such countries cannot exist.

Case 1: two neighboring regions.

Let us consider the case in which there is a minimal 5-chromatic map containing a country that has exactly two neighbors.

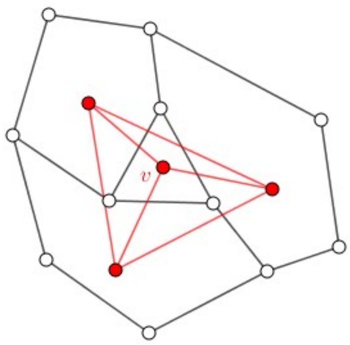


Let G be the dual graph associated to the map and let v the vertex of G corresponding to the region of the map which has only two neighbors. Consider the graph $G - v$ and a four coloring for it which exists by the minimality of G. Since v has only two adjacent vertices, we have that at

[6]The reason why we exclude f_0 and f_1 is that we are assuming that the graph is connected and that it is not possible to have one country completely surrounding another.

most two colors are used for the two vertices. Now, it is possible to assign to v one of the colors that has not been assigned to its neighbors yet. This operation leads to a four coloring of the graph G, which contradicts our assumptions.

Case 2: three neighboring regions.
Let us consider the case in which there is a minimal 5-chromatic map that contains a country with exactly three neighbors.



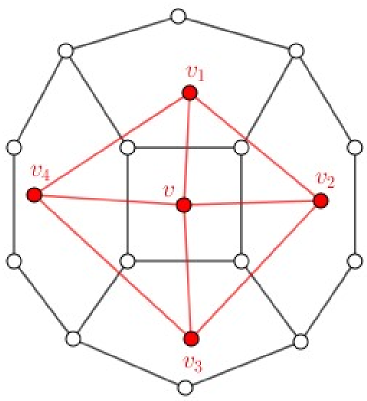
Let G be the dual graph associated to the map and let v the vertex of G corresponding to the region of the map which has only three neighbors. Consider the graph $G - v$, which is four colorable by assumption. Since v has three adjacent vertices, we have that at most three colors are needed to color these three vertices. Now, it is possible to assign to v the color (or one of the colors) that has not been assigned to its neighbors yet. This operation leads to a four coloring of the graph G, despite our initial assumptions of minimality.

Case 3: four neighboring regions.
Let us consider a minimal 5-chromatic map such that there is a country with four adjacent countries. In this case there exists a vertex v of the dual graph G which has degree four.

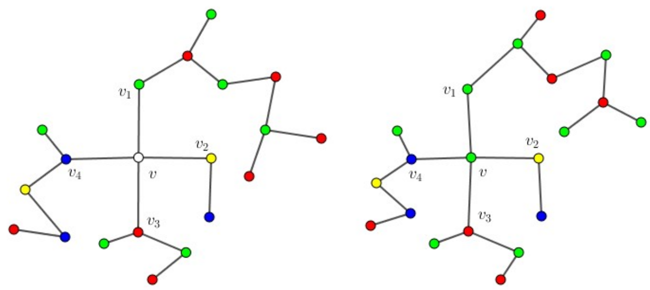
Consider the graph $G - v$ that is four colorable by assumption. There are two possibilities:

1. If less than four colors are involved in the coloring of the adjacent vertices of v, then, it is sufficient to assign to v the color (or one of the colors) which is not involved in the coloring of the neighborhood of v yet.

2. If the adjacent vertices of v in the graph G are colored with all the four different colors, then, without loss of generality, we may assume that the color i is assigned to the neighbor v_i of v with $i = 1, 2, 3, 4$, in clockwise order. Let us consider the vertices v_1 and v_3.



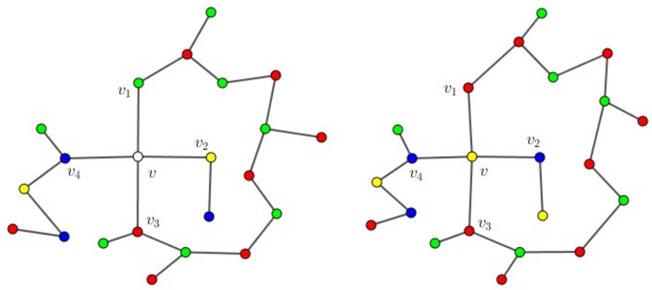
If there is no 1-3 chain[7] joining v_1 and v_3, then we consider the connected subgraph of G whose vertices are the ones that are colored either 1 or 3 together with v_1. We call such connected subgraph of G the *1-3 tree-subgraph* of G starting from v_1. Surely, v_3 does not belong to this subgraph as we are assuming that there is no 1-3 chain from v_1 to v_3. In this case, it is sufficient to interchange vertex colors of such subgraph such that v_1 turns out to be colored with color 3. It now becomes possible to assign color 1 to the vertex v, since vertex v_1 has been reassigned color 3, and thus no vertex adjacent to v turns out to be colored with color 1.



Otherwise, a 1-3 chain joining the vertices v_1 and v_3 exists, then a 2-4 chain cannot exists since G is a plane graph and the two paths would intersect. In this case, as in the previous one, it is possible to interchange the colors of the vertices of the 2-4 tree-subgraph of G starting from v_2. Notice that a 2-4 chain cannot exist, thus v_4 does not belong to the 2-4 tree-subgraph of G starting from v_2 so that v_4 will not be affected from this color interchange. This operation makes the vertex v_2 end up with color 4, so that none of the four neighbors

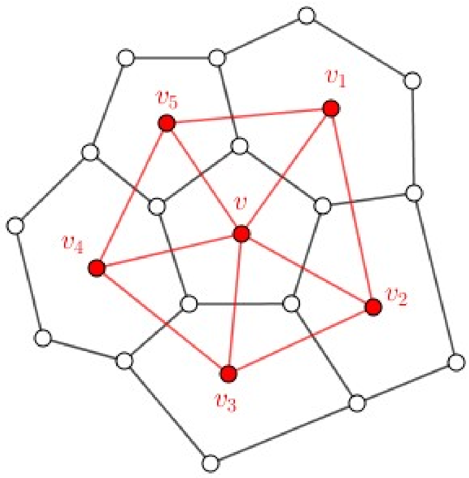
[7] A 1-3 chain is a path that alternates a 1-colored vertex with a 3-colored one, i.e. a path that joins vertices whose assigned colors may be only either 1 or 3 .

of v is colored with color 2, which can be finally assigned to the vertex v.



After that Kempe tried to generalize this smart argument to a country with five neighbors in order to complete the proof. However, it did not work in this case and some years later, Percy John Heawood pointed out the error that Kempe committed. Let us see how Kempe's fault has been pointed out by Heawood.

Consider the dual graph G of a plane map that contains a country with five neighbors.



Let v_i with $i = 1, 2, 3, 4, 5$ be the vertices adjacent to v, in clockwise order. As in the previous case there are two possibilities.

1. If the neighbors of v are colored with less than four colors, than it is sufficient to color v with a color that has not been used in the coloring of the neighbors of v yet.

2. Otherwise, since the graph G is a minimal 5-chromatic graph, there exist a coloring of $G - v$ that involves exactly four colors. Without loss of generality, assign the colors 1, 2, 1, 3, 4 to the neighbors of v

in clockwise order starting from v_1 i.e. we assign color 1 to v_1, color 2 to v_2, color 1 to v_3, color 3 to v_4 and color 4 to v_5.

