Graph Theory

A Development from the 4-Color Problem

By Martin Aigner
Professor at Freie Universität Berlin

With 147 figures, 170 exercises and many examples

Originally published as GRAPHENTHEORIE: Eine Entwicklung aus dem 4-Farben Problem

Translated from the German and edited by L. Boron, C. Christenson and B. Smith, with the collaboration of the author.

BCS Associates, Moscow, Idaho USA

1987

Manufactured in the United States of America.

ISBN: 0-914351-03-6

Library of Congress Catalog Card Number: 86-72291

TABLE OF CONTENTS

TRANSLATORS' PREFACE

We have striven to render a faithful, yet flowing translation of Dr. M. Aigner's beautiful exposition of graph theory. Our effort has been improved immeasurably by the generous and meticulous participation of the author in the editing. All English-readers, worldwide, will surely welcome this volume.

L. F. Boron

C. O. Christenson

B. A. Smith

FOREWORD

During recent years I have been asked with increasing frequence by students why an area of mathematics developed in a particular way (mostly as presented in the lectures) and not otherwise, what the principle motivation was, and how the development proceeds further. Among other factors such as applicability, or connections with other areas, of particular interest was the role which the famous classical problems play in the development of a theory.

The recent successful and novel solution of the 4-color problem was a welcome occasion for me to study the precise influence of this universally known problem on mathematics, chiefly on graph theory. Perhaps more sharply than elsewhere, opinions differ on the 4-color problem. Some say that the mathematics influenced by the 4-color problem is marginal and the solution with its enormous computer support is just plain frightful from an esthetic standpoint. On the other hand many are of the opinion that the 4-color problem almost alone permitted an entire discipline, graph theory, to arise as rarely occurs to this extent, and that the solution with its varied aspects, both intrinsically mathematical and extra-mathematical, points far into the future. Work on this book convinced me that it is rather the second interpretation that proves right—and it is my hope that I have succeeded in my presentation to give sufficient arguments for this.

The present book will be two things: First, an introduction to graph theory, which contains nearly all the important concepts and results, and secondly, a presentation of the role that the 4-color problem played in the develpment of graph theory. In order to accomplish this, the book is split into 3 parts: Part I (Introduction), Part II (Theme), and Part III (Finale).

Part I describes the origin of the 4-color problem, the early attempts at a solution, and the first theoretical groundwork that was developed for the solution. Great importance is placed on the presentation of the difficulties that the researchers of the early period had, and the dead ends that they ran into. This is presented not so much to satisfy historical curiosity, but rather to point out to the reader how one proceeds using these difficulties in ones own mathematical work. Part II is the "meat" of the book. Proceeding from the theortical contributions of the early period, we develop 5 essential chapters of graph theory, that in their totality present the above-mentioned introduction to graph theory. Finally, Part III brings a presentation of the final solution of the 4-color problem and a discussion of the questions arising from the solution. Since it is a question here of a gigantic abundance of individual results, the description must of necessity come out sketchly. It is my hope that nevertheless the fundamental ideas appear clearly. Although the book is not structured chronologically for reasons of readability, the dates can be outlined approximately as follows: Part I comprises the time 1850–1930, Part II the span 1930–1965 and Part III the years since 1965.

So much for the contents. As was already pointed out, the middle part, after an explanation of the fundamental concepts, can be used as a 1-semester introduction to graph theory. A number of exercises follow each chapter. Exercises whose working out is particularly recommended are designated by a °, and more difficult problems with a *. At the end of the book, the reader will find a bibleography by chapter with references to the accompanying lectures.

My special thanks go to my colleagues T. Andrease and R. -H. Schulz who have read and essentially improved the various chapters at all phases of creation. Furthermore, I thank Mrs. Barrett of the II. Mathematischen Institut of the Freie Universität Berlin for the careful preparation of the manuscript and Teubner Verlag for the pleasant collaboration.

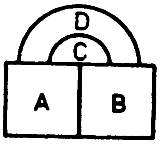
Martin Aigner

Berlin, October 1983

PART I: INTRODUCTION

1. ORIGIN AND "SOLUTION"

On October 23, 1852 Augustus De Morgan, professor at University College in London, wrote to his colleague Sir William Hamilton: "A student of mine asked me today to give him a reason for a fact which I did not know was a fact—and do not yet. He says that if a figure be anyhow divided and the compartments differently coloured so that figures with any portion of common boundary *line* are differently coloured—four colours may be wanted, but not more." At the same time De Morgan gave an example in which 4 colors were necessary (Figure 1.1), and added: "Now it does seem that drawing three compartments with common boundary ABC two and two you cannot make a fourth take boundary from all, except by inclosing one. But it is tricky work, and I am not sure of the convolutions."



A,B,C,D, are colors

Figure 1.1

Well, it was indeed not easy and convolutions arose, and many still arise—it is exactly on those that this book will report. Incidentally: Hamilton was not in the least interested, although the **4-color conjecture**, as it was soon called, was quickly well known. The student of whom De Morgan spoke, was Frederick Guthrie, who later let it be known, that the original proposer of the problem was his brother Francis.

Let us take a closer look at the 4-color problem! By "adjacent" countries, Francis Guthrie must have meant those which have a line in common and not just a corner (or even finitely many). Otherwise one already needs 7 colors for the map in Figure 1.2.

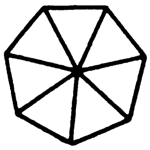


Figure 1.2

By a "country" one must understand a connected region. For if a country were permitted to have two components, then with little effort one could devise a map with 5 countries, each two of which are neighboring, so that 5 colors would be necessary (Figure 1.3).

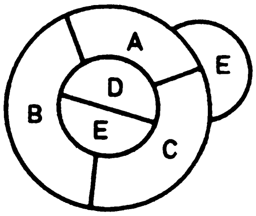


Figure 1.3

De Morgan obtained a proof (this is the first concrete contribution towards the solution of the 4-color problem) that it is impossible to give 5 countries, any two of which are adjacent (this was already asked in the form of a puzzle by Möbius around 1840). However, this says nothing about the correctness of the 4-color conjecture. It is by no means true that the maximal number of pairwise adjacent countries is equal to the number of necessary colors. (This incorrect assumption was the starting point of many false proposed solutions.) As an example, the map in Figure 1.2 does not have 3 pairwise adjacent countries. However, as one immediately sees, 2 colors do not suffice.

After De Morgan, it was Cayley more than anyone else, who contributed to publicizing the problem. At an 1878 session of the London Mathematical Society he presented the problem and discussed the difficulties that arise. Within a year A. Kempe, a lawyer and an associate of the Mathematics Society, published a treatise, in which the 4-color conjecture appeared to be answered affirmatively. The solution was received with enthusiasm. Kempe was elected a Fellow of the Royal Society, further improvements appeared, and the "4-color theorem" was a generally recognized mathematical fact.

Although Kempe's proof contained an error, which was first pointed out by Heawood in 1890 (to whom this was almost painful), it brought together nearly all the ideas which led to the definitive solution 100 years later. We will do well then, to take a closer look at his argument. For this, however, we must clarify what we understand by a map.

A **graph** G consists of a finite set V of **vertices**, a finite set E of **edges**, and a relation, which assigns to each edge e exactly two (different or the same) vertices u and v which we call the **endpoints** of e. Usually the vertices u and v are different; if $u = v$, then we call e a **loop** at u. If e has endpoints u and v, then we say that e **joins** u and v.

We say that vertices are **adjacent**, if they are joined by an edge. If $v \in V$, we denote all the vertices that are neighbors of v by $N(v)$. If v is an endpoint of e, then v and e are called

incident, and we often denote this concept, somewhat incorrectly, by $v \in e$. A **vertex** that is not incident with any edge, is called **isolated**. We call two **edges** incident if they have a common endpoint. If two vertices are joined by the edges $e_1, e_2, \ldots, e_t$, $t \geq 2$, then we call $e_1, \ldots, e_t$ **multiple edges**. If $e \in E$ has endpoints u and v, then for convenience we use the notation $e = uv$ throughout. Naturally, this notation is ambiguous if u and v are joined by several edges, but this will never lead to difficulties. For many problems, loops and multiple edges are irrelevant, so that a special name was coined for graphs without loops and multiple edges, **simple graphs**. For simple graphs, the above notation $e = uv$ is well defined.

One can best illustrate a graph $G(V, E)$ by a diagram, in which one designates the vertices as points of the plane and the edges as Jordan curves, i.e. continuous curves between the endpoints that do not cross themselves (which naturally in the case of a loop are closed). See Figure 1.4, in which the edges are given by their endpoints:

EXAMPLE. $V = \{1,2,3,4,5,6\}$, $E = \{11,12,25,34,36,46,46\}$.

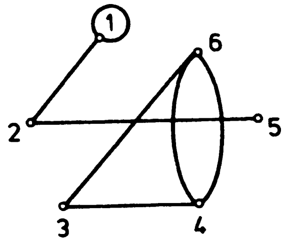


Figure 1.4

If $v_0, v_1, v_2, \ldots, v_t$ are vertices with $v_{i-1}v_i \in E$ for all i, then we call $W = (v_0, v_1, \ldots, v_t)$ a **trail** from v_0 to v_t. The **length**, $\ell(W) = t$, of the trail is the number of edges contained in it. W is called a **path** if all vertices v_i are distinct. It is clear that each trail from u to v contains a uv-path. A trail $W = (v_0, v_1, \ldots, v_t)$ in which all the edges are distinct, as well as all the vertices except $v_0 = v_t$, is called a **circuit**. Again t is the **length of the circuit**. For example, the graph in Figure 1.4 contains a circuit of length 3, namely $(3,4,6,3)$. The loops are precisely the circuits of length one. Circuits of length 2, as $(4,6,4)$ in Figure 1.4, are generated by multiple edges between two vertices.

A graph is called **connected** if any two vertices are joined by a path, otherwise the graph is called **disconnected**. Since the accessibility of vertices by paths is clearly an equivalence relation on V, each graph decomposes into its connected **components**. For example, the graph in Figure 1.4 decomposes into the components $\{1,2,5\}$ and $\{3,4,6\}$.

We have already mentioned that we usually illustrate a graph G by a point-line-system in the plane. If there exists a representation in which the edges of G only meet at vertices and never in between, then we say that the abstract graph G is **planar**, and we call the graph embedded in the plane a **plane graph**. Thus: A plane graph consists of a set of points of the plane and a set of edges (= Jordan curves), which intersect only at the corresponding endpoints. Remark: Isolated vertices play virtually no role in embeddings, they can be put anywhere, so long as no edge runs through an isolated point.

The distinction between "planar" and "plane" appears to be artificial at first sight, but we will see later that a planar graph can in general be embedded in several (nonisomorphic) ways into the plane.

EXAMPLE. *The representation of the graph in Figure 1.4 is not a plane representation, since e.g. the lines 25 and 36 intersect in an interior point. But the graph is clearly planar; we only need to pull the components apart (Figure 1.5).*

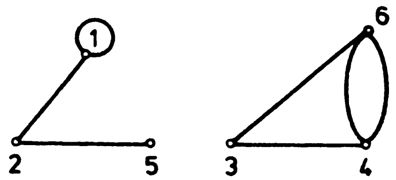


Figure 1.5

If $G(V, E)$ is a plane graph, then the lines of the graph decompose the plane into finitely many connected regions (of which exactly one, the "outer" region is unbounded), which we will call the countries R of G. The basis for that is the famous **Jordan Curve Theorem**: A closed Jordan curve C decomposes the plane into two connected regions, of which exactly one is unbounded. That is, two points of the plane can be joined by a Jordan curve that does not meet C, precisely when the points either both lie inside, or both lie outside of C (Figure 1.6). As self-evident as this theorem is, its proof is anything but easy (see e.g. the book of Aleksandrov). We will make extensive use of this fundamental theorem as we proceed.

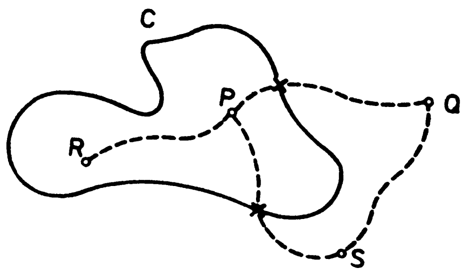


Figure 1.6

Conversely, given a "map" $\mathcal{L}$, then we can associate a plane graph $G(\mathcal{L})$ with $\mathcal{L}$. The **vertices** of $G(\mathcal{L})$ are the intersection points of distinct boundaries, and the **edges** are the boundary lines running between intersection points. (An exception is caused by islands, which are bounded by only one country. Here one must pick an arbitrary point u on the boundary, thus creating a loop at u.) It is clear that the countries of $G(\mathcal{L})$ are exactly the original countries of the map $\mathcal{L}$, whereby we also are taking into consideration the "outer" country. We call $G(\mathcal{L})$ the **skeleton** of $\mathcal{L}$.

EXAMPLE.

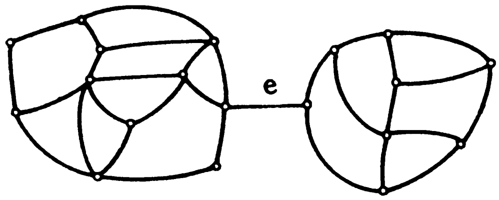


Figure 1.7

The skeleton of the map pictured in Figure 1.7 contains 19 vertices, 31 edges, and 14 countries (including the outer country). With this we come to the fundamental definition.

DEFINITION. *A* **map** $\mathcal{L}(V,E,R)$ *is a plane graph* $G(V,E)$ *together with the countries* R *that correspond to* G.

Countries that have a common border are called **neighboring**. If the edge e is on the boundary of the country F, then we say that e and F are **incident** and often denote this with $e \in F$; analogously, we speak of vertex-country incidence. We will now clarify what we mean by isomorphic graphs, resp., maps.

DEFINITION. *Two graphs* $G(V,E)$ *and* $G'(V',E')$ *are called* **isomorphic**, *denoted* $G \cong G'$, *if there are bijections* $\varphi_V : V \to V'$, $\varphi_E : E \to E'$ *such that* $v \in V$ *is an endpoint of* $e \in E$ *precisely when* $\varphi_V(v)$ *is an endpoint of* $\varphi_E(e)$. *Thus the vertex-edge-incidence is exactly preserved. Two maps* $\mathcal{L}(V,E,R)$ *and* $\mathcal{L}'(V',E',R')$ *are called* **isomorphic**, $\mathcal{L} \cong \mathcal{L}'$, *if there are bijections* $\varphi_V : V \to V'$, $\varphi_E : E \to E'$, $\varphi_R : R \to R'$, *which exactly preserve the vertex-edge-, edge-country-, and vertex-country-incidence. In particular, then* $G(\mathcal{L}) \cong G(\mathcal{L}')$.

From a combinatorial point of view, isomorphic graphs, resp., isomorphic maps are completely the same structurally.

From now on we will consider the outer country to be just like the interior countries. By stereographic projection one sees immediately that any country F can be specified as "outer" (just choose the pole in F); so there is really no distinction between the notions of outer and inner.

A **coloring** of $\mathcal{L}(V, E, R)$ is a mapping $f : R \to M$ from R into an arbitrary (**color-**) set M, so that countries which have a common boundary edge, always correspond to different colors. The smallest number of colors which are needed for that is called the **chromatic number** $\chi(\mathcal{L})$ of $\mathcal{L}$. If $\chi(\mathcal{L}) = n$, then we say that $\mathcal{L}$ is **n-chromatic**.

With these concepts, we can now give an exact formulation of the 4-color conjecture.

4-COLOR CONJECTURE. *Every map $\mathcal{L}$ has chromatic number $\chi(\mathcal{L}) \leq 4$.*

Back to Kempe. His first idea was to single out several special types of maps, and then to show that the 4-colorability of these "normal" maps implies the 4-colorability of all maps. For example, we can assume that the **skeleton** $G(\mathcal{L})$ is a **connected** graph. If $G(\mathcal{L})$ is not connected then by introducing additional edges we can construct a map $\mathcal{L}'$ whose skeleton is connected. If $\mathcal{L}'$ is 4-colorable, then so is $\mathcal{L}$, by simply removing the added edges. (The connecting edges are denoted by dashed lines in Figure 1.8.)

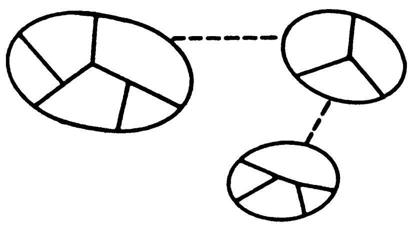


Figure 1.8

We can also omit loops, since the corresponding country is adjacent to only *one* other country. Thus after coloring the remainder, one of the 3 available colors can be used. Furthermore, we can assume that $G(\mathcal{L})$, resp., $\mathcal{L}$ contains no bridges: An edge e of an arbitrary connected graph G is called a **bridge** if G has two components after e is removed. Equivalently: e is a bridge if e does not lie on a circuit (proof?). In a plane graph, bridges are also characterized by the fact that they bound the same country on both sides, as can be seen by again using the Jordan Curve Theorem (consider the bridge e in Figure 1.7). Now if $\mathcal{L}$ contains a bridge e, then we shrink e to a point, i.e. we identify the two endpoints. If the resulting map $\mathcal{L}'$ is 4-colorable, then clearly so is $\mathcal{L}$. In particular, after the removal of all bridges we see that every vertex is incident to at least two edges. Finally, we can omit all vertices that are only incident to two edges by simply fusing the edges (Figure 1.9). This operation clearly makes no difference on the number of colors needed.

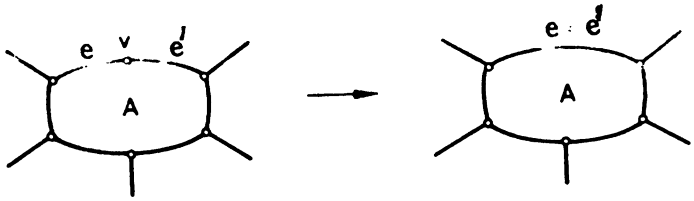


Figure 1.9

DEFINITION. *A map $\mathcal{L}$ is called* **normal** *if the skeleton graph $G(\mathcal{L})$ is connected, contains no loops or bridges, and each vertex is incident with at least 3 edges.*

As a consequence of our observations thus far we note:

1.1 THEOREM. *If the 4-color conjecture is correct for all normal maps, it holds in general.*

We keep in mind that in a normal map each edge has two distinct endpoints (no loops!) and bounds two distinct countries (no bridges!).

Now, what was Kempe's idea for a proof? He used induction on the number of countries. In the *first step* he showed that there must be a country F that is adjacent to at most 5 other countries $A_1, \ldots, A_t$, $t \leq 5$. He removed F by extending the boundaries (Figure 1.10), thus creating a new map $\mathcal{L}'$ with one less country. He now proved in the *second step* that the 4-coloring of $\mathcal{L}'$ that exists by the induction assumption can be so altered that $A_1, \ldots, A_t$ receive at most 3 different colors. The free 4th color can now be used upon the reintroduction of F, and the proof is complete.

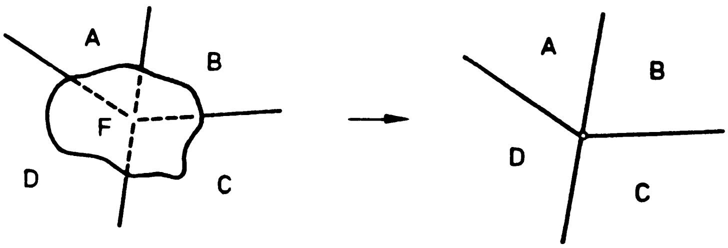


Figure 1.10

Kempe's first step is a consequence of Euler's polyhedral formula, which we will discuss first. But before that, we need a few definitions. Let v be a vertex of the graph $G(V, E)$. The **degree** $d(v)$ is the number of edges which have v as an endpoint, wherein loops at v are counted twice. We begin with a result that was proven by Euler in 1736. Quite likely this was the earliest contribution to graph theory proper.

1.2 LEMMA. *For a graph $G(V,E)$,*

$$2|E| = \sum_{v \in V} d(v).$$

If in particular $d(v) \geq 3$ for all $v \in V$, then we have $2|E| \geq 3|V|$.