If there exist no 2-4 chain that joins v_2 and v_5 then we consider the 2-4 tree-subgraph of G starting from v_2. Surely v_5 does not belong to this subgraph as we are assuming that there is no 2-4 chain from v_2 to v_5. Now we can interchange the colors of vertices of the 2-4 tree-subgraph of G starting from v_2 such that v_2 turns out to be colored with color 4. In this way, it is possible to assign color 2 to the vertex v, since color 4 has been reassigned to the vertex v_2 and thus no vertex adjacent to v is colored with color 2. This operation leads to a contradiction since four colors would be enough to color the whole graph G.

On the other hand, suppose that a 2-4 chain joining v_2 and v_5 exists. Analogously, we may assume that a 2-3 chain joining v_2 and v_4 exists. This fact together with the planarity of G imply that cannot exist neither a 1-3 chain from v_1 to v_4 nor a 1-4 chain joining v_3 and v_5. Until this point Kempe's work is impeccable, however his conclusion is not as flawless. According to Kempe, it would be enough to interchange the colors of the vertices of the 1-3 tree-subgraph of G starting from v_1 and the ones of the 1-4 tree-subgraph of G starting from v_3 in order to end up with v_1 colored with color 3 and v_3 with color 4. By doing this, none of the vertices adjacent to v would be colored with color 1, hence color 1 can be assigned to the vertex v. This would imply a contradiction since four colors would be enough to color the whole graph G.

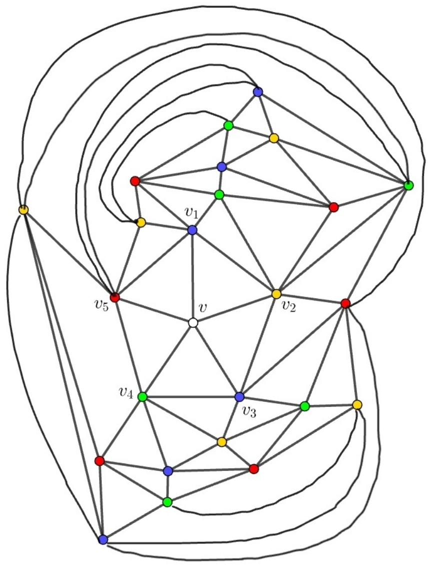
Kempe, thanks to this argument, was convinced that he proved the Four Color Theorem. For instance, he showed that in any minimal 5-chromatic map there should be at least one country with two or three or four or five neighboring countries and simultaneously that in such a map it was not possible to have any of the above possibilities.

Few years later, Heawood pointed out with a counterexample (figure below) that interchanging colors of the 1-3 tree-subgraph starting from v_1 can "produce" a 1-4 chain joining v_3 and v_5 and no color is available to be assigned to v. Heawood's example shows that Kempe's technique is not complete, in the sense that it may not always work .

It is important to underline, thanks to Kempe's work[8], that (5.3.3) may be rephrased (for minimal 5-chromatic cubic maps) as follows:

$$f_5 = \sum_{i=7}^{i_{max}} (i-6) f_i + 12 \tag{5.3.4}$$

[8] According to what we just stated above Kempe proved that a country with less than four neighbors cannot appear in a minimal 5-chromatic cubic map



Indeed, Kempe was able to show correctly that countries with two, three and four neighboring countries cannot appear in a minimal 5-chromatic cubic map. This fact implies that in (5.3.3) we have $f_2 = 0, f_3 = 0$ and $f_4 = 0$ (since the number of countries with i neighboring countries for $i = 2, 3, 4$ is zero) so that (5.3.4) is obtained.

5.4 Reducibility and Discharging

In the following section we analyse these two paramount methods. First of all, it would be useful to explain what a configuration is: *configurations* are technical devices that allow us to capture the structure of a small part of a larger triangulation. As we already saw, Kempe proved that there must be certain configurations in a plane map i.e. there must be a country with either two, three, four or five border lines (neighbors). This idea may explain the other important feature: *unavoidability*. An unavoidable set is a set of configurations that must occur in any plane map.

In conclusion, Kempe found an unavoidable set of configurations consisting of a country with either two, three, four or five neighbors, that is, every plane map must contain at least one of these four configurations.

5.4.1 Reducibility

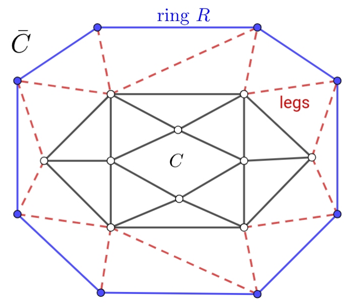
A configuration is said to be *reducible* if any coloring of the rest of the graph can be extended onto the configuration. The idea is to apply this concept to minimal 5-chromatic maps and their associated graphs. If a graph contains a reducible configuration it cannot be contained in a minimal 5-chromatic one. Indeed, the subgraph of a minimal 5-chromatic graph from which the configuration has been removed is four colorable (as otherwise the original would not be a minimal one). Since we are assuming that the configuration is reducible we have that the coloring of the rest of the graph can be extended onto the configuration, making the whole graph four colorable despite the fact that is was a 5-chromatic one. In particular, we will find it useful to characterise a reducible configuration as one that cannot be contained in a minimal 5-chromatic graph.

Notice that, because of the correspondence between maps and graphs described above, we will say indifferently that a reducible configuration cannot be contained in a minimal 5-chromatic map or graph.

It is now clear that Kempe correctly showed that three of his four configurations are reducible but he failed to show the reducibility of a country with five neighbors. The reason why his attempt is important is that he understood that finding an unavoidable set of reducible configurations would be sufficient to prove the theorem. Moreover, he set the fashion that all the later mathematicians followed to prove the conjecture, that is, first to find an unavoidable set of configurations (i.e. if a minimal 5-chromatic map existed, then it would certainly contain at least one of its configurations), and finally to show that each configuration in such set is reducible, hence it cannot be contained in a minimal 5-chromatic map and such a minimal 5-chromatic counterexample cannot exist.

Heesch was the first, after Kempe, who conjectured that an unavoidable set of reducible configurations could be found. Moreover, he observed that at least one of the methods to reduce a graph consisted in a sufficiently mechanical procedure to be done by computers. One of Heesch's students, Karl Dürre, wrote the first computer program to test the so call *D-reducibility*. In order to introduce what being D-reducible means, let us consider a planar map. We have already seen that it is always possible to consider the normal map and the graph which is associated to the original map. Moreover we can consider the dual graph so that it results a triangulation. To show that a configuration C is D-reducible is to show that if the assumption that C is contained in a minimal 5-chromatic triangulation Δ leads to a contradiction: if we consider $\Delta - C$, (i.e. if we remove the configuration C from the triangulation Δ), then it must be 4-colorable (we are considering a minimal 5-chromatic map). Thus, if the 4-coloring can be extended to C i.e. the triangulation Δ can be 4-colorable, then it contradicts our assumption. To show that a configuration is D-reducible is exactly to end up with such contradiction which implies that the original triangulation (henceforth the original map) was not a minimal 5-chromatic one.