Proof. We count the pairs (v,e) where $v \in V$, $e \in E$ and $v \in e$. Since each edge has exactly 2 endpoints (loops are counted twice!), the number on the one hand is $2|E|$. On the other hand, each vertex $v \in V$ contributes $d(v)$, thus the sum is also equal to $\sum_{v \in V} d(v)$. ▪

A connected graph without circuits is called a **tree**. The trees are one of the few classes of graphs that were studied prior to the 4-color problem, primarily by Kirchhoff in connection with electrical networks and by Cayley who initiated the enumeration of trees. Figure 1.11 depicts all trees with at most 5 vertices.

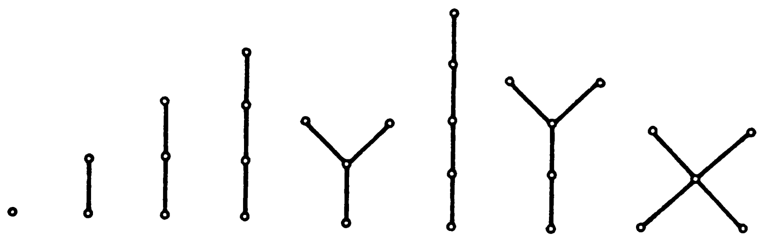


Figure 1.11

1.3 LEMMA. *For a tree $G(V,E)$, we have $|E| = |V| - 1$.*

Proof. By induction on $|V|$. For $|V| = 1$ there is nothing to prove. Now let $|V| \geq 2$ and $e \in E$. Since G contains no circuit, e must be a bridge. The graph $G' = G - e$ which arises on removal of e has two components $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ which are clearly again trees with $|V_1| < |V|$ and $|V_2| < |V|$. By the induction hypothesis we obtain $|E_i| = |V_i| - 1$, $i = 1,2$. Hence, $|E| = |E_1| + |E_2| + 1 = (|V_1| - 1) + (|V_2| - 1) + 1 = |V| - 1$. ▪

1.4 EULER'S FORMULA. *Let $G(V,E)$ be a plane connected graph with the countries R. Then:*

$$|V| - |E| + |R| = 2.$$

Proof. We use induction on the number of edges. For $|E| = 0$, G consists of a single vertex, so that $|V| = |R| = 1$. Thus $|V| - |E| + |R| = 2$. Now let $|E| \geq 1$. Assume that $e \in E$ is not a bridge. Then $G' = G - e$ is still connected, and since by the removal of e, the two countries it bounds will meld, we obtain $|V'| = |V|$, $|E'| = |E| - 1$, and $|R'| = |R| - 1$. By induction, the formula also holds for G. If on the other hand G has only bridges, then G is clearly a tree. Then however $|R| = 1$, and the formula follows from 1.3. ▪

We now have all the tools (1.2 and 1.4 had, of course, long been known) which are needed to perform Kempe's first step.

1.5 THEOREM. *Let $\mathcal{L}$ be a normal map. Let us denote the number of countries which have exactly i boundaries by p_i, $i \geq 2$. Then*

$$\sum_{i\geq 2}(6-i)p_i \geq 12.$$

In particular, one of the numbers p_2, p_3, p_4, p_5 must be positive, i.e. there must exist a country that has at most 5 boundaries.

Proof. Each country has at least 2 boundaries (no loops!), thus we have $|R| = p_2 + p_3 + p_4 + \ldots$. Since each edge bounds exactly 2 countries (no bridges!), we obtain with the same counting method as in 1.2:

$$(*) \qquad 2|E| = 2p_2 + 3p_3 + 4p_4 + \ldots.$$

In a normal map each vertex has degree ≥ 3, therefore $3|V| \leq 2|E|$ by 1.2. We now multiply the Euler Formula by 6 and use $(*)$ to obtain

$$\begin{aligned} 12 &= 6|V| - 6|E| + 6|R| = (6|V| - 4|E|) + (6|R| - 2|E|) \\ &\leq 6|R| - 2|E| = 6(p_2 + p_3 + p_4 + \ldots) - (2p_2 + 3p_3 + 4p_4 + \ldots) \\ &= \sum_{i\geq 2}(6-i)p_i. \end{aligned}$$

Since the sum is at least 12 and $(6-i)p_i \leq 0$ for $i \geq 6$, it follows that one of the numbers p_2 through p_5 must be positive. ∎

EXAMPLE. *We can now answer the previously mentioned puzzle of Möbius, whether a map $\mathcal{L}$ exists in which each of 5 countries borders on each of the others. In such a map $\mathcal{L}$ we would have $p_4 = 5$ and $p_i = 0$ for $i \neq 4$, implying $12 \leq 2p_4 = 10$. Thus no such map exists.*

Let us now analyze the second step of Kempe's proof. Suppose the 4-color conjecture has been proven for maps with at most $r-1$ countries, and that $\mathcal{L}$ is a normal map with $|R| = r$. We know that $p_2+p_3+p_4+p_5 > 0$. If there is a country F with at most 3 boundaries, then we can certainly draw Kempe's conclusion, by using the 4th color for F. The first interesting case arises then when $p_2 = p_3 = 0$, but a country F exists with 4 boundaries. Let the bordering countries be A, B, C, and D. We color the map $\mathcal{L}'$, arising from the contraction of F, with 4 colors (see Figure 1.10). If we do not use all four colors, we are again finished.

Thus assume that A was colored **red**, B **blue**, C **green**, and D **white**. We consider A and C. Either there is no chain of neighboring countries from A to C which are alternately colored red and green, or there is such a red-green-chain from A to C. In the latter case, there can not be a blue-white-chain from B to D. For two such chains must meet in a common country, which is impossible. In Figure 1.12 a red-green-chain runs from A to C in a clockwise direction.

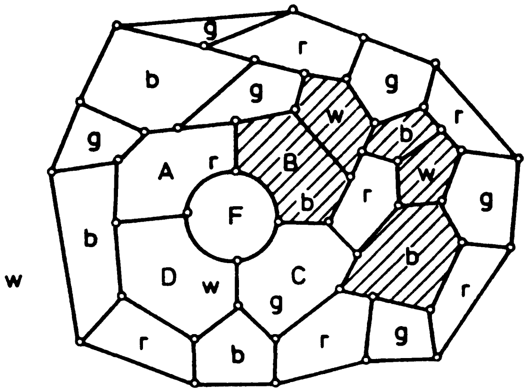


Figure 1.12

Thus: Either there is no red-green-chain from A to C or there is no blue-white-chain from B to D. And now Kempe's conclusion: Let us assume without loss of generality that there is no blue-white-chain from B to D. We set up a list of all countries which can be reached by a blue-white-chain from B. (In Figure 1.12 these countries are shaded.) If we exchange the colors blue and white in this list, there results again an admissible coloring of $\mathcal{L}'$. Since D is clearly not in this list, we have colored $\mathcal{L}'$ so that around F a color is left out (in our example, blue) which now can be applied to F.

This clever idea using color chains (which, as we will see, formed the basis for many other attempted proofs, and are today called **Kempe chains**) was also applied by Kempe to the remaining case $p_2 = p_3 = p_4 = 0$, and $p_5 > 0$. Here, however, he made a mistake which was first pointed out 10 years later. His method of proof will now be recounted. The reader may himself find the weak point (or wait until the next chapter).

Suppose then that F is adjacent to 5 countries A, B, C, D, and E. If the contracted map $\mathcal{L}'$ has a 4-coloring in which at most 3 colors are used in the adjacent countries, then we are through. If all 4 colors are needed, then Figure 1.13 displays the situation with no loss of generality.

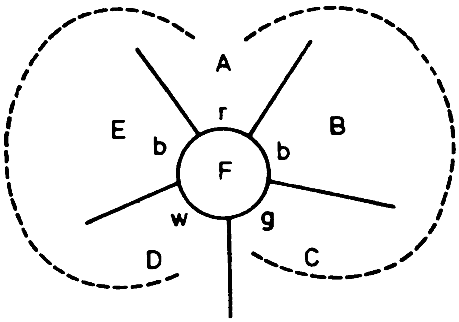


Figure 1.13

The following terminology is convenient: By the blue-white component of B we understand all countries which can be reached from B by a blue-white chain; and analogously for the other colors. If no red-green chain exists from A to C, then as above we can spare one color by interchanging red and green. The same holds if there is no red-white chain from A to D. We assume then that A and C are joined by a red-green-chain, and A and D by a red-white chain (indicated by the two arcs in Figure 1.13). Then B is separated from D, and E from C. This means, however, that the blue-white component, L_1, to which B belongs can not contain D or E. Similarly, the blue-green component L_2 which contains E can not also contain C or B. Thus in L_1 we can interchange blue and white, and in L_2 the colors blue and green, without changing the coloring of C or D. After this recoloring is finished, A is colored red, B and D white, and C and E green. Thus blue remains for use on the contracted country F.

While Kempe's argument was faulty, his proof nevertheless contained two decisive ideas: By 1.5, each normal map must contain a country with at most 5 neighboring countries. We can thus say that the set of configurations in Figure 1.14 is **unavoidable** in the sense that each normal map *must* contain a configuration from this set.

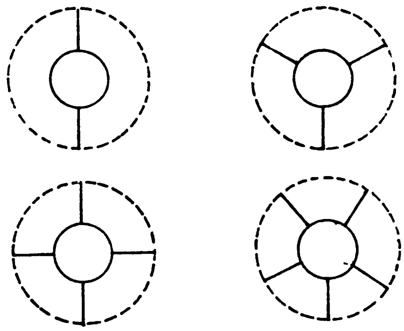


Figure 1.14

The induction step can be thought of as the search for a minimal 5-chromatic map, i.e. one with the minimal number of countries. A configuration is called **reducible** if it can be shown that it is not possible for it to appear in a minimal 5-chromatic map. Kempe's chain method shows that the first 3 configurations of Figure 1.14 are reducible. Unfortunately, it is not conclusive for the last. However the program is clear (and has finally proved successful): *Find an unavoidable set whose configurations are all reducible.* If such a set is found, then no minimal 5-chromatic map, and thus no 5-chromatic map, can exist. That is, all maps can be colored with at most 4 colors.

As already mentioned, Kempe's proof was generally accepted. Thus it is not surprising that further invalid improvements as well as new invalid proofs followed. One of these, which was given by P. G. Tait, merits careful study.

Cayley and Kempe had already noted that for the proof of the 4-color conjecture the class of normal maps can be reduced still further. We call a map $\mathcal{L}(V, E, R)$, respectively, the skeleton $G(\mathcal{L})$, **cubic** if $d(v) = 3$ for all $v \in V$. Now if $\mathcal{L}$ is an arbitrary normal map, then to $\mathcal{L}$ we can associate a cubic map $\mathcal{L}'$ which needs at least as many colors.

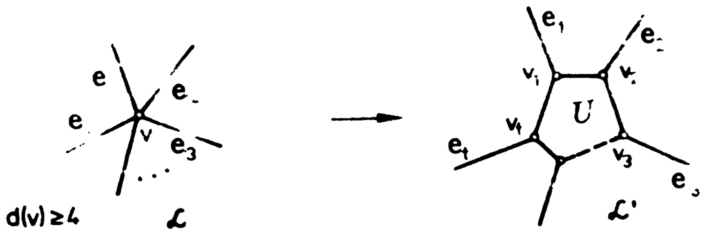


Figure 1.15

We replace each vertex $v \in V$ with $d(v) \geq 4$ by a country U, as indicated in Figure 1.15. If the cubic map $\mathcal{L}'$ constructed in this manner is 4-colorable, then so is $\mathcal{L}$. Thus we see: *The 4-color conjecture is true if all normal cubic maps are 4-colorable.*

Observe that for a cubic map, a loop must be adjacent to a bridge. That is, a cubic map without bridges has no loops. Since for a cubic map $3|V| = 2|E|$ holds, we have equality in 1.5, as follows immediately from the proof given there.

1.6 COROLLARY. *If $\mathcal{L}$ is a normal cubic map and p_i is the number of countries which have exactly i boundaries, $i \geq 2$, then*

$$\sum_{i \geq 2} (6 - i) p_i = 12.$$

Tait now reduced the 4-colorability of cubic maps to the 3-colorability of the edges. By an **edge coloring** we understand a coloring of the edges such that incident edges receive different colors.

1.7 THEOREM (Tait). *A cubic map $\mathcal{L}$ without bridges is 4-colorable if and only if its edges are 3-colorable.*

Proof. Let $\mathcal{L}$ be 4-colorable. We take Klein's 4-group $\{0, a_1, a_2, a_3\}$ as the coloring set, where 0 is the neutral element, and for all i, $a_i + a_i = 0$, $a_i + a_j = a_k$ with $k \neq i, j$ for all $i \neq j$. Now if the 4-coloring is given, then we take as the color of an edge e, the sum of the colors of the countries incident with e. From that it follows immediately that the edges are all colored by $\{a_1, a_2, a_3\}$ and that incident edges receive different colors.

Conversely, we now assume that the edges are 3-colored using $\{a_1, a_2, a_3\}$. We pick a country F_0 and color it with the color 0. Every other country will now be colored in the following way: Let C be an arbitrary Jordan curve that joins the interior of F_0 with the interior of F and does not go through a vertex of $\mathcal{L}$. Then let the color of F be the sum of the colors of all the edges crossed by C. We must first show that each such curve yields the same color, or equivalently, that the sum of the colors of all edges crossed by a closed Jordan curve (which misses vertices) is 0.

Let S be such a curve, $c_1, c_2, \ldots, c_t$ the colors of the edges crossed by S (if edges are crossed several times, insert them into the list every time they are crossed), and $d_1, \ldots, d_n$ the colors of the edges lying entirely in the interior of S. If $c(v)$ is the sum of the 3 edges incident to the vertex v, then we always have $c(v) = a_1 + a_2 + a_3 = 0$. Thus also $\sum c(v) = 0$ when we sum over all the vertices lying in the interior of S. In $\sum c(v)$ each inner edge is counted twice. Thus its value is $2(d_1 + \ldots + d_n) = 0$. Consider an edge that is crossed by S. If it is crossed an odd number of times then one endpoint lies in the interior and the other in the exterior of S. If it is crossed an even number of times, then two cases are possible (both endpoints in the interior, or both in the exterior). Since $a_i + a_i = 0$ for all i, we conclude that both the expressions $\sum c(v)$, resp., $c_1 + \ldots + c_t$, give precisely the sum of the colors of the edges which lead from the interior to the exterior. Therefore we have $\sum c(v) = c_1 + \ldots + c_t$, and thus $\sum_{i=1}^{t} c_i = 0$. Finally, we must still prove that adjacent countries A and B have different colors. If C is a curve from the interior of F_0 to the interior of A, then we can extend C over the boundary e into the interior of B. The color $c(B)$ of B is then the sum of the colors $c(A)$ and $c(e)$ of e and thus $c(B) \neq c(A)$ since $c(e) \neq 0$. ∎

EXAMPLE. *Figure 1.16 illustrates the second part of the proof, whereby for clarity a_i is replaced by i, thus* $1 + 1 = 0$, $1 + 2 = 3$, *etc.*

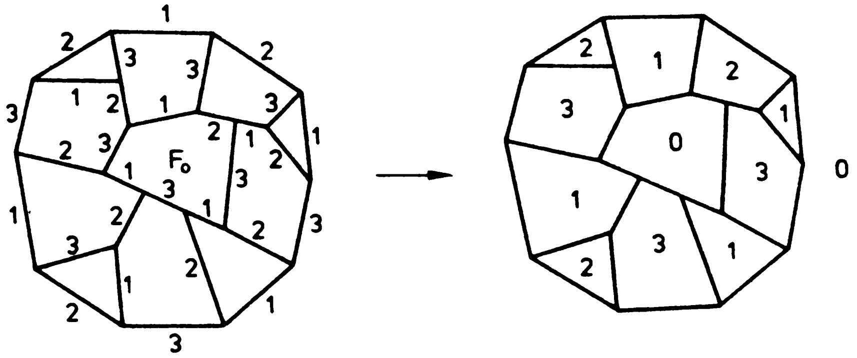


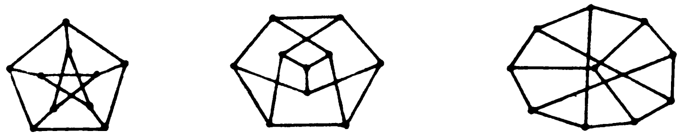
Figure 1.16

Tait assumed the 3-coloring of the edges of a cubic map to be an "elementary" theorem, easily proven by induction, which by 1.7 would then imply the 4-color conjecture. However, this theorem is in no way elementary, it is exactly as difficult as the 4-color theorem. Heawood made this clear in 1890, and we now turn to this new episode of the 4-color problem.

EXERCISES FOR CHAPTER 1

1°. The n-dimensional **cube graph** Q_n has as vertices all n-tuples $(a_1, a_2, \ldots, a_n)$ with $a_i = 0$ or 1, where two vertices are adjacent if they differ as n-tuples in exactly one place. Draw the graphs Q_1 through Q_4. How many vertices and edges does Q_n have? What is the degree of each vertex in Q_n? Which Q_n's are planar?

2. Which of the 3 graphs pictured are isomorphic?



3°. Show the equivalence: An edge e is a bridge $\Leftrightarrow$ e does not lie on a circuit. If the graph is plane, then furthermore: e is a bridge $\Leftrightarrow$ e is bounded by the same country on both sides.

4. Let G be a simple graph with at least 2 vertices. Show that G always contains two vertices with the same degree.

5. Show that the number of vertices of odd degree is always even.

6. Draw all nonisomorphic trees with 6 and 7 vertices. (There are 17.)

7°. Prove the following equivalence: $G(V, E)$ is a tree $\Leftrightarrow$ G is connected with $|E| = |V| - 1$ $\Leftrightarrow$ each two vertices are joined by exactly one path.

8. Show: Every tree with $|V| \geq 2$ has at least 2 vertices of degree 1. Which trees have exactly 2 vertices of degree 1? Which have exactly 3 vertices of degree 1?

9. Let G be a simple graph with p vertices and q edges. Show: If $q > \binom{p-1}{2}$, then G is connected.

10. Draw all cubic graphs (plane or not) up to 6 vertices.

11°. Assume $\mathcal{L}(V, E, R)$ is a map without bridges or loops in which each vertex has degree exactly d, $d \geq 2$, and each country has exactly f boundaries, $f \geq 2$. Using 1.4 determine which pairs (d, f) are possible and construct a corresponding map for each possible pair.

12°. Prove another theorem of Tait: A normal map is 2-colorable precisely when the degree of each vertex is even.

13. Using induction, conclude from the fact that each normal map $\mathcal{L}$ has a country with at most 5 boundaries (1.5), that $\chi(\mathcal{L}) \leq 6$ holds for all maps.

14°. Where did Kempe err?

15. Find a 3-edge-coloring of the map in Figure 1.12.

16. Puzzle: 9 people stand together at a reception. Some are known to one another, others are not (acquaintance will always be assumed as two sided). Assertion: Either there are 4 people each of which is acquainted with each of the others, or there are 3 people of which none knows any of the others.

17. Computer exercise: Suggest a data structure which represents a graph economically for input and output.

18. Computer exercise: Describe an algorithm which determines whether 2 vertices lie in the same component.

19. Computer exercise: Develop a program for generating all trees in a given graph.

20*. Somewhat harder: Given p labeled vertices $v_1, \ldots, v_p$, we want to determine the number of *different* trees G with $V = \{v_1, \ldots, v_p\}$. Two trees are called the same in this context, if they have exactly the same edges. For $p = 3$, there are 3 distinct trees (which are all isomorphic as graphs). Show Cayley's result: On p labeled vertices there are exactly p^{p-2} different trees. (Hint: Induction.)

2. ERRORS AND HOPES

In a work that appeared in 1890, P. J. Heawood analyzed Kempe's proof of the 4-color conjecture and laid bare a fatal defect. Let us turn back the pages to the last remaining case: A country F is surrounded by 5 countries $A, B, \ldots, E$, which are colored with the colors red, blue, green, white and blue. We assume that a red-green-chain exists from A to C and a red-white-chain exists from A to D. Thereby we can interchange the colors in the blue-white component that contains B, and also in the blue-green component which contains E. While it is true that *each* of these interchanges again gives an admissible coloring (whereby, however, all 4 colors still appear around F), it may happen after the completion of *both* interchanges that two blue countries now touch each other (which previously were colored white, resp., green), which is inadmissible. The map in Figure 2.1 demonstrates this fact.