Before seeing how this can be checked, let us introduce some terminology. Let C be a configuration of a planar triangulation Δ and call $\bar{C}$ the larger subgraph of Δ obtained by adding the vertices adjacent to C and the edges joining them. The new edges joining a vertex of C with one of $\bar{C} - C$ are called *legs*, and the part of the triangulation composed by the added vertices along with the all the edges joining them is called the *ring* R of C.



We write "n-ring" in the place of "a ring of size n" to denote the amount of vertices the ring is composed by. If we consider a four coloring of a ring vertices, say ϕ, we have that it is possible to obtain *equivalent* four colorings by permuting the colors. For this reason we let ϕ_n the set of all the *inequivalent* four colorings of a n-ring. Now, let us introduce what an *initially good coloring* is. Take any four coloring for the ring R of a configuration C, we say that it is initially good (with respect to C) if it can be extended to a coloring of the ringed configuration $\bar{C}$.

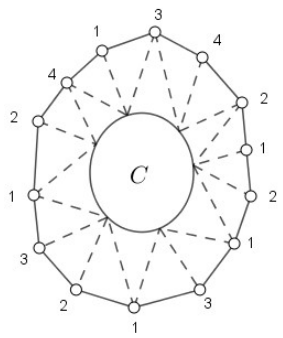
It is now clear what our aim is: it suffices to take a minimal 5-chromatic triangulation Δ, and show that all the (inequivalent) four colorings of R are initially good with respect to C, thus get a contradiction as we are assuming that the triangulation is not four colorable. This is exactly showing D-reducibility.

Consider a minimal 5-chromatic triangulation Δ which contains a configuration C and list all the four colorings for the ring of C. The clue is to show that each of them can be "*transformed"* into an initially good one. The technique to "transform" a four coloring is as follows:

- Let μ be a four coloring of the ring R of a configuration C of Δ. There must exist such a coloring since by assumption we have that Δ is a minimal 5-chromatic triangulation, thus R can be seen as the restriction $\Delta - C$ of the triangulation. We may denote the four colors as numbers as we did above, hence we will use $\{1, 2, 3, 4\}$ to indicate the set of colors.

  For example consider the 14-ring of a configuration C and let $\mu = 12312413421213$ one of its four coloring.

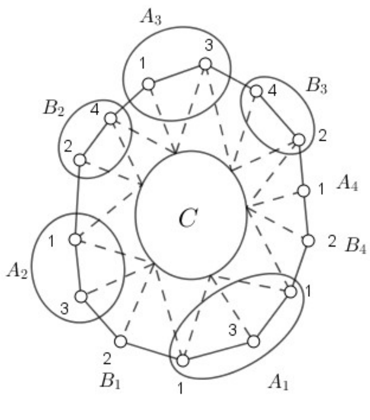
- After that, choose a pairing of colors between all the different possibilities
  $$\{(1, 3)(2, 4), (1, 4)(2, 3), (3, 4)(1, 2)\}$$
  In our example we will consider the pair (1,3) and (2,4).



- Now, let us define the *Kempe components* of $\Delta - C$. G is a Kempe component of $\Delta - C$ if it is a subgraph of Δ which contains all the vertices of $\Delta - C$ and such that two vertices are joined by an edge if they are adjacent in $\Delta - C$ as well as they have paired colors. Moreover, we define the *Kempe component* of R as the subgraph of G intersected with R.

- Label the Kempe components of R as $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ such that every A_i component is colored with the first pair of colors and every B_i component with the other pair of colors, with $i = 1, \ldots, m$.

  For example we can consider the following Kempe components of R with respect to the pair (1,3) and (2,4) (see figure below.).

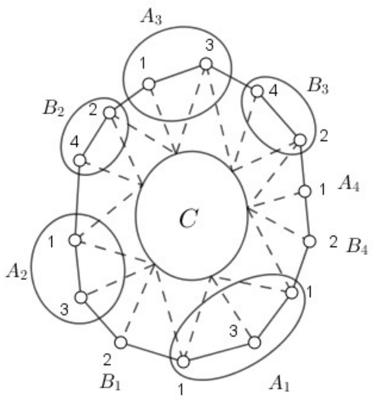


  It is important to notice that if A_i and A_j for some $i, j \in \{1 \ldots, m\}$ are in the same Kempe component of $\Delta - C$, then any B_h and B_k, with A_j and $i \leq h < j \leq k$, cannot lie in the same component of $\Delta - C$, because any two paired color path joining B_h and B_k would be blocked from the other two paired color path that links A_i.

  Furthermore, it is clear that whenever two paired colors of any Kempe component of $\Delta - C$ are interchanged, then it would produce a new four coloring of $\Delta - C$ since no pair of adjacent vertices could end up with the same color.

- Assume that A_i and A_{i+1} are in the same Kempe component of $\Delta - C$, then we can switch the colors of the vertices of the B_i Kempe component of R without affecting the colors of any other ring vertices. This operation leads to a new four coloring, say ψ of the ring R. Let us suppose that ψ is an initially good four coloring for the ring R, thus we can conclude that Δ is four colorable.

  For example we can switch the colors of the vertices of B_2, so that a new four coloring of the ring arises.



To briefly summarize the procedure, given a four coloring of $\Delta - C$, when restricted to the ring R, it induces the four coloring μ, that can be changed into the initially good four coloring ψ, therefore it is extendable to the configuration C and hence to the whole planar triangulation Δ, leading to a contradiction. Ultimately, to show the D-reducibility of a configuration C is exactly to show that every four coloring of $\Delta - C$ can be transformed into a four coloring that is extendible to C.

Few years later Birkhoff noticed that it is not necessary to show that each of the four colorings of $\Delta - C$ is extendible to C, but it suffices to consider particular subsets of ϕ_n and to show that they are initially good colorings with respect to C. A configuration that satisfies this more sophisticated way of testing reducibility is called *C-reducible*.

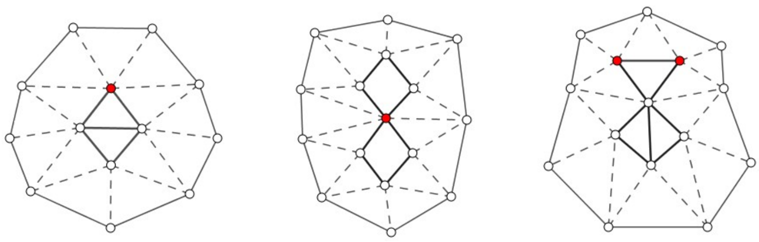
Reduction Obstacles

Heesch, during his studies, observed that a number of distinct phenomena, occurring into configurations, provide clues to the likelihood of successful reducibility. He noticed that, under particular conditions, no reducible configuration were never been found. These particular situations are known as *“reduction obstacles”*. In particular, if one wanted reducible configurations, they should avoid these reduction obstacles. The three arrangements of vertices that Heesch found to appear in any reducible configuration are the following (in the figure below the three cases are shown respectively):

1. a “$\geq$ 4-legger vertex”, that is a vertex of any degree d which is connected to fewer than $d-3$ other vertices of the configuration.

2. A “$\geq$ 3-legger cut vertex”, that is any vertex of degree d that is connected to fewer than $d-2$ other vertices of the configuration whose removal disconnect the configuration.

3. A “*hanging pair*” which consists in two vertices of degree five of the configuration that are joined to each other and to other vertices of the ring, but such that have only one other common neighbor in the configuration.



Heesch stated an important rule that allows to test a configuration for reducibility obstacles. Let us see the main steps to test a given configuration:

1. Whenever a $\geq$ 4-legger vertex occurs in the confguration it may be removed (along with all the incident edges).

2. Whenever a hanging pair appears, then both vertices may be removed to form a smaller configuration.

3. Whenever a $\geq$ 3-legger cut vertex occurs, it may be removed to form a couple of smaller configurations.

At any stage, one or more vertices are removed from the configuration. The test ends when none of the previous steps can be applied to the configuration, i.e. the final configuration is either empty or it does not contain any reduction obstacles. If all the resulting configurations are empty or known to be non-reducible than the original configuration it is certainly not reducible.

On the other hand, it has been investigate that for a n-ring, the likelihood of reducibility rapidly increases with the number of vertices inside the ring. To be more precise, if any configuration, with m vertices, satisfies

$$m > \frac{3n}{2} - 6$$

then it almost certainly reducible since the chance to have obstacles to reducibility sharply decreases.

5.4.2 Discharging

The *discharging* process, created by Heesch, is a method that allows to construct an unavoidable set of configurations. The first example of unavoidable set, as we already said above, is the one found by Kempe, whose elements are countries with two, three, four and five neighbors. Unfortunately, despite Kempe's incorrect attempt, this set was not proved to be reducible.

The crucial idea of this technique is to assign to every vertex of a triangulation a charge depending on the degree of the vertex and by moving the charge as on an electrical network, ending up with a (possibly reducible) unavoidable set of configurations for a minimal 5-chromatic planar triangulation. To deeply describe this discharging process let us recall that Kempe was able to prove that a vertex of degree less than five cannot occur in a minimal 5-chromatic planar triangulation. According to (5.3.4) and what we said above, it holds the following:

$$v_5 - v_7 - \cdots - (M-6)v_M = 12 \tag{5.4.1}$$

with v_i the number of vertices of degree i for $i = 5, \ldots, M$. At this point, once that an initial charge is assigned to each vertex, the aim is to set a list of discharging rules in order to move the charge around.

The previous formula (5.4.1) sets some constraints that a final distribution of a configuration must comply with, so that the number of possible configurations which can occur is finite. The finite possibilities of rearranging the initial charge form (what we previously called) an unavoidable set.

The main idea is to assign to every vertex of degree i an initial charge of $6 - i$ units, such that the only initially positively charged vertices are those of degree five. An initial negative charge is assigned to each of the other vertices (except those of degree six) which are called *major* vertices. According to (5.4.1), we have that the total initial charge of the triangulation is positive. We now want to move the charge along the vertices in such a way that a vertex positively charged transfers some of its charge to a vertex that is negatively charged. In order to do that, a set of rules is laid down so that charge transfer is subject to such rules. The result of applying this discharging process to a triangulation change the initial distribution of charge into a final one, which must comply with some constraints since the triangulation's total charge must stay positive.

Let us give an example in order to clarify the strategy to obtain an unavoidable set of configurations.

Example. Let us consider the procedure which consists in a single rule: a degree-five vertex transfers $\frac{1}{5}$ unit of charge to each of its major neighbors.

Our goal is to find the set of all possible configurations that can be formed by moving the charge between the vertices according to the rule described above. We claim that such corresponding (unavoidable) set consists of two configurations: the first one is a pair of degree-five vertices joined by an edge and the second one is a degree-five vertex joined by an edge to a degree-six vertex. These two configurations are shown in the figure below where each vertex's degree is indicated.

Let us prove that, according to that rule, these two configurations are the unique situations that may happen, as far as we want to preserve a final positive charge.

We know from above that for all v in the triangulation an initially charge of $6 - deg\ v$ is assigned, so that vertices of degree five are the only ones with an initially positive charge of 1. Let us study what final configurations are possible by varying the degrees of a vertex.[9]

Case 1: degree-five vertex. According to the unique rule established above, the only way that a degree-five vertex can end up with positive charge is having at least a non-major neighbor, as otherwise the whole charge of 1 would be transferred to its adjacent vertices. Indeed, if a degree-five vertex had only degree-seven neighbors, according to the rule, a charge of $\frac{1}{5}$ would be transferred from the degree-five vertex to its five adjacent vertices so that the degree-five vertex would lose an amount of charge equal to $\frac{1}{5} \cdot 5 = 1$, thus it would end up completely discharged.[10] It is for that reason that a degree-five vertex (under the rule established above) must have at least a neighbor of either degree five (the situation corresponding to the first configuration) or six (corresponding to the second configuration). These two possibilities are fully identified by those of the set described above.

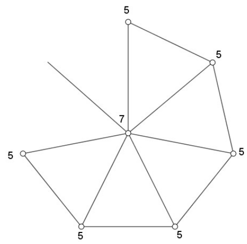
Case 2: degree-six vertex. No possible configurations are available in this case as a degree-six vertex cannot lose or gain charge since it is already discharged.

Case 3: degree-seven vertex. A degree-seven vertex, which has an initial charge of -1, can only become positive if it has at least six neighbors that are degree-five vertices. In such situation at least two of them are joined by an edge so that the first configuration of the unavoidable set occurs.

Case 4: degree-eight (or higher) vertex. A vertex of degree eight or higher cannot become positive even if all of its neighbors are degree-five vertices. This can be easily seen by examine a degree-eight vertex (the

[9] Notice that it is non-sense studying vertices of degree lower than five since they cannot occur in a minimal 5-chromatic planar triangulation.

[10] Same situation would happen if the degree-five vertex had only major neighbors.



reasoning for vertices with higher degrees follows the same idea). An initial charge of -2 is assigned to each vertex of degree eight, so that even if it had all degree-five neighbors the maximum charge it can gain is 8 times $\frac{1}{5}$, that is less than two, hence it cannot end up with a positive charge.

Since all the possible degrees for a vertex have been surveyed, the two (non reducible) configurations form an unavoidable set, i.e. at least a member of the two-configurations set will be found in every minimal 5-chromatic planar triangulation.

In conclusion, the work made by Appel and Haken focused on the development of the set of rules defining the discharging process, which could supply a not-too-large unavoidable set, whose elements can all be shown to be reducible configurations. In order to improve the set of unavoidable configurations, the discharging method has been developed in the years that followed.[11]

In the following we present the earliest version of the discharging procedure made by Haken and Appel (the first version only provided for one rule).

Rule. Whenever the triangulation contains an edge that joins a vertex of degree five with a major vertex a charge of $\frac{1}{2}$ of unit is transferred along the edge from the degree-5 vertex to the major one (such a transfer is called *regular discharging* or also *R-discharging*).