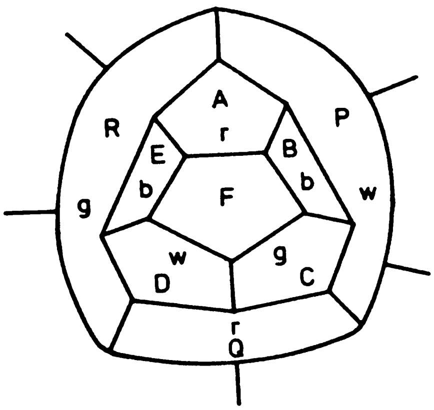


Figure 2.1

Since P belongs to the blue-white component of B, P obtains the new color blue, and since R belongs to the blue-green component of E, R also obtains the color blue—however, P and R have a common border!

Thus, the proof is incomplete—and Kempe immediately acknowledged his error. Heawood's work initially was little noted—he himself had perceived it as somewhat destructive. Completely untrue! For in it are contained two fundamental theorems which had a decided influence on the subsequent research.

First, Heawood showed that using Kempe chains, not the 4-colorability, to be sure, but the 5-colorability of all maps can be ascertained.

2.1 5-COLOR THEOREM (Heawood). *Every map can be colored with at most 5 colors.*

Proof. The argument is as before. We can assume inductively that a country F exists for whose 5 neighbors $A, \ldots, E$ all 5 colors r, b, g, w, s (in that order) are used. If there does not exist an r-g-chain from A to C then by an interchange we can save one of the colors. In the other case there is no b-w-chain from B to D and we can save a color. ∎

He next considered the coloring problem for maps on arbitrary surfaces. We have already hinted that by stereographic projection each plane map can be realized on the sphere, and conversely. We now consider the **torus**, T. Kempe had already constructed a map on T where 6 colors were needed. The map in Figure 2.2 consists of 7 countries, which are pairwise adjacent, so that 7 colors are needed. In the diagram we view the torus from above, thus we must imagine the individual countries as being extended around the tube.

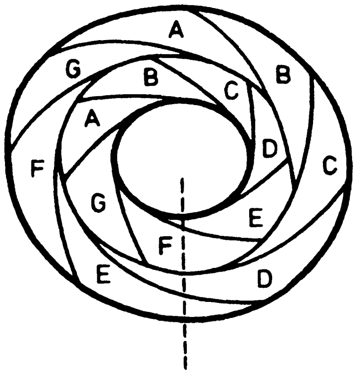


Figure 2.2

Slicing the torus along the dotted line and unrolling the cylinder so obtained, we obtain a plane realization in which the upper and lower edges are to be identified, as well as the left and right edges. In particular then, the "corners" are all one and the same point of the torus. In this plane realization, featured in Figure 2.3, the pairwise adjacency of the 7 countries is now easier to see.

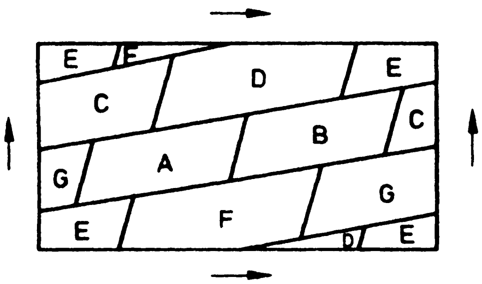


Figure 2.3

What do we mean by a closed surface? In topology, one learns the concept of an n-dimensional manifold. By that one understands a topological space X, which is Hausdorff and connected, and in which each point has a neighborhood that is homeomorphic to an open n-dimensional unit ball. The real line and the circle are examples of 1-dimensional manifolds. A **2-dimensional closed surface** (and it is only such that we consider) is a compact 2-dimensional manifold. Our first goal will be to classify all closed surfaces up to homeomorphism (i.e. up to bicontinuous transformations). For example, the 2-sphere is homeomorphic to an ellipsoid, but not homeomorphic to the torus. How this classification is carried out appears in most topology books (see Fréchet-Fan or Ringel). For our purposes it is sufficient to note the result. First of all it is shown that the surfaces split into two large classes: the **orientable** surfaces and the **nonorientable**. One calls a surface (or more generally a manifold) S **orientable**, if for every closed Jordan curve C in S, a rotation direction (e.g. clockwise) remains preserved as we go around C once. If this is not the case, then S is called **nonorientable**. Loosely speaking, orientable means that the surface is **two-sided**. One can best make this clear by the following: A fly that crawls around a ball on the inside of the skin, can never reach the outside. The same holds for the torus. These surfaces thus have two distinct sides. Two-sided surfaces of course had long been known, and had been extensively studied since Euler. Around the middle of the 19th century mathematicians noticed (above all, Möbius), that orientability is not necessarily a consequence of the topological properties of a surface. The Möbius strip in Figure 2.4 is one-sided: the fly can crawl around starting at P and arrive at the "other side". The Möbius strip, to be sure, is not a closed surface, but closed nonorientable surfaces are mathematically not difficult to describe, if not entirely easy to imagine.

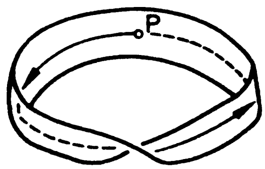


Figure 2.4

Within the class of orientable, resp., nonorientable surfaces there is still another structural invariant, the **genus** of the surface. **Classification Theorem**: *Two closed surfaces are homeomorphic if and only if they have the same orientability character and the same genus.*

In the class of orientable surfaces, the genus is an integer $h \geq 0$. The 2-sphere S_0 has genus 0, and we obtain a model of the oriented surface S_h of genus h, if we attach h "handles" to the 2-sphere. Figure 2.5 shows a model of S_3.

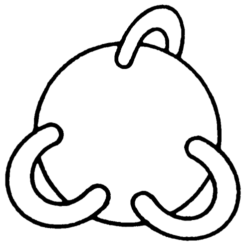


Figure 2.5

The torus is thus a model of S_1.

Among the class of nonorientable surfaces, the genus is a natural number $k \geq 1$. We obtain a model of the nonorientable surface N_k of genus k, if we cut out k disks from the 2-sphere, and in these k holes identify oppositely lying points. One also says that one places k "crosscaps" on the 2-sphere. Figure 2.6 shows a model of N_3. It is a useful convention to consider the 2-sphere also as the nonorientable surface N_0 of genus 0.

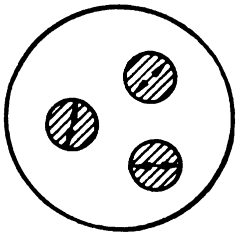


Figure 2.6

Back to the coloring problem. If we are given a closed surface S, then we understand by a **map** $\mathcal{L}$ on S, analogously to our previous discussion, a connected graph $G(\mathcal{L})$ which is embedded in S so that the edges (homeomorphs of $\overline{\mathbb{R}}$) meet each other only at endpoints, and the corresponding countries (homeomorphs of $\mathbb{R}^2$) are simply connected. The **chromatic number** $\chi(\mathcal{L})$ of $\mathcal{L}$ is again the minimal number of colors that are necessary in order to admissibly color $\mathcal{L}$. For a given surface S, it does not seem implausible that maps exist on S whose chromatic numbers are arbitrarily large. However the opposite is the case—and this is what Heawood showed in his work. *There is an upper bound for $\chi(\mathcal{L})$, which depends only on the surface S*—and this is true for any surface S.

One of the most famous formulas in all of mathematics says that the number $|V| - |E| + |R|$ is the same for every map $\mathcal{L}(V, E, R)$ embedded in S. Thus the number depends only on the surface S and not on $\mathcal{L}$. Therefore we can speak without ambiguity of the **Euler characteristic** $e(S) = |V| - |E| + |R|$ of the surface S. In 1.4 we have proven that $e(S_2) = 2$ for the plane, and thus for the 2-sphere. In general:

2.2 EULER-POINCARÉ FORMULA. *The Euler characteristic for S_h, resp., N_k is*

$$\text{i) } e(S_h) = 2 - 2h \qquad (h \geq 0),$$

$$\text{ii) } e(N_k) = 2 - k \qquad (k \geq 1).$$

Sketch of proof. If we attach a handle to an orientable surface, then we lower the characteristic by 2. As an example, consider the surface in Figure 2.7. There three edges and three countries are added, while two countries are removed. On the handle itself the proof proceeds as in 1.4. Hence if we attach h handles to S_0, then the characteristic will be reduced to $2-2h$. The nonorientable case is treated in exactly the same way. ■

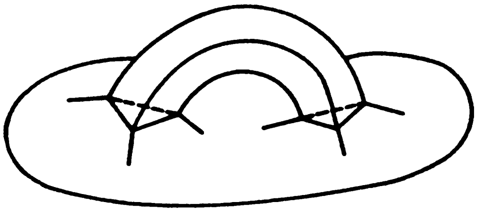


Figure 2.7

Hence the Euler characteristics for orientable surfaces are 2,0,-2,-4,..., and for nonorientable surfaces 1,0,-1,-2,....

EXAMPLE. *The nonorientable surface N_1 (which is homeomorphic to the real projective plane) can best be visualized as a disk in which the diametrically opposite boundary points are identified. Figure 2.8 shows a map with 6 countries on N_1, in which any two countries are adjacent. The map $\mathcal{L}$ has 10 vertices, 15 edges and 6 countries, which agrees with* $10 - 15 + 6 = 1$ $(= e(N_1))$.

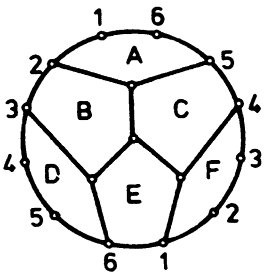


Figure 2.8

With these preparations we are ready to prove Heawood's coloring theorem. Let S be a closed surface. The **chromatic number** $\chi(S)$ of S is defined as the minimal number of colors that suffice to color *each* map $\mathcal{L}$ on S. Thus $\chi(S) = \max \chi(\mathcal{L})$, the max taken over all maps $\mathcal{L}$ on S. By Heawood's first result, 2.1, we know e.g. that $\chi(S_0) = 4$ or 5. For a general surface, however, it is not even apparent whether the chromatic number must be finite.

2.3 COLORING THEOREM OF HEAWOOD. *Let S be a closed surface with Euler characteristic $e(S) \leq 1$. Then*

$$\chi(S) \leq \left\lfloor \frac{7+\sqrt{49-24e(S)}}{2} \right\rfloor,$$

where $\lfloor x \rfloor$ denotes the greatest integer $\leq x$.

Proof. For $e \leq 2$ let $h(e)$ be defined by

$$h(e) = \frac{7+\sqrt{49-24e}}{2}.$$

We have $h^2(e) - 7h(e) + 6e = 0$, hence

$$6\left(1 - \frac{e}{h(e)}\right) = h(e) - 1.$$

Let $\mathcal{L}$ be a map on S. We may assume as in the first chapter that $\mathcal{L}$ is normal, since by the introduction of edges, the coloring number can only increase. Suppose $\mathcal{L}$ has p vertices, q edges, and r countries. Let d denote the average number of edges that bound a country. Thus we have $3p \leq 2q = dr$, $p - q + r = e(S)$, and hence $q \leq 3(q-p) = 3(r - e(S))$. It follows that

$$d = \frac{2q}{r} \leq \frac{6(r-e(S))}{r} = 6\left(1 - \frac{e(S)}{r}\right).$$

CLAIM: *$\mathcal{L}$ can be colored with $\lfloor h(e(S)) \rfloor$ colors.* We use induction on r. If $r \leq h(e(S))$, i.e. $r \leq \lfloor h(e(S)) \rfloor$, then the assertion is trivial. Thus let $r > h(e(S))$.

Case i), $e = e(S) \leq 0$. Then we have

$$d \leq 6\left(1 - \frac{e}{r}\right) \leq 6\left(1 - \frac{e}{h(e)}\right) = h(e) - 1.$$

Thus there is a country F, which borders on at most $\lfloor h(e) \rfloor - 1$ other countries. Contracting F to a point as usual, one color remains unused which can be applied to F.

Case ii), $e = e(S) = 1$. (Thus $S = N_1$.) Here

$$d \leq 6\left(1 - \frac{1}{r}\right) < 6 = h(1).$$

We can now reach the same conclusion using contractions. ∎

Why did we exclude the case $e(S) = 2$, i.e. the 2-sphere? Here we have $d \leq 6(1 - 2/r) < 6$, but $h(2) = 4$. Thus we can conclude with the usual induction proof only that $\chi(S_0) \leq 6$, an even weaker result than the already established Theorem 2.1.

Substituting 2.2 into formula 2.3, we obtain:

2.4 COROLLARY.

$$\text{i)}\quad \chi(S_h) \le \left\lfloor \frac{7+\sqrt{1+48h}}{2} \right\rfloor \qquad (h \ge 1).$$

$$\text{ii)}\quad \chi(N_k) \le \left\lfloor \frac{7+\sqrt{1+24k}}{2} \right\rfloor \qquad (k \ge 1).$$

Observe that for $h = 0$, we would exactly obtain $\chi(S_0) \le 4$—but, unfortunately, the proof does not work in this case.

EXAMPLES. *For the torus S_1 and the projective plane N_1, 2.4 yields $\chi(S_1) \le 7$ and $\chi(N_1) \le 6$. Since we have already constructed corresponding maximal chromatic maps in Figures 2.2 and 2.8, we can now assert equality for both numbers.*

Heawood's result appeared to imply that the 4-color conjecture was almost certainly true. Furthermore, it suggested that an intensive study of the colorability of surfaces of higher genus might also provide new insight about the 2-sphere, resp., the plane. In particular, the question was immediately taken up, as to how close the upper bound in 2.4 is to the real chromatic number. Since the development and the ultimate solution of this question is one of the most fascinating mathematical efforts, we will go into it in some detail, somewhat anticipating the historical course of events.

Heawood also committed an error. He clearly knew that to prove equality in 2.4(i) (Heawood considered only the orientable case) one must indeed construct a map on S which needs $\lfloor h(e(S)) \rfloor$ colors. However, he only carried this out for the torus and was obviously under the impression that the general case went similarly. Not really: It took 78 years until the proof was finished (naturally with the exception of S_0), and it is still being improved.

In 1891 Lothar Heffter pointed out the incompleteness of Heawood's argument, and he was also the first to show equality in 2.4(i) for a few additional values of h. In particular, he contributed two important new ideas.

Let us consider the surface S_h. In order to calculate $\chi(S_h)$, we must determine the *maximal* chromatic number $\chi(\mathcal{L})$ of all maps on S_h. It is easy to see that $\chi(S_h) \le \chi(S_{h+1})$, as also $\chi(N_k) \le \chi(N_{k+1})$. Heffter now turned the problem around. Let $\mathcal{L}_n$ be a map with n countries, which all border on one another. What is the *minimal* genus h (or k in the nonorientable case), so that $\mathcal{L}_n$ can be embedded in S_h (resp., N_k)? This minimal number will be denoted by $\gamma(n)$ in the orientable case and by $\bar{\gamma}(n)$ in the nonorientable case. Since $\mathcal{L}_5$ can not be embedded in the plane, but in the torus and in N_1, we have e.g. $\gamma(5) = \bar{\gamma}(5) = 1$. Naturally, $\gamma(4) = \bar{\gamma}(4) = 0$.

2.5 LEMMA. *For* $n \geq 3$,

$$\text{i)} \quad \gamma(n) \geq \left\lceil \frac{(n-3)(n-4)}{12} \right\rceil.$$

$$\text{ii)} \quad \bar{\gamma}(n) \geq \left\lceil \frac{(n-3)(n-4)}{6} \right\rceil.$$

where $\lceil x \rceil$ *is the smallest integer* $\geq x$.

Proof. Let $\mathcal{L}_n$ be embedded in S_h, where we can assume as usual that each vertex touches at least 3 edges. From $|V| - |E| + |R| = 2 - 2h$, $3|V| \leq 2|E|$ we obtain $|E| \leq 3|R| - 6 + 6h$. Furthermore, by $|R| = n$, $|E| \geq \binom{n}{2}$, we have

$$\frac{n(n-1)}{2} \leq 3n - 6 + 6h, \quad \text{or}$$

$$h \geq \frac{n(n-1) - 6n + 12}{12}$$
$$= \frac{(n-3)(n-4)}{12}.$$

Thus $\frac{(n-3)(n-4)}{12}$ is a lower bound for each such h, and thus also for the minimal $\gamma(n)$. The proof of ii) is completely analogous. ∎

2.5 is, as the proof also suggests, a counterpart to 2.3 and thus no great leap forward. However, what is really helpful is the fact that equality for all n in 2.5 implies also equality in 2.3 for all S. It is indeed certainly easier (if still difficult enough), to embed a *given* map $\mathcal{L}_n$ into a *particular* $S_{\gamma(n)}$ (resp., $N_{\bar{\gamma}(n)}$) than for *each* surface, to investigate the colorability in principle of infinitely many maps. First we consider the orientable case.

2.6 LEMMA. *The equality in* 2.5(i) *for all* $n \geq 3$ *implies the equality in* 2.4(i) *for all* $h \geq 1$.

Proof. Suppose S_h is given and $\mathcal{L}_n$ is embedded in S_h with maximal n. Then by 2.4(i)

$$n \leq \chi(S_h) \leq \left\lfloor \frac{7 + \sqrt{1 + 48h}}{2} \right\rfloor.$$

By the choice of n we infer $\gamma(n+1) > h$. Assuming then equality in 2.5(i), we obtain

$$\left\lceil \frac{(n-2)(n-3)}{12} \right\rceil > h \quad \text{or}$$

$$n^2 - 5n + 6 - 12h = \left(n - \frac{5+\sqrt{1+48h}}{2}\right)\left(n - \frac{5-\sqrt{1+48h}}{2}\right) > 0.$$

It follows that

$$n > \frac{5+\sqrt{1+48h}}{2},$$

and hence

$$\frac{7+\sqrt{1+48h}}{2} - 1 < \chi(S_h) \leq \frac{7+\sqrt{1+48h}}{2},$$

i.e.

$$\chi(S_h) = \left\lfloor \frac{7+\sqrt{1+48h}}{2} \right\rfloor. \quad \blacksquare$$

As we shall see in a moment, equality 2.5(i) for oriented surfaces was indeed established for all n— and hence also in 2.4(i). The nonorientable case, however, contains *one* exception, $n = 7$. Franklin showed in 1934, that $\mathcal{L}_7$ can *not* be embedded in N_2, even though 2 is the bound in 2.5(ii) for $n = 7$. Since $\bar{\gamma}(7) = 3$, this exception implies the one exception $\chi(N_2) = 6$ in Heawood's bound 2.4(ii). This exceptional surface N_2 is called Klein's bottle (although, since it has only one side, it would be somewhat impractical as a bottle). Apart from this exception, the proof of 2.7 is completely analogous to 2.6.

2.7 LEMMA. *The equality in* 2.5(ii) *for all $n \neq 7$ and $\bar{\gamma}(7) = 3$ imply equality in* 2.4(ii) *for all $k \geq 1$, $k \neq 2$, and $\chi(N_2) = 6$.*

Heffter himself proved the equality in 2.5(i) for $n \leq 12$ and for a remarkable sequence, whose first elements are $n =$19, 31, 55, 67, 139, ... And this is where things remained for a long time.

Heffter's second idea was to use the concept of **duality** of maps. If $\mathcal{L}$ is a map on a surface S, then we can construct a **dual map** $\mathcal{L}^*$ in the following way: In the interior of each country we place a vertex, and we join two such vertices v^* and w^* by an edge e^* in $\mathcal{L}^*$ precisely when the corresponding countries U and W have a common border e in $\mathcal{L}$, whereby we cross e^* with e. (If e is a bridge in the country U, then we draw a loop e^* at v^*.) Figure 2.9 shows this construction in the plane and on the torus.