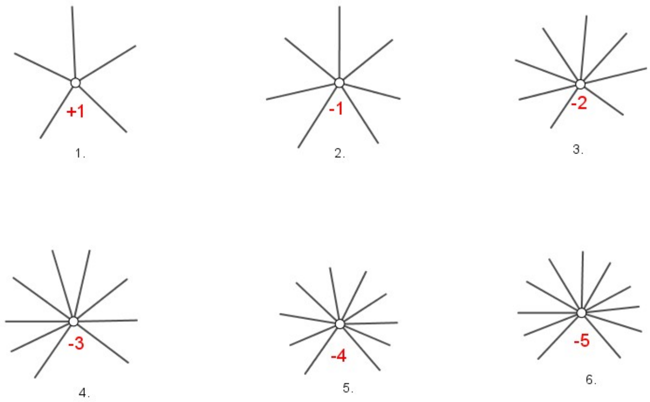
It can be seen that, under such rule, the possibilities to get a final positive distribution of charge on a vertex v of the triangulation are the following:

1. v has degree 5 and has either 0 or 1 major neighbors;
2. v has degree 7 and has between 3 and 7 neighbors of degree five;
3. v has degree 8 and has between 5 and 8 neighbors of degree five;
4. v has degree 9 and has between 7 and 9 neighbors of degree five;

[11] In 1976 the unavoidable set consisted of 1936 configurations. Since then, Appel and Haken found and eliminated about 100 redundancies, that are those configuration that were accidentally listed twice or were a part of other configurations. Moreover, thanks to the new discharging techniques, they found other 352 configurations that were not really needed. Finally, in 1985, 16 more redundancies were found, bringing the size of the set down to 1476 configurations.

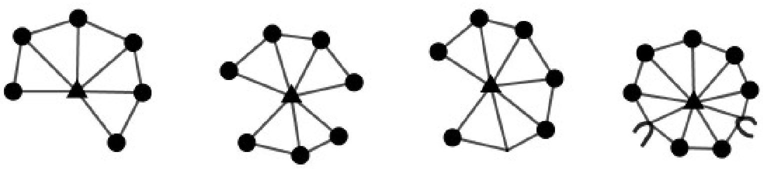
5. v has degree 10 and has either 9 or 10 neighbors of degree five;

6. v has degree 11 and has 11 neighbors of degree five.

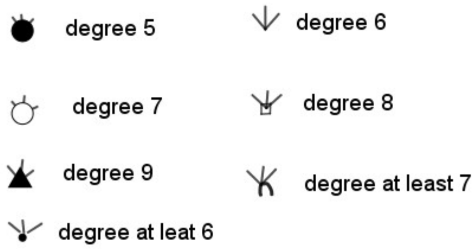


Since the total charge is positive, at least one of these situations occurs in the triangulation, thus the set of all the unavoidable configurations may be construct collecting all these cases.

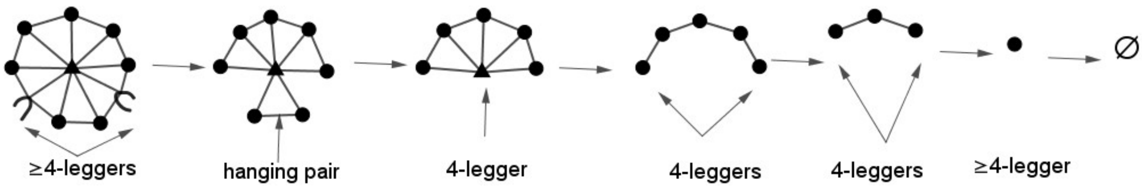
The problem is that, once all these cases are listed, proving the reducibility of all the configurations may not be always possible, so that the unavoidable set would not result reducible. For example, consider the case number 4: the figure below shows the set of all the configurations that satisfy the condition of the case considered, thus at least one of those must occur. In order to indicate the degree of each vertex in a configuration, we



use symbols that Heesch introduced and their meanings are explained by the figure below.



The problem is that the first three configurations shown in the figure above are reducible, while the last one, once the Heesch's rule is applied to it, it result to be non-reducible. All the stages of Heesh's rule are illustrated below.



Appel and Haken spent most of their work in boosting such discharging procedure. They begun by considering some exceptions to avoid "bad" cases as the one shown above. In some situation, a charge of less than $\frac{1}{2}$ is moved from a 5-vertex to a major one; such a transfer is called a *small-discharging* or *S-discharging*. Analogously, in some situations a charge greater than $\frac{1}{2}$ is transferred from a 5-vertex to a major one and those are called *large dischargings* or *L-dischargings*.

By introducing these new rules, in spite of removing some "bad" configurations (in terms of reducibility), they added some others of larger ring size. The two mathematicians, in order to restrict the number of configurations of the unavoidable set and the maximal ring-size that occurs among its members, expanded the discharging procedure with the *transversal dischargings* or *T-discharging*.

Appel and Haken finally produced an unavoidable set of 1825 reducible configurations, which they later made decrease to 1476 elements.

5.5 From a Computer-Checked to a Fully Formalized Proof of the Four Color Theorem

This theorem has an important role in the sense that it has been the first theorem that needed computers to be involved in its proof. Most of the mathematical community was skeptical and distrustful about the use of computers into a mathematical proof. The reason of this rejection is that mathematicians pretend to find a solution as well as to understand its accuracy. Moreover, they do not even stress about whether the proof is correct but rather whether the proof is a valid proof. Finally, with the possibility of a computing error, mathematicians do not feel comfortable relying on a machine to do their work as they would be if it were a simple hand-proof. For all these reasons many mathematicians are still looking for a "better" hand-checkable proof for this problem.

5.5.1 Appel and Haken's Computer-Verified Proof

In order to find an avoidable set of reducible configurations, Appel and Haken, as we already showed above, used two particular methods: reducibility and discharging.

The two mathematicians used a computer program to search for these configurations. They begun to work at a discharging algorithm in order to get an unavoidable set of configurations which had to be tested for reducibility.

Whenever a configuration failed to be reducible, they modified the algorithm in order to get a better discharging procedure that allowed them to construct an unavoidable set that avoids configurations which were previously proved to be non-reducible. First, in order to simplify the problem, they considered a restricted type of configurations, those called geographically good[12]. The first run of the computer program in 1972 gave a lot of information on these particular configurations although many flaws and problems arose. It took almost six months to write a program that would provide a set of unavoidable geographically good configurations.

In 1975 Appel and Haken improved the algorithm by extending it to obstacle free configurations rather than geographically good ones. The size of the unavoidable set doubled; however, for the first time, the two mathematicians figured out that there was a chance to find an unavoidable set of configurations that could be tested for reducibility.

Although the computer work was necessary to find the final unavoidable set, Apple and Haken were able to specify by hand the correct discharging procedure and prove that this procedure produced such unavoidable set of configurations. Indeed, a surveyable proof that the set is unavoidable is presented in [1].

On the other hand, to conclude the proof of the "Four Color Theorem", it is necessary to prove that all the configurations of the unavoidable set are reducible. Appel and Haken established that all the configurations in the unavoidable set previously found were reducible by computer programming. This procedure, unlike the previous one, is not replaceable with a hand-checking work since most of the configurations have a large ring-size (13 or 14). They begun writing a program to test for reducibility, working with the assembler language[13] for the "University of Illinois IBM 360" computer. After some time Appel and Haken started a cooperation with John Koch, a graduated student in computer science, who wrote a program to check for reducibility up to 11-ring size configurations. Despite the collaboration brought important results in testing reducibility, other problems were found, thus the discharging procedure has been improved one more time.

In June 1976, they finally completed the construction of the unavoidable set of reducible configurations, so that the "Four Color Theorem" was proved. The final discharging procedure required the hand analysis of almost 10 000 neighborhoods other than the machine analysis of about 2 000

[12] Configurations that do not contain the first two Heesch's reduction obstacles.