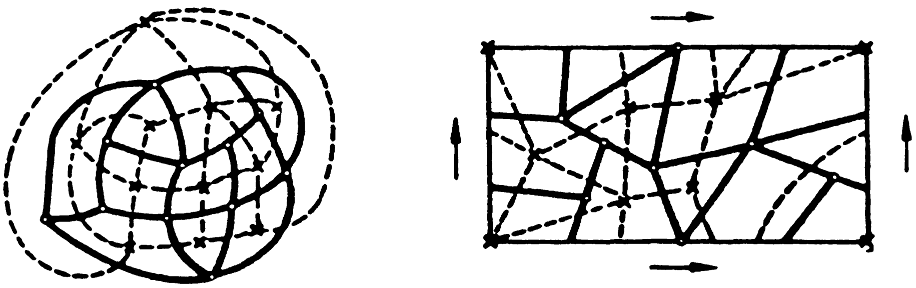


Figure 2.9

For the map $\mathcal{L}^\star(V^\star, E^\star, R^\star)$ that is dual to $\mathcal{L}(V, E, R)$ we clearly have $|V^\star| = |R|$, $|E^\star| = |E|$, $|R^\star| = |U|$. Since we assume that the skeleton $G = G(\mathcal{L})$ of a map $\mathcal{L}$ is always connected, then obviously $\mathcal{L}^{\star\star} \cong \mathcal{L}$ using the natural correspondences. Now we consider the skeleton $G^\star = G(\mathcal{L}^\star)$ of $\mathcal{L}^\star$ (indicated by dotted lines in the figure). $G^\star$ is called the **dual graph** corresponding to G, and $G^{\star\star} \cong G$. A **coloring of the countries** of $\mathcal{L}$ corresponds precisely to a **vertex coloring** of $G^\star$, in the sense that adjacent vertices obtain different colors. Let us make this more precise.

DEFINITION. *Let $G(V, E)$ be an arbitrary graph. A* **coloring** *of G is a mapping $f : V \to M$ (M the color set), such that $uv \in E$ implies $f(u) \neq f(v)$. The* **chromatic number** *$\chi(G)$ is the smallest number of colors which is needed for a coloring.*

Hence we have: $\chi(\mathcal{L}) = \chi(G^\star)$. *Coloring problems for maps are therefore equivalent to coloring problems for graphs*, and these are, mostly, substantially easier to deal with. It is for this reason that in later chapters we will almost always consider the graph version of the 4-color problem. If $\mathcal{L}$, resp., $G(\mathcal{L})$ is embedded in the surface S, so also is $\mathcal{L}^\star$, resp., $G^\star = G(\mathcal{L}^\star)$. In particular, $\mathcal{L}$ is a plane map exactly when $G^\star$ is a plane graph. Hence we note:

2.8 THEOREM. *The 4-color conjecture is equivalent to the conjecture that all planar graphs are 4-colorable.*

On the basis of the relation $\mathcal{L} \leftrightarrow G^\star$ (maps $\leftrightarrow$ plane graphs) we can turn theorems about plane maps, by "dualization", into theorems about graphs. The first example is the dual form of 1.5, resp., 1.6. Bridges in $\mathcal{L}$ obviously correspond to loops in $G^\star$ and conversely. Thus we obtain:

2.9 THEOREM. *Let $G(V,E)$ be a simple, connected, plane graph with at least 3 edges. (Observe that each country must then have at least 3 borders.) Denote by p_i the number of vertices of degree i, then $\sum_{i\geq 1}(6-i)p_i \geq 12$; and we have equality if G has no bridges and all countries have exactly three borders. In particular, there must always exist a vertex of degree ≤ 5.*

Back to a general surface S. A graph G is called **embeddable** in S if G is isomorphic to a graph whose vertices are points in S and whose edges are Jordan curves in S which intersect only at corresponding endpoints. In order to determine $\chi(S)$, we can by our remarks above consider also the maximum of $\chi(G)$ over all graphs embeddable in S. We define in the obvious way the **genus** $\gamma(G)$ (resp., the **nonorientable genus** $\bar{\gamma}(G)$) of a graph G as the smallest number h (resp., k) such that G can be embedded in S_h (resp., N_k). The question arises then whether every graph can be embedded in some surface S_h (resp., N_k). However that is clear. We draw G first on a 2-sphere, whereby edges may cross. For each edge uv we can insert a handle over which the edge runs. These handles clearly can be deformed so that no more crossings appear. The argument is similar in the nonorientable case. The graph $G^\star$ derived from the map $\mathcal{L}_n$ has by our preceding discussion n vertices, which are all pairwise adjacent. We call $G^\star$ the **complete graph** and denote it by K_n. The numbers $\gamma(n)$, resp. $\bar{\gamma}(n)$, defined before are then precisely the genus $\gamma(K_n)$, resp. $\bar{\gamma}(K_n)$.

As a summarization of Heffter's ideas, we can reformulate 2.5 through 2.7:

2.10 THEOREM. *For $n \geq 3$,*

$$\gamma(K_n) \geq \left\lceil \frac{(n-3)(n-4)}{12} \right\rceil,$$

$$\bar{\gamma}(K_n) \geq \left\lceil \frac{(n-3)(n-4)}{6} \right\rceil.$$

If equality holds for all n except $\bar{\gamma}(K_7) = 3$, then equality holds in Heawood's coloring theorem 2.3 except $\chi(N_2) = 6$ (always excluding S_0).

In this formulation as a graph embedding, the determination of $\gamma(K_n)$, resp. $\bar{\gamma}(K_n)$ became known as the **thread problem**, mentioned first in the book of Hilbert-Cohn-Vossen: Can we pick n points on S_h and join all of these points with $\binom{n}{2}$ threads such that no two of them cross? Figure 2.10 shows a thread figure, the dual of Figure 2.3, on the torus with $n = 7$.

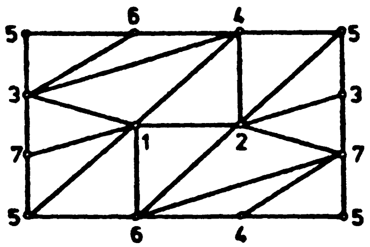


Figure 2.10

How did the story continue? In 1910 Tietze considered for the first time the nonorientable case and proved 2.4(ii). He proved that $\chi(N_2) \geq 6$, while the upper bound is actually 7, as we know. The fact that $\chi(N_2) = 6$ was, as previously mentioned, proven by Franklin, who determined some other values. Until 1939 it was known that equality holds in 2.4(ii) for $k \leq 7$, $k \neq 2$. Finally it was Ringel who in 1954 gave the complete solution:

$$\bar{\gamma}(n) = \left\lceil \frac{(n-3)(n-4)}{6} \right\rceil$$

always holds with the exception of $\bar{\gamma}(7) = 3$. Hence the Heawood coloring theorem was finally proven in the nonorientable case.

The orientable case was considerably more difficult. After Heffter, who, as mentioned, proved equality in 2.5(i) for $n \leq 12$ and some other values, it languished there until 1952, when Ringel proved equality for $n = 13$. In 1954, it was again Ringel who gave the corresponding proof for all n of the form $n = 12k + 5$ and in 1961 for all $n = 12k + r$ with $r = 3, 7, 10$. After that, several mathematicians had part in the solution, above all Ringel, Youngs, and Gustin. At the end of February 1968 only the single case $n = 30$ was still open, and this was given the coup de grâce by J. Mayer, a professor of French literature.

The thread problem was solved, and with it, as we established in 2.10, the Heawood coloring theorem. Well, not entirely. The case of the 2-sphere or equivalently the plane, was still open. Many results were obtained in the hope of coming closer to a solution. However the question was still undecided: Is $\chi(S_0) = 4$ or 5? And so, we will again pick up the thread of the development of the 4-color problem.

EXERCISES FOR CHAPTER 2

1. Verify the remark after 2.1, that each plane map can be realized on the 2-sphere, and conversely.

2°. We call a graph $G(V, E)$ **bipartite** if the vertex set V decomposes into two disjoint parts V_1, V_2, so that all edges have one endpoint in V_1 and the other in V_2. If $|V_1| = m$, $|V_2| = n$ and if G contains all edges between V_1 and V_2, then G is called **complete bipartite**,and is denoted $K_{m,n}$. E.g. $K_{2,2}$ is isomorphic to a circuit of length 4. Show: a. $K_{3,3}$ is not planar, either directly from the Jordan Curve Theorem, or using 1.4. b. $K_{3,3}$ can be embedded without intersections in the torus. Can $K_{3,4}$ and $K_{4,4}$ also be embedded in the torus?

3. Consider the Petersen graph P (Figure 3.15). Show: a. P is not planar, b. the edges of P can not be 3-colored, c. P can be embedded in the torus.

4. Embed the complete graph K_6 in the projective plane N_1.

5. Can K_5 or $K_{3,3}$ be embedded in the Möbius strip?

6°. The **Platonic Graphs** are the skeleton graphs of the Platonic bodies: tetrahedron, cube, octahedron, dodecahedron, icosahedron. The first two are isomorphic to K_4 , resp., Q_3 (see exercise 1.1), the last three are pictured in Figure 3.16, resp., 5.9. Draw the dual graph for each. Which are self dual, i.e. isomorphic to the dual graph?

7. Embed K_6 in the torus and draw the dual graph. What properties does this graph have?

8°. Given the two plane graphs

show: $G \cong H$, but $G^* \ncong H^*$. Is 7 the smallest number of vertices where this is possible?

9. A **wheel**, W_n, is a circuit C_n of length n together with another vertex, which is joined with all of the circuit vertices (Example: $W_3 \cong K_4$). Show that all wheels are planar and as a plane graph are self dual. What other graphs are self dual?

10°. If we go from a plane graph $G = G(\mathcal{L})$ to its dual $G^* = G(\mathcal{L}^*)$ then we obtain corresponding pairs of concepts. Example: number of borders $\leftrightarrow$ degree of vertex. Find additional pairs of concepts. Dualize the statement in exercise 1.12.

11. Prove another coloring theorem of Heawood: A normal cubic plane map is 3-colorable $\Leftrightarrow$ all countries have an even number of borders. (Hint: Dualize!)

12. What does a map look like, whose vertices can be admissibly colored with 2 colors, and also the countries?

13. Show with the help of 2.10 that there is no complete graph of orientable genus 7. What is the next number that is not the genus of a K_p? Determine similar missing values in the nonorientable genus.

14°. Show

$$\gamma(K_{m,n}) \geq \left\lceil \frac{(m-2)(n-2)}{4} \right\rceil$$

for all m, n. Find the corresponding inequality for $\tilde{\gamma}(K_{m,n})$. Determine $\gamma(K_{4,4})$.

15. Puzzle. We already know that $K_{4,4}$ is not planar. Every realization of $K_{4,4}$ in the plane therefore must have crossing points. What is the smallest possible number of crossings?

16. Show: Every simple graph can be realized in $\mathbb{R}^3$ without crossing points (vertices = points, edges = Jordan curves).

17*. Somewhat harder: Show that each simple planar graph can be embedded in the plane so that all edges are straight lines. (Hint: Induction!)

3. NEW BEGINNINGS

Although around 1900 the 4-color problem was still considered a combinatorial or topological curiosity which perhaps could be solved in one lucky stroke, we can also perceive some initial ideas for treating the problem within a theoretical framework. Essential ground was broken in five directions, which we will now discuss.

Heawood's work of 1898 and most of his later work proposed an *arithmetization* of the problem. He first proved, building on Tait's coloring theorem 1.7, that a normal cubic map is 4-colorable, under the hypothesis that the number of boundaries around each country is divisible by 3. (By a map we naturally again mean a map in the plane.) As a generalization of this remark he showed:

3.1 THEOREM (Heawood). *Let $\mathcal{L}(V, E, R)$ be a normal cubic map. $\mathcal{L}$ can be 4-colored exactly when a function $h : V \to \{1, -1\}$ exists such that for all countries F, the sum $\sum_{v \in F} h(v)$, taken over all the vertices on the boundary of F is divisible by 3—in symbols*

$$\sum_{v \in F} h(v) \equiv 0 \pmod 3.$$

Proof. If $\mathcal{L}$ is 4-colorable, then by 1.7 the edges can be colored with 3 colors α, β, γ. Around a vertex v the colors appear in a clockwise sense either in the sequence $\alpha \to \beta \to \gamma$, or $\alpha \to \gamma \to \beta$. In the first case we set $h(v) = 1$, in the second $h(v) = -1$. We must show that the function h constructed in that way satisfies the conditions of the theorem. Let F be a country and e a border of F. We run through the boundary of F in a counterclockwise direction starting with e. If l and m are adjacent borders, then the color of m is uniquely determined by the color of l and the value $h(v)$ of the common vertex. For example if l has the color β and if $h(v) = -1$, then the color of m must be α. Since, after one trip around, the color of e again must result, it follows that

$$\sum_{v \in F} h(v) \equiv 0 \pmod 3$$

as asserted (see Figure 3.1).

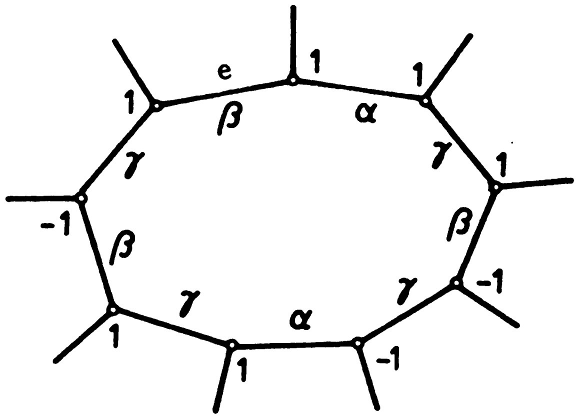


Figure 3.1

If conversely $h : E \to \{1, -1\}$ is given, then one constructs an edge coloring by starting at an arbitrary edge and continuing the coloring as in Figure 3.1. ■

3.1 puts the 4-color problem in a purely arithmetic form. Let a normal cubic map $\mathcal{L}(V, E, R)$ be given. With each vertex v_i we associate a variable x_i, and for each country F we introduce a congruence

$$x_i + x_j + x_k + \ldots + x_l \equiv 0 \pmod 3$$

where $v_i, v_j, v_k, \ldots, v_l$ are the vertices on the boundary of F. We obtain in this manner a system of congruences (in which each x_i appears in exactly 3 congruences), and it follows from 3.1 that the 4-color conjecture is correct if each such system of congruences has a solution with $x_i = 1$ or -1 for all i. Heawood himself investigated such congruences in general in a series of papers. He came no closer to a solution, and his methods were not taken up by other mathematicians. Still it was the first significant attempt to bring in methods that were not purely combinatorial or topological.

A similar, *geometrically* motivated, approach was suggested by Veblen in a paper in 1912. He began by describing a map $\mathcal{L}(V, E, R)$ by means of a pair of **incidence matrices**, an idea which had already appeared in Kirchhoff's work in 1847.

As usual, let $\mathcal{L}$ be normal (it is sufficient to assume that $\mathcal{L}$ has no bridges) with p vertices, q edges and r countries. We enumerate the vertices $V = \{v_1, \ldots, v_p\}$ as well as $E = \{e_1, \ldots, e_q\}$, and $R = \{F_1, \ldots, F_r\}$. The **vertex-edge-incidence matrix** A is a 0,1-matrix with p rows and q columns in which each row corresponds to a vertex and each column corresponds to an edge. If v_i is incident with e_j a 1 is placed in the i,jth position, and a 0 is placed otherwise. The **country-edge-incidence matrix** B, with r rows and q columns, is defined analogously. Figure 3.2 shows a map and the corresponding incidence matrices.

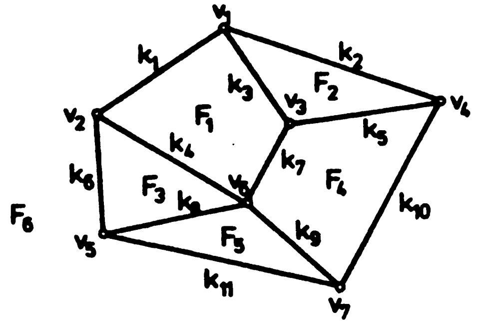


Figure 3.2

$$A = \begin{pmatrix} 1&1&1&0&0&0&0&0&0&0&0 \\ 1&0&0&1&0&1&0&0&0&0&0 \\ 0&0&1&0&1&0&1&0&0&0&0 \\ 0&1&0&0&1&0&0&0&0&1&0 \\ 0&0&0&0&0&1&0&1&0&0&1 \\ 0&0&0&1&0&0&1&1&1&0&0 \\ 0&0&0&0&0&0&0&0&1&1&1 \end{pmatrix} \qquad B = \begin{pmatrix} 1&0&1&1&0&0&1&0&0&0&0 \\ 0&1&1&0&1&0&0&0&0&0&0 \\ 0&0&0&1&0&1&0&1&0&0&0 \\ 0&0&0&0&1&0&1&0&1&1&0 \\ 0&0&0&0&0&0&0&1&1&0&1 \\ 1&1&0&0&0&1&0&0&0&1&1 \end{pmatrix}$$

We consider A and B as matrices over the field $GF(2)$ that has two elements 0, 1; i.e. we add and multiply by the rules $0+0=0$, $0+1=1+0=1$, $1+1=0$, $0\times 0=0\times 1=1\times 0=0$, $1\times 1=1$. Each row of A or B, and in general each 0,1-vector $c=(c_1,\ldots,c_q)$ can be unambiguously identified with a set C of edges, those for which $c_i=1$. Thus $C=\{e_i\in E: c_i=1\}$. The addition of vectors $(c_1,\ldots,c_q)+(d_1,\ldots,d_q)=(c_1+d_1,\ldots,c_q+d_q)$ can immediately be carried over to the sets C and D. Since c_i+d_i is 1 exactly when $e_i\in C$ and $e_i\notin D$, or $e_i\in D$ and $e_i\notin C$, we see that the **sum** $C+D=(C-D)\cup(D-C)$ is precisely the symmetric difference of C and D. The empty set naturally corresponds to the null vector $(0,0,\ldots,0)$. By the arithmetic rules in $GF(2)$, it follows that the inner product $\sum_{i=1}^{q} c_i d_i = 1$ or $=0$, depending on whether the intersection $C\cap D$ contains an odd or even number of edges.

Let us now consider the system of linear equations $Ax=0$. A vector c is a solution precisely when the inner product of c with each row vector is 0. Since the row vectors correspond to the edges incident with a given vertex, we obtain: c is a solution precisely when the corresponding edge set C contains an even number of edges from each neighborhood $N(v)$, i.e., when each vertex v has even degree $d_C(v)$, in the subgraph $G(V,C)$ generated by C. We call such a graph a **cycle** or an **Euler graph**. Thus G is an Euler graph if each vertex has even degree.

The name "Euler Graph" originally arose from an entirely unrelated problem. Let G be a connected graph. An **Euler trail** is a closed edge trail which goes through each edge exactly once. Many puzzles are of this sort: One runs through all the edges of a figure, without picking up the pencil and arrives back at the beginning point. If G has an Euler trail, then each vertex has even degree, since if a vertex is arrived at, then it must be left along a different edge. However the converse is also true, as the reader can easily convince himself: If each vertex in G has even degree, then there is an Euler trail. Hence for connected graphs, being an Euler graph and having an Euler trail are equivalent.

The name Euler trail arises from perhaps the oldest graph theory work, the solution of the Königsberg bridge problem by Euler in 1736. Figure 3.3 shows 4 pieces of land A,B,C,D and 7 bridges joining them $a,b,\ldots,g$. Question: Is it possible starting on island A, to walk on a path in which one crosses all the bridges exactly once and ends up on A? If we translate this in an obvious way into the language of graphs, then the graph on the right in Figure 3.3 results. Since A however has odd degree (as do all the other vertices), the answer is no.

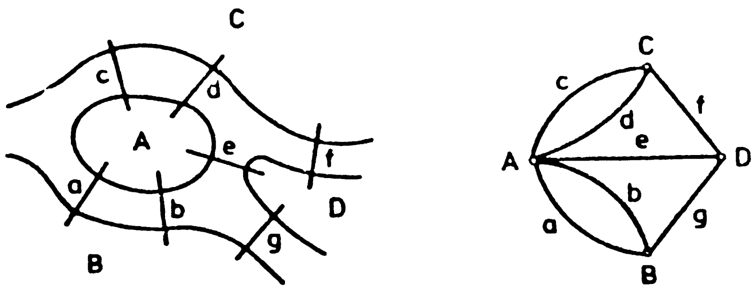


Figure 3.3

Back to our theme. We can describe Euler graphs even more concisely. If $C \neq \emptyset$ and $e \in C$, then since $d_C(v)$ is even for all $v \in V$, we can find a path starting with e, which must meet itself eventually at a previously passed vertex. We remove the edges of the circuit constructed in this way, and choose a new edge, if there are any, construct a new circuit, etc. Thus finally C decomposes into **edge disjoint** circuits, and conversely it is clear that every union of edge disjoint circuits is an Euler subgraph. In Figure 3.2 $c = (0,1,1,1,1,1,0,0,1,0,1)$ is, e.g., a solution to $Ax = 0$. The corresponding set of edges $c = \{e_2, e_3, e_4, e_5, e_6, e_9, e_{11}\}$ decomposes into the circuits $\{e_2, e_3, e_5\}$ and $\{e_4, e_6, e_{11}, e_9\}$.

If $F \in R$, the set C_F of edges incident with F is certainly a cycle, and hence the vector c_F is a solution of $Ax = 0$. We call the cycles C_F, resp. the vectors c_F, the **fundamental cycles**, resp. **fundamental solutions**. Every other solution is a linear combination of the fundamental solutions, for each circuit subdivides the plane into an inside and an outside and is hence the sum of the fundamental circuits in its interior. E.g., in Figure 3.2 the circuit $\{e_1, e_3, e_7, e_9, e_{11}, e_6\}$ corresponds to the vector $(1,0,1,0,0,1,1,0,1,0,1)$ which is the sum of the fundamental solutions $(1,0,1,1,0,0,1,0,0,0,0) + (0,0,0,1,0,1,0,1,0,0,0) + (0,0,0,0,0,0,0,1,1,0,1)$. Since each edge lies on exactly 2 fundamental cycles, $\sum_F c_F = 0$. Thus the c_F's are linearly dependent. Every proper subset however is linearly independent, since $\sum_{F \in R'} c_F = 0$ for $\emptyset \neq R' \subsetneq R$ would mean that each edge, if it appears at all, must occur in the countries of R' at least twice, which is clearly impossible. The dimension of the solution space of $Ax = 0$ thus is $r-1$ and hence the rank of the matrix A, $\mathrm{rk}(A) = q - r + 1$.

Since the fundamental cycles correspond precisely to the rows of B, we obtain without further ado, the matrix equality $A \cdot B^T = 0$, where B^T denotes the transpose of B, and $\mathrm{rk}(B) = r - 1$.

One obtains a second system of equations when one considers $A^T y = 0$. The equations all have the form $y_a + y_b = 0$, where ab is an edge of G. By the calculation rules in $GF(2)$, we infer, $y_a = y_b$ for each solution of $A^T y = 0$, and since G is connected, all the variables must assume the same value. That is, $(0,0,...,0)$ and $(1,1,...,1)$ are the only solutions. The solution space has dimension 1 and consequently A^T has rank $\mathrm{rk}(A^T) = p - 1$. From $\mathrm{rk}(A) = \mathrm{rk}(A^T)$ it follows that $q - r + 1 = p - 1$, i.e. $p - q + r = 2$, and this is just our old Euler Formula.

Veblen now proceeded to the 4-color problem. He chose the 4 elements from the field $GF(4)$ as the color set. For two elements α, β of $GF(4)$, $\alpha + \beta = 0$ if and only if $\alpha = \beta$. Let us consider the system $B^T z = 0$ over $GF(4)$. The equations are of the form $z_F + z_G = 0$, where F and G are adjacent countries. A 4-coloring of $\mathcal{L}$ then corresponds precisely to a vector $(e_1, \ldots, e_r)$ over $GF(4)$, which does *not* satisfy any of the equations $z_F + z_G = 0$. Veblen now interpreted the vectors $(e_1, \ldots, e_r)$ as **points** in an $(r-1)$-dimensional projective space, each of the equations $z_F + z_G = 0$ as a **hyperplane**, and derived from that the equivalence of the 4-color problem to a purely geometric problem.

Veblen's ideas were advanced further by Whitney in the 1930's, and after that primarily by Tutte. For the moment, in order to make the connection to the coloring problem more transparent, we will dualize Veblen's approach.

Let us turn back to the matrices A and B. $\hat{A}$ and $\hat{B}$ shall denote the subspaces of the vector space $\mathcal{V}_q$ of all 0,1-tuples of length q over $GF(2)$ which are spanned by the rows of A resp.

B. We already know: $\dim \hat{A} = q - r + 1$ and $\dim \hat{B} = r - 1$. We also know that $\hat{B}$ consists precisely of those vectors b which are orthogonal to all $a \in \hat{A}$, i.e. for which the inner product $a \cdot b = 0$. Hence we write $\hat{B} = \hat{A}^{\perp}$. Since $\dim \hat{A} + \dim \hat{B} = q$, it also follows that $\hat{A} = \hat{B}^{\perp}$. In order to prove that $b \in \hat{B}$, therefore by linearity it is sufficient to prove that b is orthogonal to all the rows of A.

We already know that the edge sets belonging to the vectors $b \in \hat{B}$ are precisely the **Euler subgraphs**. We will now derive a further description of $\hat{B}$, from which the connection to the coloring problem will become apparent. Let $f : R \to GF(2)$ be an arbitrary mapping. To f we associate the mapping $\delta f : E \to GF(2)$ by $(\delta f)(e) = f(F) + f(G)$, where e borders the countries F and G. δf is called the **boundary** corresponding to f. By linearity, the boundaries form a subspace of $\mathcal{V}_q$, and we will show that this is precisely the subspace $\hat{B}$. If δf is given, and $v \in V$ with the corresponding row a_v of A, then $a_v \cdot \delta f = \sum_{e \ni v} (\delta f)(e) = \sum_{e \ni v} (f(F) + f(G)) = 0$ since each $f(F)$ occurs twice in the sum. Each boundary is thus in $\hat{B}$. For the reverse inclusion, it is sufficient by linearity to show that each vector c_F corresponding to a fundamental cycle C_F is also a boundary. Now we choose the function $f_F : R \to GF(2)$ with $f_F(F) = 1$ and $f_F(G) = 0$ for all $G \neq F$. Then obviously $(\delta f_F)(e) = 1$ exactly when e is incident with F. That means however $c_F = \delta f_F$, thus the claim is correct. Figure 3.4 shows a mapping $f : R \to GF(2)$. The edges of the corresponding boundaries, resp., Euler subgraphs, are indicated with heavy lines.

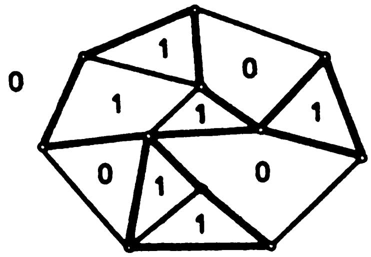


Figure 3.4

Suppose we want to color $\mathcal{L}$ with 2 colors. We choose as the color set 0,1 from $GF(2)$. A permissible coloring corresponds then to precisely a function $f : R \to \{0, 1\}$ with $f(F) \neq f(G)$ for adjacent countries F and G, or, what is the same thing, to a boundary δf (resp. Euler subgraph) with $(\delta f)(e) = 1$ for *all* $e \in E$. We obtain immediately from this a result of Tait:

3.2 THEOREM. *A bridgeless map $\mathcal{L}$ is 2-colorable if and only if the skeleton $G(\mathcal{L})$ is an Euler graph.*

Let us now color $\mathcal{L}$ with 4 colors. As the color set M we take the pairs (0,0), (0,1), (1,0), and (1,1) with 0 and 1 from $GF(2)$. A coloring $f : R \to M$ corresponds to a pair (f_1, f_2) of mappings $f_i : R \to \{0,1\}$ with $f(F) = (f_1(F), f_2(F))$. f is an admissible coloring if $f_1(F) \neq f_1(G)$ or $f_2(F) \neq f_2(G)$ for each two adjacent countries, or, what is equivalent to that, if $(\delta f_1, \delta f_2)(e) \neq (0,0)$ for all $e \in E$. Thus for the 4-colorability of $\mathcal{L}$, it is necessary and sufficient that 2 boundaries (resp. Euler subgraphs) exist which among them cover all edges e.

3.3 THEOREM. *The 4-color conjecture is true if and only if the skeleton $G(\mathcal{L})$ of every bridgeless map $\mathcal{L}$ is the union of two Euler subgraphs.*

Figure 3.5 shows a (not disjoint) decomposition of a graph into two Euler subgraphs C_1 and C_2 and the 4-coloring induced thereby. For clarity, the colors have been replaced by $a = (0,0)$, $b = (0,1)$, $c = (1,0)$ and $d = (1,1)$. The edges of C_1 are drawn with bold lines, those of C_2 are denoted by dashes. The outer country is first colored with a.

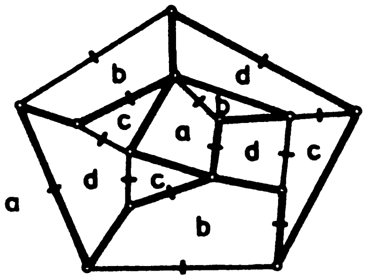


Figure 3.5

Let $G = G(\mathcal{L})$ be a plane graph with incidence matrix A. Then the dual graph $G^* = G(\mathcal{L}^*)$ has the incidence matrix B, and $\hat{B} = \hat{A}^{\perp}$ as we have already seen. Now, we can of course set up the incidence matrix A for any graph G, whether or not G is plane. But does there exist a graph that is dual to it? That is, a graph for whose incidence matrix B we have $\hat{B} = \hat{A}^{\perp}$? The answer to this question (by Whitney) will be presented in Chapter 4. It is the beginning of one of the most interesting chapters of graph theory and more precisely its generalization, the **theory of matroids** whose fundamentals we shall discuss in Chapter 8.

A third proposal on how to attack the 4-color conjecture was made by G. D. Birkhoff (who was a colleague of Veblen at Princeton) in a series of papers starting in 1913. Let $\mathcal{L}$ be a map and λ a natural number. Birkhoff developed a formula for the number $p(\mathcal{L}; \lambda)$ of colorings of $\mathcal{L}$ with at most λ colors. If one could show $p(\mathcal{L}; 4) > 0$ for all maps $\mathcal{L}$ then this would give a positive answer to the 4-color conjecture. One can thus describe Birkhoff's idea as **quantification** of the 4-color problem. Birkhoff's first important result showed that $p(\mathcal{L}; \lambda)$ is always a polynomial in λ of degree $|R|$, called the **chromatic polynomial** of $\mathcal{L}$. Since zeros of polynomials can be found by many methods from algebra and analysis, at least approximately, it was hoped that one could show that $p(\mathcal{L}; 4) \neq 0$ for all chromatic polynomials. That was not without reason—one could imagine obtaining a precise characterization of these polynomials.

We will prove Birkhoff's result (and several extensions due to Whitney) for graphs instead of maps, and indeed in complete generality for arbitrary graphs G.

3.4 THEOREM. *Let $G(V, E)$ be a graph, without multiple edges, that has p vertices and q edges. Let $p(G; \lambda)$ be the number of vertex colorings of G with at most λ colors. Then $p(G; \lambda)$ is a polynomial in λ,*

$$p(G, \lambda) = a_0\lambda^p + a_1\lambda^{p-1} + \ldots + a_p.$$

If G contains a loop, then $p(G; \lambda) \equiv 0$, otherwise:

(i) *$p(G; \lambda)$ has degree p.*

(ii) *The coefficients $a_0, a_1, \ldots, a_{p-t}$ are all $\neq 0$ and have alternating signs, where t is the number of components of G.*

(iii) *$a_0 = 1$, $a_1 = -q$, $a_i = 0$ for $i > p - t$.*

(iv) *If G has components $G_1, \ldots, G_t$, then $p(G; \lambda) = \prod_{i=1}^{t} p(G_i; \lambda)$.*

Proof. If G has a loop, then $p(G; \lambda) \equiv 0$ and there is nothing to prove. Thus suppose that G does not have a loop. We use induction on the number of edges. If $|E| = 0$ and $|V| = p$, then we can color the vertices arbitrarily, so that $p(G; \lambda) = \lambda^p$. Now suppose that $e = uv \in E$. We construct two new graphs G'_e and G''_e. G'_e has the vertices and edges of G with the exception of the edge e, which we remove. In G''_e we identify the vertices u and v to a vertex uv and join uv to all vertices which are adjacent to u or v, whereby we only draw one edge, in case multiple edges arise (see Figure 3.6).

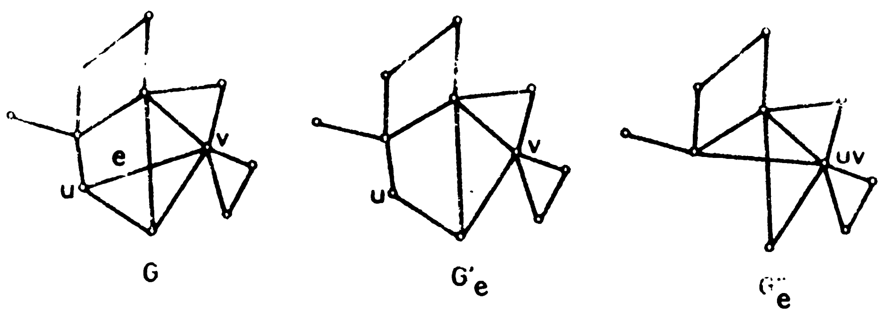
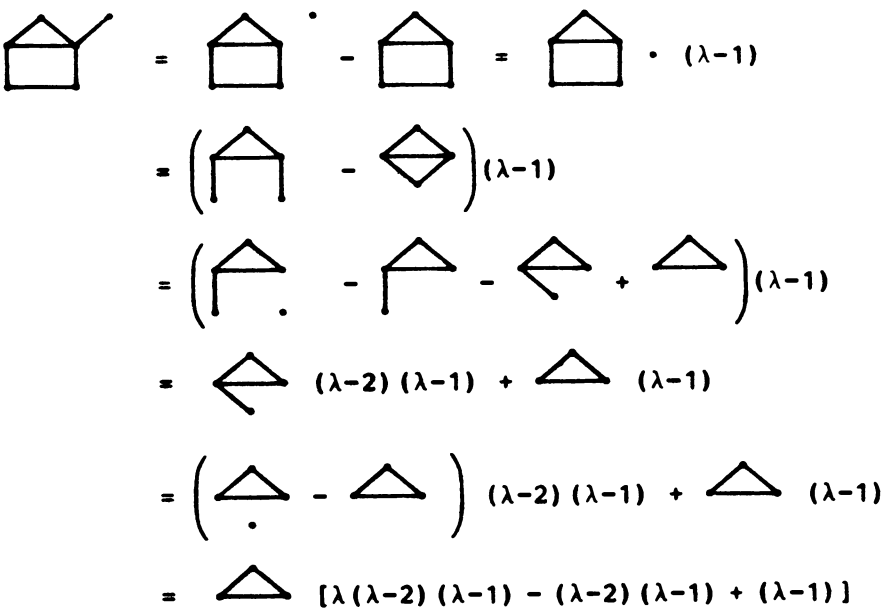


Figure 3.6

Let us consider the colorings of G'_e. They split into two disjoint classes, depending on whether u and v are given different colors, or the same color. In the first case we obtain exactly the colorings of G, in the second case we obtain an obvious one to one correspondence with the colorings of G''_e. In summary then we have $p(G'_e;\lambda) = p(G;\lambda) + p(G''_e;\lambda)$ or $p(G;\lambda) = p(G'_e;\lambda) - p(G''_e;\lambda)$. Since both graphs G'_e and G''_e have fewer than q edges, we can apply induction, from which the conditions (i) through (iii) follow without difficulty. The last condition (iv) is clear, since the coloring of the individual components may be performed independently of each other. ■

In principle, 3.4 permits the step-by-step computation of each chromatic polynomial, and thus also the determination whether $p(G;4) > 0$ holds. It is clear, however, that this inductive computation will in general be no more efficient than an ad hoc coloring.

EXAMPLE. *Let us perform the algorithmic step of the theorem in symbolic form (whereby we identify graphs and chromatic polynomials) until we have decomposed the graph into small enough graphs all of whose chromatic polynomials we know. E.g.* $p(K_p;\lambda) = \lambda(\lambda-1)(\lambda-2)\cdots(\lambda-p+1)$, *since we can color the first vertex with* λ *colors, after a color is chosen, the second vertex with* $\lambda - 1$ *colors, etc. In addition, we will make several applications of* 3.4(iv).



$$= \lambda(\lambda-1)(\lambda-2)\,.\,(\lambda-1)(\lambda(\lambda-2) - \lambda + 2 + 1)\ .$$

Hence $p(G;\lambda) = \lambda(\lambda-1)^2(\lambda-2)(\lambda^2-3\lambda+3)$. *For example then* G *has* $p(G;3) = 36$

colorings with 3 *colors and* $p(G;4) = 504$ *colorings with* 4 *colors. The reader may verify the result* $p(G;4) = 504$ *in order to appreciate the power of the algorithm.*

In his next paper Birkhoff took up Kempe's idea about the reducibility of maps. We recall that a configuration is called **reducible** if it can not appear in a hypothetical 5-chromatic map. The first 3 configurations in Figure 1.4 are reducible. We will call a 5-chromatic map $\mathcal{L}$ that has the smallest number of countries, **irreducible**. It was Birkhoff's hope to prove so many properties of an irreducible map $\mathcal{L}$, that finally $\mathcal{L}$ could be explicitly constructed, or determined to be impossible.

What are some properties of an irreducible map $\mathcal{L}$ that can be immediately asserted? We can always assume that $\mathcal{L}$ is normal. Since the three configurations previously mentioned are reducible, every country in $\mathcal{L}$ has *at least* 5 *borders*. Furthermore $\mathcal{L}$ must be *cubic*. For if $v \in V$ has degree $t \geq 4$, then there must be at least one pair A_i, A_j of countries which touch v, but which are not neighboring. If we merge A_i and A_j as shown in Figure 3.7, then we have a new map $\mathcal{L}'$ with one fewer country. Thus $\mathcal{L}'$ is 4-colorable, and hence also $\mathcal{L}$, since after reseparation of A_i and A_j we can choose the same color for A_i and A_j.

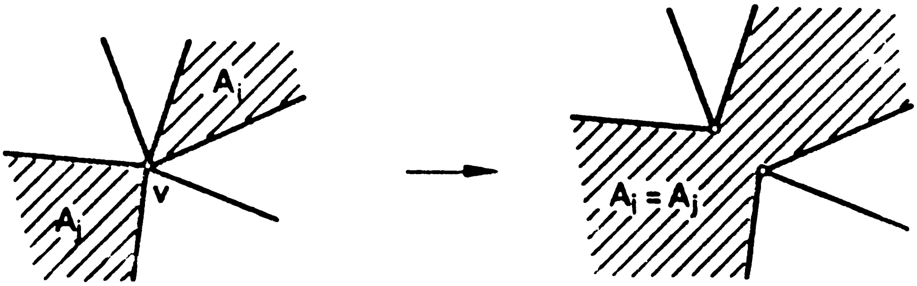


Figure 3.7

Birkhoff's most important new idea was the concept of a **ring** of countries. A subset R' of countries form a ring, if R' forms a circle in $\mathcal{L}$ which divides the plane into a nontrivial interior $\mathcal{L}_1$ and exterior $\mathcal{L}_2$ (where nontrivial means that $\mathcal{L}'$ is not just a set of countries incident with a single vertex v). Figure 3.8 shows a ring with 5 countries.

Birkhoff proved that an irreducible map can contain no ring with 4 or fewer countries, and that a 5-ring can only exist if the interior (or exterior) consists of a single country with 5 borders. 6-rings are already more complicated, but Birkhoff already gave an example of a reducible 6-ring that is pictured in Figure 3.9.

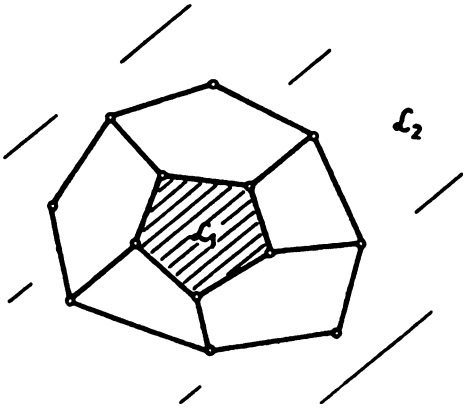


Figure 3.8

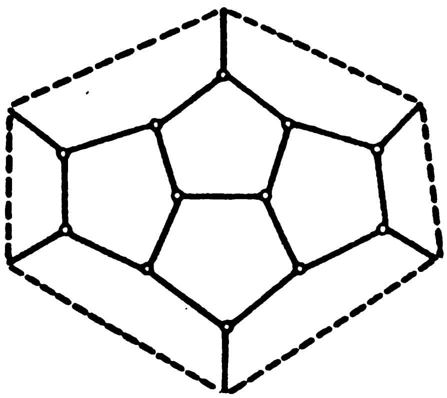


Figure 3.9

In 1.6 we introduced the formula $\sum_{i\geq 12}(6-i)p_i = 12$ for each normal cubic map, where p_i is the number of countries with exactly i borders. In particular then this relation holds for irreducible maps. It immediately follows that $p_5 = 12 + \sum_{i\geq 12}(i-6)p_i \geq 12$ (since $p_i = 0$ for $i \leq 4$) and hence $|R| \geq 12$. An irreducible map thus must contain at least 12 countries.