[13] A law-level programming language

cases. After a century of many mathematicians' effort, errors and approximately 1200 hours passed over different computers, the theorem has finally been proved. Although the discharging procedure can be checked by hand in a couple of months, it would be impossible to verify the reduction computations in this way. In fact, when the authors of this innovative proof presented it to the *Illinois Journal of Mathematics*, its referees checked the discharging procedure from the complete notes but they had to check the reducibility computations by running an independent computer program. The part of the proof concerning unavoidability did not pose any particular problem, as the proof was only guided by the use of a computer program, but then entirely manually rewritten. On the other hand, the issue concerning reducibility became problematic as the proof was only provided on microfilm, and if one wanted to check it, one would have to use super-computers, which were hardly available at the time.

The use of a computer program in this proof consists of a "simple" verification of a large collection of cases. In particular, the design of the proof and the input for a computer program are still provided by a human. What the computer does it to check whether all possible cases fall under the desired characteristic.

5.5.2 Robertson et al.'s Computer-Verified Proof

Another proof of the Four Color Theorem has been carried out by Neil Robertson, Daniel Sanders, Paul Seymour and Robin Thomas. Initially, their idea was to fully check the entire proof made by Appel and Haken; however checking the computer part would require a lot of programming and even the part that was supposed to be hand-checkable turned out to be extremely complicated. They soon gave up and started to work on their own proof. In particular, they followed the same idea of Appel and Haken.

First, they exhibited a set of configurations; they used Heesch's conjecture[14] (which later was confirmed by the team) to reduce the number of configurations of the set to check for reducibility and unavoidability down to 633 rather than almost 1400. After that, they proved that none of them can occur in a minimal 5-chromatic triangulation i.e. in a minimal counterexample of the "Four Color Theorem" (this step is exactly proving reducibility). Finally, thanks to Birkhoff's lemma which provides that every minimal counterexample to the "Four Color Theorem" is an internally 6-connected triangulation [7], they verified that at least one of the 633 configurations always occurs in a internally 6-connected planar triangulation (namely they proved unavoidability). In such a way a contradiction is given and hence the Four Color Theorem is proved as such a minimal counterexample cannot exist.

[14] A reducible configuration can be found in the second neighborhood of an overcharged vertex.

As in Happel and Haken's proof, the team of mathematicians used the help of computer programs to guide the search for the set of unavoidable configurations. In addition, they were able to supply both the fully human-performed version and the computer-made one so that it would be easier to check. They thus provided two possibilities for the reader: the first, verifiable by hand in a couple of months, and the second, checkable through a computer program in a couple of minutes. In contrast, concerning reducibility, they were "only" able to provide codes to check the correctness of the program by using special computers. This broadened the possibility for a larger number of readers to check the results independently from different devices.

The difference between Robertson's proof and that of Appel and Haken is how they handled unavoidability: Robertson et al. defined a set of 32 discharging rules, despite the one designed by Appel and Haken of about 487 (rules). Furthermore, they replaced the Apple and Haken's gigantic hand-checkable proof of unavoidability by another gigantic one, which however is formally written, therefore can be checked by a computer in a very short time.

5.5.3 Gonthier's Fully Formalized Proof

The ultimate proof of the Four Color Theorem was finally achieved in 2005 by Georges Gonthier, who provided a *formal* proof inspired by that of Robertson et al. Gonthier used the so-called proof assistant *Coq* to succeed in presenting a fully formalized version of the theorem.

Formalizing a proof normally requires much more effort than writing a traditional one, and the result is also much more verbose. On the other hand, a formalized proof leaves no room for human error (such as that committed by Kempe), since the only things to rely on and trust are the correctness of the definitions and axioms used as well as the system employed.

To formalize the very statement of the theorem as well as its proof, Gonthier was forced to resort to more sophisticated tools. Working with graphs embedded in the plane or in the sphere had become more complicated because the properties of these structures did not facilitate such formalization process. It is for this reason that Gonthier shifted the focus of his proof to the so-called "*hypermaps*". This structure allowed a better and "easier" formalization of the problem, thus enabling its combinatorial nature to emerge. It turned out that describing the property of planarity for hypermaps was even easier and that many information were intrinsically encoded. In addition, properties that hold for plane graphs can be carried over to these "new objects".[15]

[15] For example "Euler's formula" (3.1.1) can be rephrased in such a way that it can be applied to hypermaps.

In order to prove reducibility, Gonthier iterated a formalized version of Kempe's argument. Unlike the previous proofs (those of Appel and Haken and Robertson at al.), the mathematician used the fact that four coloring the faces of a cubic map is equivalent to three coloring its edges. This result goes back to Tait in 1880, and the reason why this result was employed is that it better exploits the symmetries that hypermaps provide. However, he still used Robertson's set of 633 reducible configurations.

In conclusion, Gonthier provided formal proof scripts for both the mathematical and computational parts of the proof which were run through the Coq proof checking system to verify their correctness.

Chapter 6

Towards Assisted Proofs

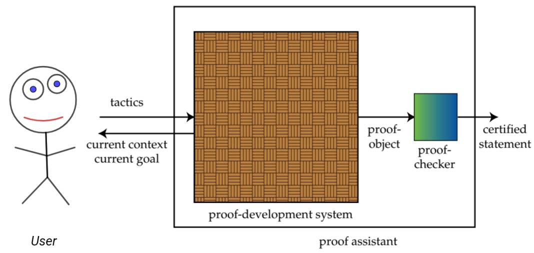
In this last chapter we look at the "Four Color Theorem" under a different prospective. The importance of this theorem does not stand in what it states rather than how the proof has been carried out over the years. We look at the proof of this theorem as the starting point of the increasingly involvement of computers in mathematical proofs.

6.1 Proof Assistants

In order to achieve greater and greater consistency in the demonstrations, many mathematicians and others began to want to verify their proofs using computers. To do that, the main idea is to translate traditional proofs into proofs expressed through a formal language, so that they can be more easily examined and surveyed by a computer. Once this formalization work has been done, the focus becomes the pure verification of each step; this task is perfectly performed by the so-called *proof-checkers*, which are small programs that can be verified by hand by a user. On the other hand, a *proof-development system* is needed by the user in order to make such formalization work more practical and suitable. It is a interactive system that helps the user to develop a proof. A *proof-assistant* consists of a proof-developer system together with a proof-checker. It is an interaction between these two features and the user, that is able to generate a "verified" proof. This interaction can take place at various levels, i.e. proof assistants encompass the whole spectrum between those which purely verify (pure proof-checkers) and those which, given a statement, try to find the proof if it exists (automated theorem provers). Both cases at the extremes are difficult to deal with, since as concern the first case the full formalization of the problem is a difficult task, while in the second one the use of more highly advanced theories becomes necessary. Let us see in more detail how they works.

Let us consider a context Γ and a statement A. First the proof-development system acts, formalizing the given context and the statement A, in order to construct a proof. Rather than creating a proof (as a term) directly, the proof-development system interacts with the user, who provides the so-called *tactics*, which are hints that serve to guide the system towards the creation of the proof-term. This is why it is considered a machine-human interaction, as the user guides the system to create a term through certain patterns.