Using Birkhoff's idea about rings, Franklin proved in 1933 that an irreducible map must contain at least 26 countries. Through the study of larger and larger rings, this number, soon called the **Birkhoff number** b, was pushed up further and further, as the following table displays:

Reynolds (1926)	$b \geq 28$
Franklin (1938)	$b \geq 32$
Winn (1940)	$b \geq 36$
Ore-Stemple (1968)	$b \geq 41$
Mayer (1978)	$b \geq 96$.

The last work of Mayer (the same professor of literature) already came too late since the 4-color problem had already been solved. It was less the Birkhoff number itself but rather the methods for reducibility that proved to be of lasting interest. Here, however, the real breakthrough did not come before the 1960's, primarily through the work of Heesch. Consequently, we will consider this next to the last chapter in the story of the 4-color problem at exactly that place in the book (Chapter 9).

The last two new developments which arose while seeking a solution to the 4-color problem had their origins in two conjectures of Tait, each of which turned out to be false.

We recall Tait's Theorem 1.7: *A bridgeless cubic map is 4-colorable if and only if its edges are 3-colorable.* We assume that the edges are colored red, blue and white. The subgraph G_R of $G(\mathcal{L})$, which consists only of the red edges, has the property that each vertex in G_R has degree 1, and the same holds for the subgraphs G_B and G_W of the blue, resp., white edges. In general, if a graph has the property that all the vertices have the same degree r, then we call G **r-regular**. Cubic just means that G is 3-regular. An edge coloring of $G(\mathcal{L})$ then determines a decomposition of $G(\mathcal{L})$ into three disjoint 1-regular subgraphs, and conversely each such decomposition gives an edge coloring with 3 colors. Since Tait was of the opinion that the 4-coloring problem was solved and thus that his version 1.7 was correct, he was naturally convinced that every 3-regular bridgeless plane graph is decomposable in that way. Indeed, he went so far as to regard this decomposition as proven for *arbitrary* 3-regular graphs without bridges. That this was false was shown (see below) by Petersen in 1898 with an example. Despite this error, the idea of the **factorization** of graphs proved to be an important concept whose beginnings we will now sketch.

DEFINITION. *Let $G(V,E)$ be an arbitrary graph. An* **r-factor** *of G is an r-regular subgraph $H(V,E')$ on the same vertex set V, where H is a proper subgraph of G. We say that G is* **r-factorable** *if G can be decomposed into r-factors $H_i(V,E_i)$ such that each edge of E is in exactly one factor H_i, i.e. $E=\bigcup_i E_i$ is a disjoint union.*

For example if $G(V,E)$ has a 1-factor, then $|V|$ must clearly be even. Hence the complete graphs K_{2n+1} can not have a 1-factor. On the other hand, each complete graph K_{2n} is 1-factorable, as the reader can easily convince himself. Figure 3.10 shows a 1-factorization of K_6.

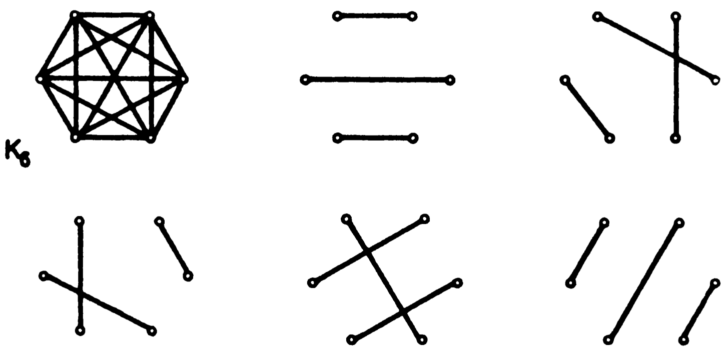


Figure 3.10

Naturally, if a graph can be decomposed into disjoint factors then it must be regular. If G is r-regular, then in the extreme case, we can hope that G decomposes into r 1-factors, and thus is 1-factorable. This means then, as we have seen, that the edges can be colored with r colors.

Petersen, in two works in 1891 and 1898, took up the problem of factorization. First he considered the case of a 2-regular graph. Each such graph decomposes into disjoint circuits, and since a circuit obviously splits into two 1-factors if and only if it has even length, we note as our first result:

3.5 THEOREM. *A 2-regular graph G is 1-factorable if and only if G consists of disjoint circuits of even length.*

With the help of 3.5 Petersen arrived at a proof of the 2-factorability of all regular graphs of even degree—we will come back to this in Chapter 6. If the degree r is odd, then things are more difficult. On the basis of Tait's Theorem 1.7, it is of course the case $r = 3$ that is of particular interest. By 1.2, a 3-regular graph G has an even number of vertices. In G there are only two kinds of factors, 1-factors and 2-factors. If G has a 1-factor, then the complementary edge set forms a 2-factor, and conversely. In order to prove that G is 1-factorable we must therefore first show the *existence of a 1-factor* and secondly show that among all 1-factors there is at least one whose complementary 2-factor decomposes into *circuits all with even length.* If we pick the 1-factor in the graph of Figure 3.11, as in (i), then the complement consists of two circuits of length 3, and thus is not decomposable further. However, if we choose the 1-factor as in (ii), then we obtain as complement a circuit of length 6, and hence a decomposition into 1-factors.

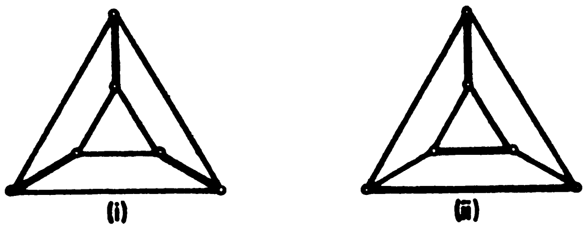


Figure 3.11

Petersen completely solved the first problem, the existence of a 1-factor. Since 1-factors can be formed in the individual components, we can restrict our attention to connected graphs. The 3-regular graph in Figure 3.12 has no 1-factors, since any such factor must include exactly one of the 3 edges in the middle, say e. Then, however, the subgraph spanned by $\{u, v, w\}$ would no longer be 1-factorable.

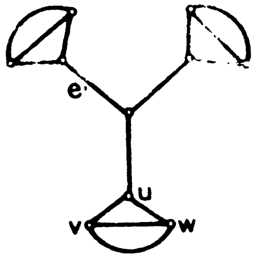


Figure 3.12

Clearly it is the 3 middle edges, all of which are bridges, which account for the difficulty. However, if we exclude bridges then we have:

3.6 THEOREM (Petersen). *A connected 3-regular graph without bridges always has a 1-factor.*

Proof. We apply induction on the number of vertices. The smallest graph of that type has 2 vertices with 3 connecting edges and trivially has a 1-factor. The essential tools for the induction step are contained in the following lemma which can be verified by a somewhat lengthy argument by considering various cases: Let $e = uv \in E$ not be a multiple edge, i.e. there are no other edges from u to v. Furthermore, let $a = ux$, $b = uy$ be the other edges incident with u and $c = vz$, $d = vw$ the other edges incident with v. We construct two new graphs G_1 and G_2, in which we remove the vertices u and v and the edges a, b, c, d, and e and insert two new edges xz, yw, resp. xw, yz. Then either G_1 or G_2 is again a connected bridgeless 3-regular graph. (See Figure 3.13.)

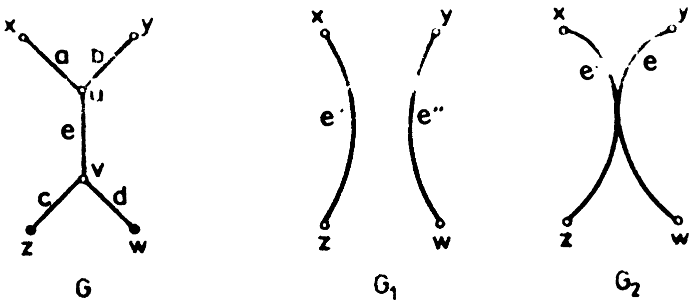


Figure 3.13

Let G be a 3-regular graph that has a 1-factor. We color the edges of the 1-factor red and the edges of the complementary 2-factor blue. A circuit in G whose edges alternately are colored red and blue is called an **alternating** circuit. Claim: Each edge of G is contained in an alternating circuit. It is sufficient to show this for blue edges, since the adjacent red edge likewise then lies in the circuit. Assume that G is the smallest graph of that type for which the claim does not hold, and the blue edge $f = tx$ is not in an alternating circuit. G has at least

3 vertices. Let $a = xu$ be the other blue edge at x and $e = uv$ the red edge at u. In Figure 3.14 the blue edges are solid and the red edges dotted. By construction, the edge f is distinct from a and e. Assume $f = b$, i.e. a and f are both edges between x and u. Let $g = xs$ be the red edge at x. We remove the vertices x and u, the two edges a and f, as well as g and e and insert a new red edge $r = sv$. The new graph satisfies all the conditions and has two fewer edges. By the choice of G, each edge including the new edge is in an alternating circuit. In this circuit we can replace r by the red-blue-red segment g,f,e by which f is contained in an alternating circuit of G. The cases $f = c$, resp., $f = d$ as well as the case where e is a multiple edge can be handled analogously. Thus we assume finally that f is different from all other edges in Figure 3.14 and that e is not a multiple edge. One of the graphs G_1 or G_2 of Figure 3.13, say G_1, then satisfies all the conditions of the theorem in which we carry over the coloring to G_1 and color xz, yw blue. Because of the minimality of G there now exists an alternating circuit in G_1 that contains f. This circuit can not contain the new blue edge $e' = xz$. If it contains $e'' = yw$, then we replace e'' in G by the edge sequence b,e,d so that in each case there results an alternating circuit C in G with $f \in C$, which contradicts the assumption.

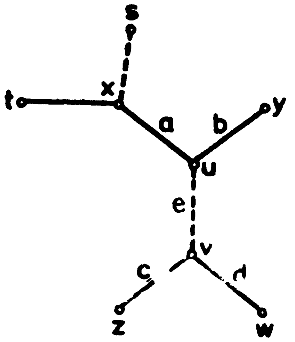


Figure 3.14

We come to the induction step. Assume the theorem is true for all graphs with fewer than p vertices. Suppose G has $p \geq 3$ vertices and satisfies the conditions of the theorem. G certainly has an edge that is not a multiple edge, for otherwise, since $p \geq 3$, there would be a vertex of degree ≥ 4. Let e be such an edge. We construct the graphs G_1 and G_2 in accordance with Figure 3.13. By the induction hypothesis one of the two, say G_1, has a 1-factor. We color the edges of G_1 red and blue as above. If both edges e' and e'' are colored blue, then we color the edges a,b,c,d in G blue and e red. If e' and e'' are colored differently, e.g. e' red and e'' blue, then in G we color the edges a and c red and b,d and e blue. Finally, if e' and e'' are both colored red, then by the above proof e' is in an alternating circuit of G_1. If we interchange the colors in this circuit, then again we obtain a red-blue coloring, where e' is now colored blue. Depending on whether e'' was in C or not, we are just in one of the two cases treated first. Hence we can proceed as before.Thus in all cases we obtain a proper red-blue-coloring of G, i.e. G does indeed decompose into a 1-factor and a 2-factor. ∎

3.9 TAIT'S CONJECTURE. *Every 3-regular polytopal graph has a Hamiltonian circuit.*

Before sketching the history of this conjecture, and showing that 3.9 indeed implies the 4-color conjecture, we will present a few basic concepts about polytopes.

A set $S \subseteq \mathbb{R}^3$ is called **convex**, if whenever two points P,Q are in S, then so is the entire line segment joining them. Intuitively speaking, convexity means that every point in a convex set can be "seen" from every other point—a natural, aesthetic principle. The 1-dimensional convex sets are clearly the intervals. Figure 3.18 shows three 2-dimensional sets and one 3-dimensional convex set.

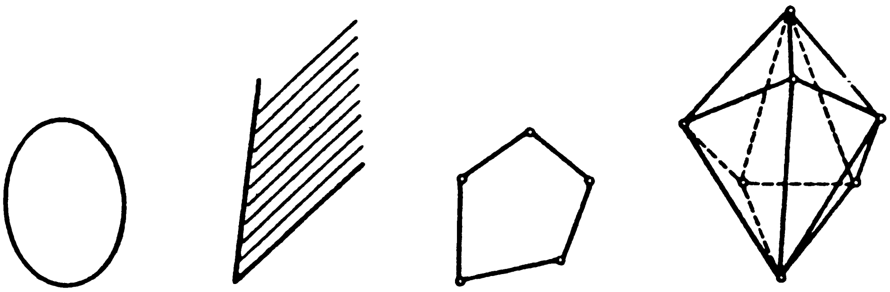


Figure 3.18

All four sets are closed (i.e. contain their boundary points), however the second through the fourth exhibit an essential difference with the first: They contain points that intuitively we would immediately designate as vertices. If we consider the vertices for a moment, then we recognize that they are distinguished from all other points in that they are not in the interior of a line segment that lies in the convex figure. Equivalently: Upon removal of the vertices, the remaining set is again convex. The first figure in Figure 3.18 also exhibits vertices by this definition, each of the (infinitely many) boundary points is a vertex. We note another fact: The second set in Figure 3.18 has only one vertex, and in contrast to the last two examples, is unbounded. With that we arrive at the main concept.

DEFINITION. *A* **convex polytope** *in $\mathbb{R}^n$ is a compact convex set with finitely many vertices.*

The last two sets in Figure 3.18 are polytopes. The pentagon is a 2-dimensional polytope (it should be clear that in the 2-dimensional case we obtain in general just the polygons). The double pyramid is a 3-dimensional polytope. Besides the **vertices**, a polytope contains **edges** and in the 3-dimensional case **faces**. The displayed double pyramid for example has 7 vertices, 15 edges, and 10 faces.

Each polytope S has a corresponding **skeleton** $G(S)$, namely the graph consisting of the vertices and edges of S with the natural incidence. Hence we come to the final definition.

DEFINITION. *A graph G is called* **polytopal** *if G is isomorphic to the skeleton graph of a 3-dimensional convex polytope.*

What do polytopal graphs have to do with the 4-color problem? First all of these graphs are **planar**. One sees this by taking a point z in a neighborhood of a face F and then looking into the interior of the polytope, using F as a "window". In this way we project all the points of S onto F. Thus the skeleton of S corresponds to a graph G in the plane determined by F. Furthermore, if z is sufficiently close to F, no edge crossings arise. Thus G is planar. The surfaces of S correspond to the countries of G, whereby the window assumes the role of the outer country. In Figure 3.19 such a plane projection of the cube is pictured.

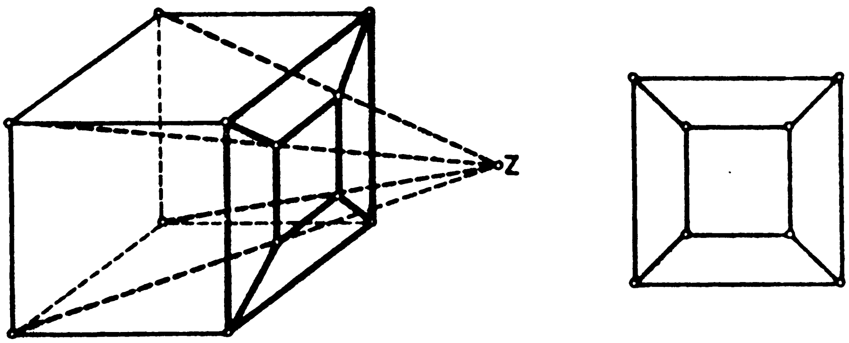


Figure 3.19

Secondly, all polytopal graphs have the property mentioned above, that the removal of 2 vertices never separates the graph. We will introduce this separation property in general. Still another word about terminology: When we say, we **remove** the vertex v from the graph G, we mean, of course, that we remove both v and all the edges incident with v. For brevity we will denote the resulting graph by $G - v$ or more generally by $G - A$ when the vertex set $A \subseteq VE$ is removed.

DEFINITION. *The* **connectivity number** $\kappa(G)$ *of a graph G is the smallest number of vertices whose removal disconnects the remaining graph. Since for complete graphs such separating vertex sets do not exist, we define* $\kappa(K_p)$ *to be* $p-1$. *If* $\kappa(G) \geq n$, *then we say that G is* **n-connected**.

A disconnected graph G then has connectivity number $\kappa(G) = 0$. If $\kappa(G) = 1$ and v is a separating vertex, then we call v a cut-vertex. The graph in Figure 3.20 has u,v and w as cut-vertices.

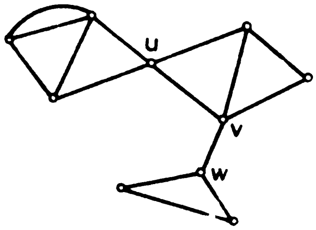


Figure 3.20

From the definition of a cut-vertex it is not hard to obtain a characterization of 2-connected graphs, whose proof will be left for the reader.

3.10 LEMMA. *A graph is 2-connected if and only if any two vertices (or equivalently any two edges) lie on a common circuit.*

A **block** of $G(V, E)$ is a maximally connected subgraph with at least 2 vertices, that does not contain any cut-vertices. A block then is a single edge, or is 2-connected. From 3.10 it follows that the blocks induce a disjoint decomposition of the edges in E.

From geometric considerations one can observe that the skeleton of each convex polytope is a 3-connected graph. What is even more interesting, however, is the fact that this connectivity property together with the planarity shown above characterize the polytopal graphs. The proof of this theorem, stated by Steinitz in 1922, merely involves elementary geometry. However, it would take us too far afield to do it, so we will be satisfied with the statement of the theorem. For those who want to learn more about convex sets and in particular the proof of the theorem, the book by Grünbaum is recommended.

3.11 THEOREM (Steinitz). *A graph G is polytopal if and only if it is planar and 3-connected.*

Let us conclude the history of Tait's conjecture, somewhat anticipating the later chapters. In 1931, Whitney showed, that for the proof of the 4-color conjecture one may restrict oneself to cubic maps $\mathcal{L}$ whose skeleton is 3-connected (see Theorem 7.1). Since by the theorem of Steinitz, $G(\mathcal{L})$ is a polytopal graph, Tait's conjecture 3.9, would, because of 3.8, therefore imply the 4-color theorem. Tait himself was apparently not completely aware of this connection. In any case the discussion about that became academic, when Tutte in 1946 constructed a polytopal 3-regular graph with 46 vertices which did not contain a Hamiltonian circuit. Subsequent authors have found smaller counterexamples of which the smallest currently has 38 vertices. That the graph in Figure 3.21 does not contain a Hamiltonian circuit can be seen by noticing that every path that goes through each of the vertices of a subgraph G_i exactly once must have a vertex v_i as first or last vertex. The continuation, however, can only continue through u, so that one gets stuck at one of the three vertices v_1, v_2, v_3.

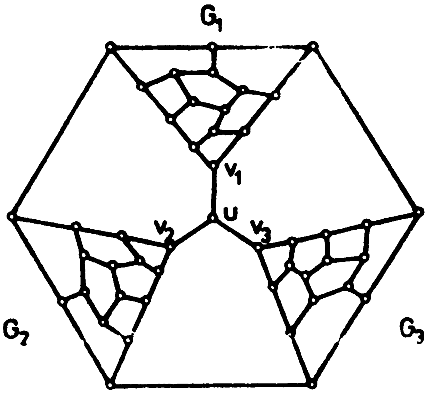


Figure 3.21

Around 1930 the study of graphs had outgrown its beginnings. Stimulated by the ever present challenge of the unanswered 4-color problem, 5 disciplines were founded, along the lines exhibited in this chapter, namely: **planarity** and more generally embedding questions, **coloring** of graphs, **factorization**, **Hamiltonian graphs**, and finally representation of graphs by matrices and more generally **matroids**. Related questions concerning connectivity and enumeration of graphs were added which we will just touch upon lightly. So, from a collection of isolated results a whole new theory, graph theory, had emerged which is documented by the fact that the first textbook on graph theory by D. König appeared in 1936. The first years of graph theory were dominated by H. Whitney, a young American mathematician who obtained fundamental results in almost all the disciplines mentioned above, which quickly pushed the new theory along. Although after a few years he again left graph theory (as is said, partly because of the hopelessness of solving the 4-color problem), his achievements remain landmarks, equalled later only by the work of W. T. Tutte. Thus the following five chapters are in large part a discussion of the ideas and results of these two mathematicians.

EXERCISES FOR CHAPTER 3

1°. Complete the second part of the proof of 3.1.

2. Construct the function $h : E \to \{1, -1\}$ according to 3.1 for the map in Figure 3.16.

3. Find the vertex-edge-, and country-edge-incidence matrices for the Platonic graphs.

4°. Show that a connected Euler graph has an Euler trail.

5. If G contains vertices of odd degree, then by exercise 4, G can not contain an Euler trail. Assume G is connected and has exactly 2 vertices of odd degree. Show that in this case a (not closed) edge trail exists which passes through each edge exactly once. What can be said when G has exactly $2h$ vertices with odd degree?

6. For which values of n are the complete graphs K_n and cubes Q_n Euler graphs? Which Platonic graphs are Euler?

7°. Describe the set of edges in the graph G that corresponds to $a \in \hat{A}$, A=incidence matrix.

8. Dualize Theorem 3.3.

9. Verify again the recursion for chromatic polynomials: $p(G; \lambda) = p(G'_e; \lambda) - p(G''_e; \lambda)$ and fill in all the details of the proof of 3.4.

10°. Compute the chromatic polynomial: a) of the circuit C_n, b) of the wheel W_n, c) of the Petersen graph.

11. For a simple graph G, the coefficient $a_1 = -q$ (cf. 3.4). What does a_1 look like when G contains multiple edges?

12. Show by a Kempe argument that an irreducible map can not contain a ring with 4 or fewer countries.

13°. Show that K_{2n} is always 1-factorable.

14. Verify the lemma at the beginning of the proof of 3.6.

15. Show that the Petersen graph does not have a Hamiltonian circuit.

16. Solve the problem of the knight's moves and discuss the analogous problems for the queen, bishop, etc. Does there also exist a solution if the knight is placed on a 7×7 board?

17°. Prove Lemma 3.10.

18. Determine the connectivity number κ for: a) the wheel W_n, b) the Petersen graph, c) the cubes Q_3, Q_4.

19. Computer problem. Assume G is Euler. Devise an algorithm which yields an Euler trail.

PART II: THEME

4. PLANARITY

Starting with this chapter, we always consider graphs instead of maps. In Chapter 2 (Theorem 2.8) we saw that the 4-color conjecture is equivalent to the following conjecture: *Every planar graph is 4-colorable.* It follows from this that the most immediate problem is a *characterization of planar graphs.*

We mentioned the Möbius puzzle in previous chapters. Its dual formulation reads as follows: Is it possible to choose five points in the plane and to connect them by edges so that no two edges cross? Stated tersely: Is the graph K_5 planar? Another well-known puzzle (mentioned, e.g., in the famous collection of puzzles by Dudeney 1917) is the "gas-water-electricity problem": Given three houses and three stations for gas, water and electricity, is it possible to lay conduits from each station to each house in such a way that no two conduits intersect? In graph theory, it is now customary to call the graph formed from two disjoint vertex sets V_1 and V_2, with $|V_1| = m$ and $|V_2| = n$, and in which the edges connect all of the vertices of V_1 with all of those of V_2, the **complete bipartite graph**, $K_{m,n}$. Hence, the puzzle asks: Is $K_{3,3}$ planar?

We can give the answer easily with the help of Euler's formula 1.4 or by means of 1.5.

4.1 LEMMA. *K_5 and $K_{3,3}$ are not planar.*

Proof. We have already considered this for K_5 at the end of 1.5. Assume that $K_{3,3}$ is planar and that $G(V, E, R)$ is a plane graph that is isomorphic to $K_{3,3}$. Since $|V| = 6$ and $|E| = 9$, we have $|V| = \frac{2}{3}|E|$. Furthermore, since $K_{3,3}$ and hence G does not contain bridges and circuits of length 2 or 3, we have that $2|E| \geq 4|R|$. It follows from Euler's formula that

$$2 = |V| - |E| + |R| \leq \frac{2|E|}{3} - |E| + \frac{|E|}{2} = \frac{|E|}{6},$$

i.e., $|E| \geq 12$; which contradicts $|E| = 9$. ∎

Considering the ease with which the nonplanarity of K_5 and $K_{3,3}$ was proven, it is all the more surprising that these two graphs alone are responsible for nonplanarity in an arbitrary graph. First, the following clearly holds:

4.2 LEMMA. *If G is planar, then so is every subgraph of G.*

Therefore, certainly neither K_5 nor $K_{3,3}$ can be a subgraph of a planar graph, but the converse is also valid. That is the content of the famous Kuratowski Theorem of 1930: If G is nonplanar, then G must contain, in a sense to be made precise, either K_5 or $K_{3,3}$.

We shall explain this "containment" shortly. Let G be an arbitrary graph, and let $e = uv$ be an edge of G. We say that we **subdivide** the edge e if we replace e by a path $u, u_1, \ldots, u_t, v$, i.e., in the new graph G' all of the added vertices u_i have degree 2. A graph H is a **subdivision** of G if H arises by the subdivision of certain edges of G. Since vertices of degree 2 obviously have no influence on embeddability, we can say:

4.3 LEMMA. *A graph is planar if and only if each of its subdivisions is planar. Therefore, in particular all subdivisions of K_5 and $K_{3,3}$ are nonplanar.*

EXAMPLE. *By using the concept of subdivision we can immediately verify the nonplanarity of the Petersen graph, P. As one can see in Figure 4.1, P contains a subdivision of $K_{3,3}$ as a subgraph—and thus by 4.1 through 4.3 is nonplanar. The vertices of the first set are designated by* I *and those of the second set by* II.

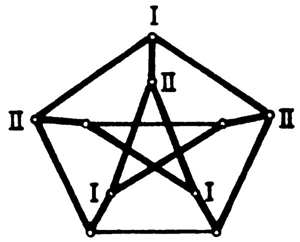


Figure 4.1

Before proving Kuratowski's theorem: G is planar if and only if G contains no subdivision of K_5 or $K_{3,3}$, we will discuss a no less famous theorem, Menger's Theorem. Historically this precedes Kuratowski's Theorem and makes its proof substantially more transparent.

A vertex set T of a graph $G(V, E)$ is called a **separating vertex set** if the graph $G - T$ is not connected. In the preceding chapter we defined the **connectivity number** $\kappa(G)$ to be the size of a smallest separating vertex set. Therefore, $\kappa(G)$ is defined to be $\min\{|T| : T \subseteq V$ a separating vertex set$\}$. If u and v are nonadjacent vertices in G, then we understand the **local connectivity number** $\kappa_G(u, v)$ to be the size of a smallest vertex set T that separates u from v, i.e. u and v must lie in different components of $G - T$. Consider the graph in Figure 4.2. Here, $\kappa_G(u, v) = 1$, but $\kappa_G(u, w) = 2$.

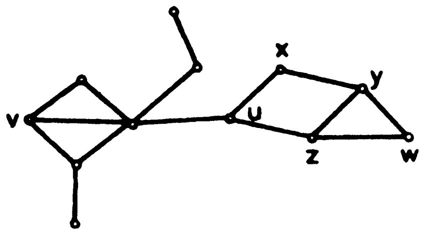


Figure 4.2

We call two paths that connect the vertices u and v **disjoint** if they have no points in common other than u and v. In Figure 4.2 e.g., (u, x, y, w) and (u, z, w) are disjoint u,w-paths. We understand by the **joining number** $\mu_G(u, v)$ the maximal number of pairwise disjoint u,v-paths. For nonadjacent vertices u,v it is obvious that $\kappa_G(u, v) \geq \mu_G(u, v)$, since each u,v-separating vertex set must clearly intersect each of the $\mu_G(u, v)$ disjoint paths. Otherwise u and v would still be joined by a path after removal of the separating vertices. That equality in fact holds is the content of the following theorem.

4.4 THEOREM (Menger). *Let u and v be two nonadjacent vertices of the graph G. Then* $\kappa_G(u, v) = \mu_G(u, v)$.

Proof. By the remark just made, it suffices to verify that $\kappa_G(u, v) \leq \mu_G(u, v)$. That is, we must show: If we need h vertices to separate u and v, then there are h disjoint u,v-paths. This is clear for $h = 0,1$. Assume the assertion is false; then let $h > 1$ be the smallest number for which a counterexample exists. Among all these counterexamples we choose one having a smallest number, p, of vertices and from among all counterexamples with p vertices, we select a graph G with a minimal number of edges. Therefore, in G there are two nonadjacent vertices u and v for which $\kappa_G(u, v) \geq h$, but $\mu_G(u, v) < h$. We collect a few properties of G:

(a) *No vertex in G is adjacent to both u and v.* For if w were such a vertex, then since $G - w$ contains fewer vertices than G, we would have $\kappa_{G-w}(u, v) = \mu_{G-w}(u, v) = h - 1$. We could add the path (u, w, v) to the $h - 1$ disjoint u,v-paths in $G - w$, which contradicts $\mu_G(u, v) < h$.

Let e be any edge in G. If we remove e, then the connectivity number $\kappa_G(u, v)$ can decrease by at most 1. But by the choice of G, it must indeed decrease. Therefore we have $\kappa_{G-e}(u, v) = \mu_{G-e}(u, v) = h - 1$. Thus, to each $e \in E$, there exists a u,v-separating vertex set $T(e)$ in $G - e$ with $|T(e)| = h - 1$. Let $e = ab$ and $a \neq u$, $a \neq v$. Since $T(e)$ does not separate the vertices u and v in G, there must be a u,v-path in G that does not meet $T(e)$. But, naturally, this path must contain e (for otherwise $T(e)$ would not be a u,v-separating vertex set in $G - e$). Therefore we have the following:

(b) $T(e) \bigcup \{a\}$ *is a u,v-separating vertex set in G.*

Let $W = \{w_1, \dots, w_h\}$ be an arbitrary (minimal) u,v-separating vertex set in G. We understand by a u,W-path a path from u to a vertex w_i of W that contains no vertices of W, except w_i. A W,v-path is defined analogously. Let P_u and P_v be the sets of all u,W-paths, resp., all W,v-paths. Each vertex w_i occurs in at least one path of P_u and in at least one path of P_v, for otherwise $W - \{w_i\}$ would already separate u from v. But a path from P_u can have no vertices in common with a path from P_v with the exception of the possible intersection point in W. Otherwise, there would be a u,v-path that does not intersect W at all. Finally, we assert that either P_u consists of exactly the paths $(u, w_1), \dots, (u, w_h)$, hence of individual edges, or P_v consists of exactly the edges $(w_1, v), \dots, (w_h, v)$. For if not, then P_u together with the edges $w_1v, \dots, w_hv$ would be a graph G_1 with $\kappa_{G_1}(u,v) = h$ and similarly P_v together with $uw_1, \dots, uw_h$ would be a graph G_2 with $\kappa_{G_2}(u,v) = h$. Since G_1 as well as G_2 contain fewer vertices than G, we would obtain h disjoint u,v-paths in G_1 as well as in G_2. But by combination of the P_u-, resp., the P_v-part, we would also have h disjoint u,v-paths in G. Thus, we note:

(c) *If $W = \{w_1, \dots, w_h\}$ is a u,v-separating vertex set, then either u is adjacent to all w_i or v is adjacent to all w_i.*

We now arrive at our conclusion. Let $P = (u, a_1, a_2, \dots, v)$ be a u,v-path of shortest length. Then, by (a), $a_1 \neq u$, $a_1 \neq v$, and $a_iv \notin E$. Similarly (by (a)) $a_2 \neq v$, $a_2 \neq u$, and $ua_2 \notin E$ by the choice of P. Let $e = a_1a_2$ and $T(e) = \{t_1, \dots, t_{h-1}\}$ be a u,v-separating vertex set in $G - e$. By (b), $T(e) \bigcup \{a_1\}$ as well as $T(e) \bigcup \{a_2\}$ separate the vertices u and v in G. Since $a_1v \notin E$ and $ua_2 \notin E$, it follows from (c) that u as well as v are adjacent to all the vertices $t_1, \dots, t_{h-1}$. Then, since $h - 1 \geq 1$, there is at least one vertex t_i that is adjacent to u and v, which contradicts (a). ∎

Figure 4.3 shows an example for 4.4. Here $\kappa_G(u,v) = \mu_G(u,v) = 3$.

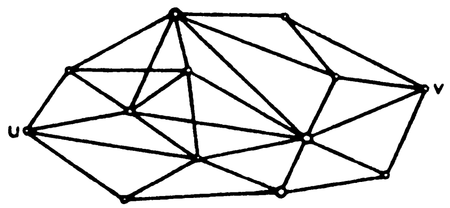


Figure 4.3

We shall return to the many implications of this theorem in Chapter 6. For the moment we only note the following consequence.

4.5 COROLLARY (Whitney). *A graph G with at least two vertices is n-connected if and only if every two vertices are joined by at least n disjoint paths.*

Proof. The theorem clearly holds for complete graphs. Suppose G is not complete. If every two vertices are joined by n disjoint paths, then surely $\kappa(G) \geq n$. Conversely, let $\kappa(G) \geq n$. Since clearly $\kappa(G) = \min \kappa_G(u,v)$ over all nonadjacent vertex pairs u,v, then by Menger's Theorem we have that $\mu_G(u,v) = \kappa_G(u,v) \geq \kappa(G) \geq n$ for all nonadjacent pairs u,v. It remains to show that $\mu_G(u,v) \geq n$ also holds for all edges $uv \in E$. The easy proof is left to the reader. ∎

Now we go back to our real theme, planarity. We make a preliminary remark before proceeding to the proof of Kuratowski's Theorem: If G is a plane 2-connected graph, then every country of G is bounded by a Jordan curve that consists entirely of edges of G. This follows directly from the Jordan curve theorem. We can now embed G so that an arbitrary country F is the outer country. For this, we merely need to map G by a stereographic projection onto the sphere, choose a point in the interior of F as "North Pole", and then project again to a plane that is tangent to the South Pole. Therefore, G can in particular be embedded so that a prescribed edge, resp., a prescribed vertex lies on the boundary of the outer country.

4.6 THEOREM (Kuratowski). *A graph G is planar if and only if G contains no subgraphs that are isomorphic to a subdivision of K_5 or $K_{3,3}$.*

Proof. Only one implication need be shown (cf. 4.3). For brevity, we say that G is in the class Π if G contains no subdivision of K_5 or $K_{3,3}$. Thus, we must show that $G \in \Pi$ implies G is planar. To this end, we use induction on the number of vertices. Every graph with at most four vertices is planar since K_4 is planar. Within the class of graphs with p vertices, we use induction on the number of edges, q. For $q = 0$, the graph is trivially planar. Hence, suppose the assertion holds for all graphs with at most p vertices, and fewer than q edges. Now let $G \in \Pi$ be a graph with $p \geq 5$ vertices and q edges. We split the proof into several cases according to the connectivity number of G.

(a) $\kappa(G) = 0$. In this case G is not connected. Since by the induction assumption each component can be embedded in the plane, this is naturally also true for G.

(b) $\kappa(G) = 1$. If v is a cut vertex, then we can decompose $G(V,E)$ into two edge-disjoint graphs $G_1(V_1,E_1)$, $G_2(V_2,E_2)$ with $V_1 \bigcup V_2 = V$, $V_1 \bigcap V_2 = \{v\}$, and $E = E_1 \bigcup E_2$. Clearly, $G_1, G_2 \in \Pi$. By the induction assumption and our remarks above we can embed G_1 and G_2 in the plane so that v is a boundary point of the outer country. Now it is clear that the two planar graphs G_1 and G_2 (perhaps by a necessary continuous map) can be joined at v so that the composite graph G is also planar.

(c) $\kappa(G) = 2$. Let $\{u,v\}$ be a separating vertex set. Then G decomposes into two graphs $G_1(V_1,E_1)$ and $G_2(V_2,E_2)$ with $V_1 \bigcup V_2 = V$, $V_1 \bigcap V_2 = \{u,v\}$, $E = E_1 \bigcup E_2$, and $E_1 \bigcap E_2 = \{uv\}$, in the case where $e = uv$ occurs in G. Since G is 2-connected, u and v are joined in each of the graphs G_i by a path W_i. From this it follows that $G_i \bigcup e$, $i = 1,2$, contains neither a subdivision of K_5 nor of $K_{3,3}$. Otherwise, $G_i \bigcup W_j \subseteq G$ $(j \neq i)$ would contain such a subdivision, in contradiction with the assumption $G \in \Pi$. Thus, by the induction hypothesis, $G_1 \bigcup e$ and $G_2 \bigcup e$ are embeddable in the plane, and indeed so that e lies on the outer boundary. There we can now join $G_1 \bigcup e$ and $G_2 \bigcup e$ to form $G \bigcup e$. Now discard e in case this edge was not originally present.

(d) $\kappa(G) \geq 3$. Let $e = uv$ be an arbitrary edge and $G' = G - e$. We distinguish the two cases: $\kappa_{G'}(u,v) = 2$ or $\kappa_{G'}(u,v) \geq 3$. We deal with the former case first. Let $\{a,b\}$ be a u,v-separating vertex set in G'. As above, G' decomposes into the subgraphs $G_1'(V_1,E_1)$ and $G_2'(V_2,E_2)$ with $V_1 \bigcup V_2 = V$, $V_1 \bigcap V_2 = \{a,b\}$, $u \in V_1$, $v \in V_2$, $E_1 \bigcup E_2 = E - \{e\}$, and $E_1 \bigcap E_2 = \{ab\}$, if this edge is present. (See Figure 4.4.)

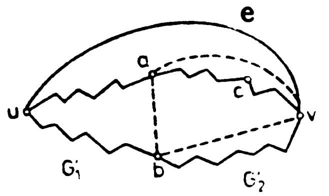


Figure 4.4

Since $p \geq 5$, there is still another vertex c in, say, G_2'. In case they are not already present, we add the edges ab, av and bv and write D for the complete graph on $\{a,b,v\}$. Finally, let G_1 and G_2 be the graphs $G_1 = G_1' \bigcup D \bigcup e$, and $G_2 = G_2' \bigcup D$. Then G_1 and G_2 have precisely D in common, and we have $u \in G_1$ and $c \in G_2$. Since $\kappa(G) \geq 3$, by Menger's theorem there are three disjoint u,c-paths. Since $\{a,b,v\}$ is a u,c-separating vertex set, these paths are decomposed (as in the proof of 4.4) into two triples P_u from u to D and P_c from D to c. Now if G_1 contained a subdivision of K_5 or $K_{3,3}$ then there would also be a subdivision in G in which possibly absent edges of D in G are replaced by paths from P_c. Our considerations for G_2 proceed completely analogously. Therefore, G_1 and G_2 are planar by the induction hypothesis. We embed G_1 in the plane. If there were a vertex in the interior of D and a vertex in the exterior of D, then by 4.4 we could join these vertices by disjoint path triples with the vertices of D. But then, together with P_c we obtain a subdivision of $K_{3,3}$ in G. Therefore D is the boundary of a country in G_1 and analogously in G_2. By our preliminary remark, we can now make D into the outer country in G_1, and thus embed G_1 (possibly by a deformation) in the interior of D in G_2 (see Figure 4.5).