After that, once the proof-term has been found, the proof-checker takes it as an input and verifies its correctness, scanning through every single formal step, using the formal rules of the logic of the context.



6.1.1 Type Theory: Proofs as Terms and Propositions as Types

Usually, as we have already seen above, mathematical results are written in a rigorous but "informal" way:

$$\text{in a situation } \Gamma \text{ it holds } A$$

$$\text{and a proof } p \text{ of this statement is given}$$

In "every-day mathematics" Γ represents a environment where A holds; they are both given informally as well as the proof p. In logic, both the context Γ and the statement A can be written in a formal language and the notion of provability can be formalized using derivation trees. It is for this reason that it is better to work with logic, which provides an even more rigorous environment and allows mathematical results to be rewritten more feasibly. Thus, the above result can be rewritten as

$$\Gamma \vdash_L A$$

$$\text{proof: } p$$

With Type Theory this aspect improves even further: propositions can be translated as *types* and proofs as *terms*. Again, going back to the previous example, we translate the proposition A (logic language) as a type $[A]$ (type theory) and the proof p as a term $[p]$ (type theory). It then becomes possible to translate informal mathematical results into more formal ones using the language of logic and finally, by using Type Theory, to further improve the rigour and the formality. The correspondence explained above is fully shown below:

$$\text{If } \Gamma \text{ holds, then } p \text{ is a proof of } A \iff \Gamma \vdash A \text{ with proof } p \iff [\Gamma] \vdash [p] : [A]$$

The last item represents the mathematical problem formalized in Type Theory. The most interesting feature of this powerful theory is that it can be rephrase in a even better way, which a proof-checker can take as input

$$\text{Type}_\Gamma([p]) = [A]$$

This means that the proof-checker generates a type of $[p]$ in the context $[\Gamma]$, if it finds one (otherwise it returns false), and it "compares" it to the type $[A]$. A consequence of this correspondence is that "checking the proof" becomes the same as "checking the type". This feature is known as *proof-as-terms*.

Let us go back to the above-mentioned *proposition-as-types* feature, which identifies exactly this relationship between logic propositions and types. In particular, a correspondence is established between logic and function theory. For example, proving the logical implication $A \supset B$ is equivalently translated as a method of obtaining a proof of B from one of A. It forms a function space and such relation can better be seen as $A \to B = \{f \mid \forall a : A.f(a) : B\}$. Analogously, a proof of A & B is identified by a pair $\langle p, q\rangle$ where p is a proof of A and q is a proof of B. It embodies the Cartesian Product, thus $A \times B = \{\langle p, q\rangle \mid p : A \text{ and } q : B\}$. A proof of $A + B$ is translated by the couple $\langle b, p\rangle$, where b is either 0 or 1 and p is a proof of A or B respectively. This is interpreted as the disjoint union $A + B = \{\langle 0, p\rangle \mid p : A\} \cup \{\langle 1, p\rangle \mid p : B\}$.

In conclusion, Type Theory plays a paramount role in formalizing proofs, in order to transfer the problem from the proof-checking to what we called above the type-checking.

6.1.2 Reliability

In this subsection we deal with the problem of reliability concerning proof-checkers. Since they are programs themselves, they need to be verified in order for the output to be considered consistent. Surely, being a good programmer or the fact that the program is often tested are both factors that underlie the confidence we place in the correct running of a proof-checker. But how can we be sure that there are no typos from the user, or that the system responds correctly to what it is supposed to do, or that it does exactly what we intend it to do? In the following, other ways to ensure the reliability of a checked proof are listed.

- **Independent Logic:** As we have already seen above it is extremely convenient to work with logic, since it supplies a method that allows to formalize mathematics. It becomes very useful to supply a system-independent description of the logic, so that we can establish whether our mathematical definitions of objects fully fit our purpose or whether the steps of the proof make sense. Indeed, it provides an independent way to express mathematical features, so that the reliability would not depend only on that of the system. Thus, it becomes extremely important to describe and define a logical system-independent language.

- **Checking the System Itself:** Since the proof assistant is "just" a program, it should be verified itself. First of all, due to the fact that the formalization is independent (see the point above) it constitutes a secure part of the program. It is enough to believe in the formalization that it is obtained independently of the system itself. Next, it is necessary to prove that when the program verifies some statement then it means that a demonstration tree has been found in the logic, verifying that all tactics are consistent and that they are derivable via some inference rules of the logic.

- **Small Kernel:** It is important for the system to have a small kernel, so that the code that rules the system can be "easily" manually checked. In such a way the user needs to have confidence into a small likely trustable kernel only.

- **De Bruijn Criterion:** A proof assistant satisfies the de Bruijn criterion if the proof terms (generated by the system) can be independently checked by a small program, which can be verified manually. In other words, if it can output a proof which is checkable by a much simpler program

6.1.3 The Concept of Mathematical Proof and Some Philosophical Considerations

From a philosophical point of view, mathematics has always been seen as analytical truth, however mathematical error is old as mathematics itself. In order to see the problem from a better perspective, let us see what the mathematical community means by the notion of *proof*.

In mathematics a proof is absolute. It consists in a sequence of steps that, starting from a collection of assumptions, leads to the thesis.

However, most of the proofs we usually deal with are not written in a formal fashion. Traditional mathematical proofs are rigorous but they do not explicit every single detail, thus often they rely on intuitive arguments, omitting basic logical steps.

In mathematics a proof plays two different roles: it is needed to *convince* the reader that a proposition holds and to *explain* why the proof itself works.

As far as concern the first feature, it is basically to verify the correctness of every step that leads to the conclusion. On the other hand, the second point concerns the more difficult question of providing a reason why a process works or where an idea comes from.

Often the first task is totally satisfied by traditional proofs, which, starting with hypotheses, are concerned with verifying the logical steps of an argument. The second task is more complicated and laborious, and it represents a limit for traditional proofs. Proof assistants are systems that make it possible to investigate this second feature as well. It is clear that proof assistants are capable of accomplishing the first task as well, and even in less time.

The reason why the Four Color Theorem was dealt with lies not so much in the result itself as in the method in which it was achieved. Indeed, the importance of this theorem is mainly the idea behind its proof and how it has been reached. In other words, the Four Color Theorem can be seen as the starting point which later became the "gymnasium" for the development and the application of computer programming in mathematical proofs.

Dwelling on this topic in [6] the author divides the involvement of computers in proofs into two categories (which may overlap): *benign* and *pseudo-benign*. Within the first category Appel places all proofs that can be hand-checked by a human in a quite short amount of time. This is the case in which, for example, the computer is trivially used as a calculator, thus when there is the possibility of verifying the result manually. In this case, as well as in general, computer programs are used because of the speed they can provide. In addition, human work is likely to be affected by tiredness and minor distracting errors, which can be more easily avoided by the use of a computer. The second category that Appel discusses is that of the so-called pseudo-benign proofs. This covers all those proofs that can be checkable by hand by a single mathematician in a lifetime work. This is exactly the case of the proof of the Four Color Theorem. It involves so many cases to check that if one wished to spend all the life in surveying all of them, one could. Thus, a pseudo-benign proof would require either the search for a shorter proof or the existence of a human checker of unerring patience, accurate and never tired, nothing more than a computer. Now, it comes clear the reason why the Four Color Theorem plays such an important role: before that, proofs would only be considered correct if they could be understood and checked in their entirety by a human.

Chapter 7

Conclusion

In 1852 Guthrie, probably unaware of what he would begin, provided a new problem to the mathematical world. In the following years this conjecture has been the focus of research and study for several mathematicians. A first result was achieved by Kempe, who, although he made a mistake, provided all his successors with more than a cue to work on. His technique represented the basis of all later attempts to prove the Four Color Theorem. The first real attempt that led to the solution of the problem was that conducted by Appel and Haken who, after working on it for several years and with the help of several mathematicians, succeeded in providing a proof to the conundrum proposed by Guthrie almost a century earlier. The importance of this theorem is not the result itself but how it was achieved. The proof proposed by Appel and Haken consisted of the checking of millions of cases, work that would have been difficult to accomplish by hand. For this reason, for the first time, they involved the use of computer programming in the proof of a theorem. This fact aroused, besides a stir, quite a bit of skepticism within the mathematical community, which could not immediately accept it easily.

After this first attempt there have been further ones including the one by Robertson et al., which took up and partly developed the previous one. In addition, to reduce the number of cases to be checked, they provided a code to verify the computer program involved in the automation of the proof. Even this attempt, was limited in the eyes of the mathematical world since the code provided could not be tested by any computer, undermining the reliability of the proof.

Finally, in 2005, Georges Gonthier and his team succeeded in entirely formalizing the Four Color Theorem and through the proof assistant Coq, solving the more than a century-old problem. Gonthier's proof is based on a formalization of the one previously formulated by Robertson et al. In order to achieve this result Gonthier makes use of hypermaps, which are technical devices that allow the problem to be framed from a more useful point of view for the purpose of solving the problem.

As we have just briefly summarized, this important theorem, although its easy-looking statement, has a long history and it is this on which we wanted to focus on. The Four Color Theorem has played a fundamental role in the history of mathematics, since it turned out to be the "propelling" theorem for the involvement of proof assistants in proofs. This development, as we have already seen, was initially slow and gradual as the mathematical world turned out to be skeptical of these new systems. Initially, computer programming was introduced only for a multiple-case verification task, resulting suspicious in any case at the time. After several corrections to the many attempts to prove the theorem, the idea of using a computer for a mathematical proof became increasingly consolidated among mathematicians, who gradually developed the technique. In 2005 Gonthier succeeded in presenting a proof of the theorem in which the computer no longer were merely used as a case-checker, but rather as a proof-developer.

Being able to accept an entirely computerized proof became the subject of a long discussion among mathematicians and other scientists. Topics as reliability and correctness for a proof became among the most important issues to be resolved.

Proof assistants have now become devices that support and in some cases even guide human effort in proving theorems. The problem of reliability has been tackled by the employment of an independent logic as well as using a small kernel for the system.

In conclusion, the Four Color Theorem is not only proved theorem, but it also launched and promoted the use of proof assistants in the search for proving theorems.

Bibliography

[1] Kenneth Appel and Wolfgag Haken. *Every Planar Map is Four Colorable*. Vol. 67. American Mathematical Society, 1989.

[2] Kenneth Appel and Wolfgag Haken. "Every Planar Map is Four Colorable, Part I: Discharging". In: *Bulletin of the American Mathematical Society* 21 (1977), pp. 429–490.

[3] Kenneth Appel and Wolfgag Haken. "Every Planar Map is Four Colorable, Part II: Reducibility". In: *Bulletin of the American Mathematical Society* 21 (1977), pp. 491–567.

[4] Kenneth Appel and Wolfgang Haken. "The four color proof suffices." In: *The Mathematical Intelligencer* 8 (1986), pp. 10–20.

[5] Kenneth Appel and Wolfgang Haken. "The Solution of the Four-Color-Map Problem". In: *Scientific American* 237.4 (1977), pp. 108–121.

[6] Kenneth I. Appel. "The Use of the Computer in the Proof of the Four Color Theorem". In: *Proceedings of the American Philosophical Society* 128.1 (1984), pp. 35–39.

[7] George D. Birkhoff. "The Reducibility of Maps". In: *American Journal of Mathematics* 35.2 (1913), pp. 115–128.

[8] Leonhard Euler. "Elementa doctrinae solidorum". In: *Novi Commentarii academiae scientiarum Petropolitanae* 4 (1758), pp. 109–140.

[9] Herman Geuvers. "Proof assistants: History, ideas and future". In: *Sadhana* 34 (2009), pp. 3–25.

[10] Frank Harary. *Graph Theory*. Addison Wesley, 1969.

[11] Herman Geuvers Henk Barendregt. "Proof-Assistants Using Dependent Type Systems". In: *Handbook of Automated Reasoning* (2001), pp. 1149–1238.

[12] A. B. Kempe. "On the Geographical Problem of the Four Colours". In: *American Journal of Mathematics* 2 (1879), pp. 193–200.

[13] John Mitchem. "On the History and Solution of the Four-Color Map Problem". In: *The Two-Year College Mathematics Journal* 12.2 (1981), pp. 108–116.

[14] Paul Seymour Neil Robertson Daniel Sanders and Robin Thomas. "The Four-Colour Theorem". In: *Journal of combinatorial theory* Series B.70 (1997), pp. 2–44.

[15] Thomas Tymoczko. "The Four Color Problem and Its Philosophical Significance". In: *The Journal of Philosophy* 76.2 (1979), pp. 57–83.

Acknowledgements

Un primo doveroso ringraziamento va al mio relatore, dott. Samuele Maschio, il quale mi ha seguita e indirizzata verso la stesura di questa tesi. Lo ringrazio in quanto mi ha permesso di concludere il percorso della Magistrale, affiancandomi nello studio e nell'approfondimento di questo argomento. Lo ringrazio per avermi insegnato un metodo da seguire per il (per me difficile) processo di scrittura di un elaborato e per avermi accompagnata in modo costante nella realizzazione di questo lavoro.

In secondo luogo ringrazio la mia famiglia: papà, mamma, Giulia e Beatrice, i quali, ognuno a modo proprio, mi supportano. A volte basta la consapevolezza che qualcuno abbia stima e creda in ciò su cui si investe, non per forza essendo sempre presenti, ma lasciando la certezza che potrebbero esserlo sempre. Questo aspetto è presente in ciascuno di loro, e sebbene mi venga difficile mostrarlo, gliene sono grata e sono consapevole di quanto sia più difficle farlo in queste modalità. Inoltre ringrazio anche tutti i miei familiari: zie e zii, cugini e cugine.

Desidero continuare ringraziando le persone che in questi ultimi anni mi sono state più vicine. Grazie Anni che sei sempre tanto entusiasta e felice, che sei coraggiosa e fan numero uno. Grazie ad "Ave Maria", a le "Squadrate", a "Or Chebello", a "-Roccaiolo", a "Slovenia e una tenda per tre", a "Escili silice liscie" e a tutti i compagni di Università e alle compagne di squadra. Vi ringrazio perche' da sempre, chi più e chi meno, avete condiviso con me questo percorso, e non solo. Vi ringrazio per tutte le vacanze fatte insieme, per le serate e per le tisane.

Grazie a tutti voi, che avete reso questo percorso più piacevole e che siete le persone con cui desidero festeggiare questo traguardo.