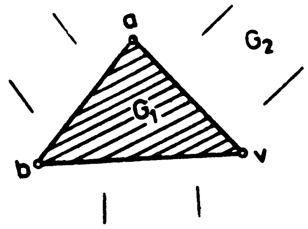


Figure 4.5

The last case, $\kappa_{G'}(u,v) \geq 3$, remains. By our induction hypothesis, we can think of G' as being embedded in the plane. By Menger's theorem, G' contains three disjoint paths W_1, W_2, and W_3 from u to v. These paths decompose the plane into three regions L_1, L_2, and L_3 as in Figure 4.6:

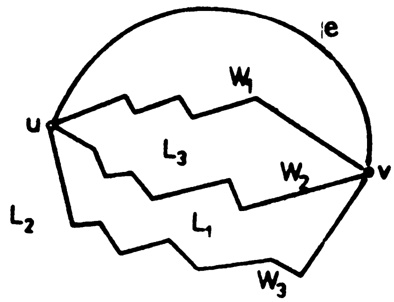


Figure 4.6

Let $F_1, \ldots, F_h$ be the countries situated in L_1 that contain u as a boundary point. If one of the F_i also contains v as a boundary point, then we could draw e as a curve from u to v inside F_i, and thus G would be planar. Otherwise, there exists, on the boundaries of F_i and inside of L_1, a path $W_{2,3}$ from a vertex $a_{2,3}$ on W_2 to a vertex $a_{3,2}$ on W_3. Similarly, there is a path $W_{1,3}$ lying entirely in L_2 from a vertex $a_{1,3}$ on W_1 to a vertex $a_{3,1}$ on W_3, and finally a path $W_{1,2}$ lying entirely in L_3 from a vertex $a_{1,2}$ on W_1 to a vertex $a_{2,1}$ on W_2. Let us consider the vertex pairs: $a_{1,2}, a_{1,3} \in W_1$, $a_{2,1}, a_{2,3} \in W_2$, $a_{3,1}, a_{3,2} \in W_3$. These pairs may be either equal or distinct—and in each case we shall verify the existence of a subdivision of K_5 or $K_{3,3}$ in contradiction to the assumption that $G \in \Pi$. The first two possibilities that all three pairs are equal (yielding a subdivision of K_5), resp., that two pairs are equal (yielding a subdivision of $K_{3,3}$), are depicted in Figure 4.7 The other two cases are disposed of as easily.

■

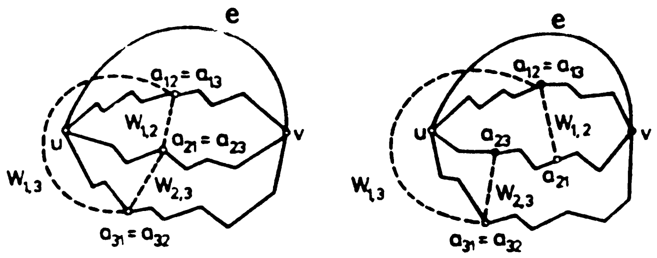


Figure 4.7

As often happens in mathematics when a theorem is "in the air", the characterization of planar graphs was also found at almost the same time in 1930 by Frink and Smith, with exactly the same conditions. Kuratowski's work, however, was the first to be published, and it is his name that has since been associated with the theorem.

In 1932 Whitney proceeded in a different direction. He took up the geometric-combinatorial ideas of Veblen, in particular, the fact that every plane graph G has a natural dual graph $G^{\star}$. Let us recall the definition: Let G be a plane graph. We obtain the **dual graph** $G^{\star}$ by placing a vertex in each country of G and joining two such new vertices if the corresponding countries are adjacent, as often as the number of common boundaries they possess. In order to make this clear, we cross the corresponding edges. (See Figure 4.8.)

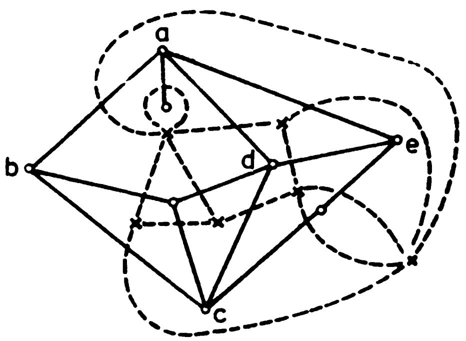


Figure 4.8

We have already noted that $G \cong G^{\star\star}$ holds if G is connected, and also only then since the dual graph is always connected. With the help of the dual, we can now answer the question posed earlier, about whether a planar graph can be embedded in the plane in various ways. Let G be planar, and suppose G_1 and G_2 are plane graphs isomorphic to G. We shall call the embeddings isomorphic if and only if isomorphism occurs also for the dual graphs $G_1^{\star}$ and $G_2^{\star}$. It is evident that a nonconnected graph can have nonisomorphic embeddings, but this is even possible for connected, and indeed for 2-connected graphs. Figure 4.9 shows two embeddings G_1 and G_2 of one and the same graph. Obviously $G_1^\star \not\cong G_2^\star$ since $G_1^\star$ has a vertex of degree 5 whereas $G_2^\star$ does not.

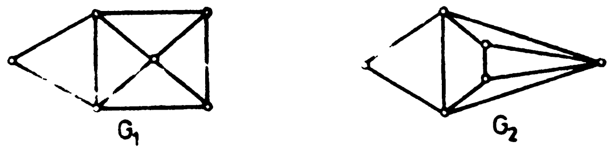


Figure 4.9

But if G is 3-connected, this can no longer happen.

4.7 THEOREM (Whitney). *A 3-connected planar graph G, with no loops, can be embedded in the plane in only one way.*

Proof. Let $\mathcal{L}_1(V, E, R_1)$ and $\mathcal{L}_2(V, E, R_2)$ be two embeddings of G with skeletons $G_1(\mathcal{L}_1) \cong G$ and $G_2(\mathcal{L}_2) \cong G$. We must show that under the identity mappings $1_V : V \to V$, and $1_E : E \to E$ of G_1 to G_2, the countries of $\mathcal{L}_1$ correspond precisely to the countries of $\mathcal{L}_2$. Let F be a country in $\mathcal{L}_1$. Since G_1 is 3-connected, the boundary edges C_1 of F form a circuit in G_1. Let C_2 be the corresponding circuit in G_2. C_2 determines a closed Jordan curve. It follows from the 3-connectedness of G_1 that $G_1 - C_1$ is connected, since by 4.5, any two vertices in G_1 are joined by at least three disjoint paths of which clearly at most two can intersect the circuit C_1. Then, since $G_1 \cong G_2$, $G_2 - C_2$ is also connected. But this excludes the possibility that vertices of G_2 occur in the interior *and* in the exterior of the closed Jordan curve C_2, i.e. C_2 must be the boundary of a country in $\mathcal{L}_2$. ■

In particular, it follows from Steinitz's Theorem 3.11 that every polytopal graph can be embedded in the plane (in the sense of the above isomorphism) in one and only one way.

Let us consider Figure 4.8. It is clear that a loop in G corresponds to a bridge in $G^\star$, and conversely. Even more: We see that the edge set of a circuit in G, e.g., of the circuit (a, b, c, d, e, a), corresponds to a minimal separating edge set in $G^\star$, and the converse is also true. Whitney used this observation in 1932 to derive a completely new combinatorial characterization of planar graphs (instead of, so to speak, the topological one in 4.6). In this connection, one was, of course, hopeful of solving the (combinatorial) coloring problem by means of a combinatorial interpretation of the underlying structure. We first give the necessary definitions.

DEFINITION. *Let $G(V, E)$ be an arbitrary graph. A* **polygon** *is the edge set of a circuit. $B \subseteq E$ is called a* **separating edge set** *if G has at least one more component upon the removal of B. Furthermore, $B \subseteq E$ is called a* **minimal separating edge set** *or a* **bond** *if B is a separating edge set but no proper subset of B has this property.*

A bond just barely holds the graph together, hence the name. If we remove a bond, then the number of components increases by exactly one. Obviously the edges incident with a vertex v always contain a bond. Another example is shown in Figure 4.10. The dark edges constitute a bond.

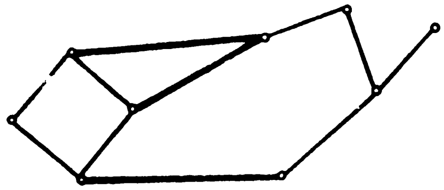


Figure 4.10

DEFINITION. *Let $G(V, E)$ be an arbitrary graph. A graph $G^\star(V^\star, E^\star)$ is called a* **dual graph to $G(V, E)$ in the sense of Whitney**, *briefly* **W-dual**, *if there is a bijection $\phi : E \to E^\star$ such that $C \subseteq E$ is a polygon in G if and only if $C^\star = \phi(C) \subseteq E^\star$ is a bond in $G^\star$.*

We have already remarked that in the case of a planar graph G the (geometric) dual graph $G^\star$ is also a W-dual. (Here, ϕ is the mapping that assigns the crossing edge $e^\star$ of $G^\star$ to each edge e of G.) Thus every planar graph has a W-dual. Now, Whitney showed that conversely, the existence of a W-dual implies the planarity of G. The proof follows from Kuratowski's Theorem and the following three lemmas.

4.8 LEMMA. *Let $G(V, E)$ be given and let $G(V, A)$ be the subgraph spanned by the edge set $A \subseteq E$. If $G(V, E)$ has a W-dual, then so has $G(V, A)$.*

Proof. Since we obtain the subgraph $G(V, A)$ step by step by omitting the edges of $E - A$, it suffices to show that the graph $G(V, E - \{e\})$ has a W-dual for each edge e. Let $G^\star(V^\star, E^\star)$ be a W-dual of G with corresponding mapping $\phi : E \to E^\star$. We set $S^\star = \phi(S)$ for each edge set $S \subseteq E$. We denote the graph that arises from $G^\star$ by **contraction** of the edge $e^\star$ (i.e., by identification of the two end vertices), by $G^\star/e^\star$. Therefore, $G^\star/e^\star$ is a graph with the edge set $E^\star - \{e^\star\}$.

CLAIM: *$G^\star/e^\star$ is a W-dual of $G(V, E - \{e\})$ via the restriction $\phi|_{E-\{e\}} : E - \{e\} \to E^\star - \{e^\star\}$.* Let $C \subseteq E - \{e\}$ be a polygon. Since $C^\star$ is a bond in $G^\star$, then $C^\star$ is naturally also a separating edge set in $G^\star/e^\star$. But $C^\star$ must also be minimal separating since $e^\star$ must lie in one of the components arising upon removal of $C^\star$ and separation is not influenced at all by contraction. The converse is shown as easily. ∎

Figure 4.11 illustrates 4.8. For simplicity we choose the same letters (without asterisks) for the corresponding edges in G and $G^\star$. We construct $G^\star$ as the geometric dual, the interior vertex of $G^\star$ corresponds to the outer country of G.

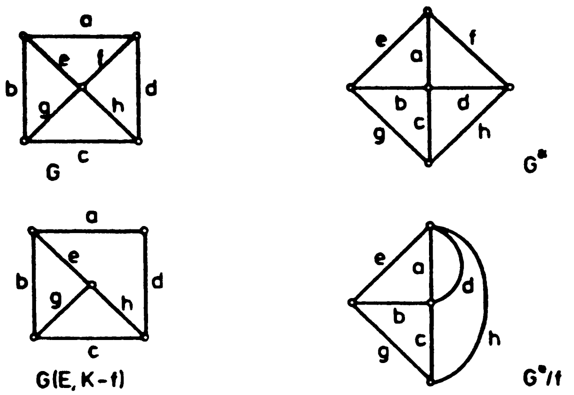


Figure 4.11

4.9 LEMMA. *If a subdivision of a graph G has a W-dual, then so does G.*

Proof. Since we can construct a subdivision step by step by the addition of the vertices of degree 2, it suffices to consider the subdivision of a single edge $e = ab$ into the two edges $e_1 = ac$, and $e_2 = cb$. Let H be the graph obtained in this way on the edge set $E_1 = E - \{e\} \bigcup \{e_1, e_2\}$ and let $H^\star$ be a W-dual with the mapping $\phi : E_1 \to E_1^\star$.

CLAIM: *$H^\star - e_2^\star$ is a W-dual of G via the mapping $\psi : E \to E_1^\star - \{e_2^\star\}$, defined by $\psi(x) = \phi(x)$ for all $x \neq e$, and $\psi(e) = e_1^\star$.* (The reader will note that in order to go from H to G we contract the edge e_2. Thus ψ is exactly the inverse mapping of ϕ from 4.8.) Let C be a polygon in G. If C does not contain the edge e, then C is also a polygon in H, and thus $C^\star$ a bond in $H^\star$. Hence $C^\star$ is also a separating edge set in $H^\star - e_2^\star$. If $C^\star$ were not minimal separating, then one of the components arising upon removal of $C^\star$ would have to contain $e_2^\star$ as a bridge (see Figure 4.12). But then $C_1^\star \bigcup \{e_2^\star\}$ would be a bond in $H^\star$, and hence $C_1 \bigcup \{e_2\}$ a polygon in H. But this can not be, since then $e_1 \in C_1$ must hold, contrary to the assumption that $e \notin C$.

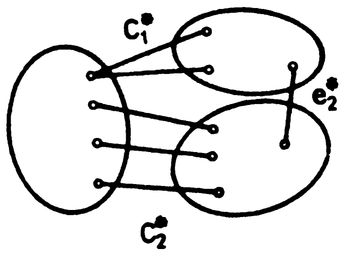


Figure 4.12

$$C^\star = C_1^\star \bigcup C_2^\star, \quad C_1^\star \neq \emptyset, \quad C_2^\star \neq \emptyset$$

Now let $e \in C$, and C be a polygon in G. Then $C_1 = (C - e) \bigcup \{e_1, e_2\}$ is a polygon in H, thus $C_1^\star$ is a bond in $H^\star$. It follows immediately that $C_1^\star - e_2^\star = \psi(C)$ is a bond in $H^\star - e_2^\star$. We leave the proof of the converse to the reader. ∎

In order to complete the proof of Whitney's Theorem, we verify that neither K_5 nor $K_{3,3}$ has a W-dual, and apply 4.6.

4.10 LEMMA. *Neither K_5 nor $K_{3,3}$ possess a W-dual.*

Proof. We content ourselves with a proof for K_5. Let $G(V,E) \cong K_5$ and $G^\star(V^\star, E^\star)$ be a W-dual. Since in the definition of a W-dual, the vertices of degree 0 are completely irrelevant, we can assume that $G^\star$ has no such vertices. Furthermore: If we join the components together at a vertex, then the resultant graph has exactly the same bonds as the original graph. Therefore, we can assume that $G^\star$ is connected. Since each edge e of G lies in a polygon, then each edge $e^\star$ lies in a bond of $G^\star$. Hence $G^\star$ has no loops since a loop can clearly not contribute to a separation. $G^\star$ also has no multiple edges. Indeed, if $a^\star$, $b^\star$ were multiple edges, then every bond $C^\star$ would contain either $a^\star$ and $b^\star$ or neither of the two edges. For G this means that for each polygon C either $\{a,b\} \subseteq C$ or $\{a,b\} \bigcap C = \emptyset$. But this can not be, since in $G \cong K_5$, for each pair of edges a, b, there is a polygon that contains a but not b. Therefore $G^\star$ is a simple connected graph. In G, we have $|C| \geq 3$ for all polygons C; hence the same for all bonds $C^\star$ in $G^\star$. Since the edges of each neighborhood $N(v^\star)$ in $G^\star$ contain a bond, it follows that $d(v^\star) \geq 3$ for all $v^\star \in V^\star$ and thus

$$20 = \sum_{v^\star \in V^\star} d(v^\star) \geq 3p^\star.$$

Therefore, for the number, $p^\star$, of vertices of $G^\star$ we have the inequality $p^\star \leq 6$. Let $C = \{a,b,c\}$ be a polygon in G corresponding to the bond $C^\star = \{a^\star, b^\star, c^\star\}$ in $G^\star$. Let $V_1^\star$ and $V_2^\star$ be the vertex sets of the components arising upon removal of $C^\star$, where we can assume that $|V_1^\star| \leq |V_2^\star|$, and let $G_1^\star$, $G_2^\star$ be the subgraphs on these vertices. (See Figure 4.13.)

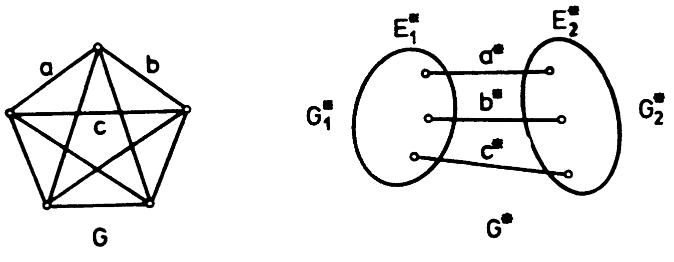


Figure 4.13

Since $|V^\star| \leq 6$, we have that $|V_1^\star| \leq 3$. If $|V_1^\star| = 3$ (and thus $|V_2^\star| = 3$), then since $d(v^\star) \geq 3$, both subgraphs $G_1^\star$ and $G_2^\star$ would have to be isomorphic to K_3 ($G^\star$ is simple !), and we obtain the contradiction $|E^\star| = 9 < 10$. If $|V^\star| = 2$, then the condition $d(v^\star) \geq 3$ would be violated for at least one of the two vertices in $V_1^\star$. Therefore we conclude that $V_1^\star = \{v^\star\}$ and $C^\star = N(v^\star)$ for a vertex $v^\star \in V^\star$. These considerations hold for every 3-polygon C in G. Since there are 10 such polygons in $G \cong K_5$, but only at most 6 vertices in $G^\star$, we again have a contradiction. ∎

We can thus finally formulate Whitney's Theorem:

4.11 THEOREM (Whitney). *A graph G is planar if and only if G has a W-dual.*

Whitney proved 4.11 without using Kuratowski's theorem, and later deduced 4.6 from his characterization. The proof of 4.10 which was not entirely easy even for K_5, indicates that the nonexistence of a W-dual for large graphs can be verified only with difficulty. Despite this, the idea of W-duality is of fundamental significance. It forms the basis for the concept of a matroid, which encompasses an abundance of combinatorial structures.

We can draw another interesting consequence from 4.11. We say that H is a **contraction** of G if H arises from G by means of a sequence of edge contractions as in Lemmas 4.8 and 4.9. We can prove, exactly as in 4.9, that if G has a W-dual, then every contraction of G also does. The next corollary follows directly from this observation and our previous results:

4.12 COROLLARY (Wagner). *A graph G is planar if and only if G does not have any subgraph that can be contracted to K_5 or $K_{3,3}$.*

In the Petersen graph P (see Figure 4.1), if we contract the 5 edges that lead from the exterior circuit to the interior circuit, then we obtain the complete graph K_5 so that using 4.12 we have again verified the nonplanarity of P.

A third way of characterizing planarity of graphs was provided by Mac Lane in a paper in 1937. His point of departure was the observation that in a planar 2-connected graph there is a family Z_0 of polygons, namely the boundary edges of the individual countries, which generate all Euler graphs Z in the algebraic sense of Veblen (Chapter 3). They have the additional property that each edge lies in exactly two polygons of Z_0. To prove that this, in fact, characterizes planar graphs, we first recall the necessary concepts from Chapter 3, and immediately generalize them to arbitrary graphs.

Let $G(V,E)$ be a graph without loops. We index the vertices $V = \{v_1, \dots, v_p\}$, the edges $E = \{e_1, \dots, e_q\}$ and keep this indexing fixed. Furthermore, we consider all circuits $Z = \{Z_1, \dots, Z_r\}$ and again keep the indexing fixed. We now associate two matrices with G, the vertex-edge-incidence matrix A and the circuit-edge-incidence matrix B. $A = \left(a_{ij}\right)$ and $B = \left(b_{ij}\right)$ are 0,1-matrices with

$$a_{ij} = \begin{cases} 1, & \text{if } v_i \in e_j, \\ 0, & \text{if } v_i \notin e_j; \end{cases} \qquad b_{lj} = \begin{cases} 1, & \text{if } e_j \in Z_l, \\ 0, & \text{if } e_j \notin Z_l; \end{cases}$$

for $i = 1, \dots, p$; $j = 1, \dots, q$; $l = 1, \dots, r$.

As in Chapter 3, we identify the 0,1-vectors of length q with the corresponding edge set and think of A and B as matrices over $GF(2)$. We again denote the subspaces spanned by the rows of A resp., B by $\hat{A}$ and $\hat{B}$. Let us describe the vectors (resp., edge sets) of these two subspaces. For $\hat{B}$ we have already done this in Chapter 3. The vectors of $\hat{B}$ correspond precisely to the **Euler subgraphs** (cycles). Furthermore, we have already seen there that $\hat{A}^\perp = \hat{B}$ and hence because of the finite rank, $\hat{B}^\perp = \hat{A}$ also holds. Now, how can we graph-theoretically characterize the vectors of $\hat{A}$? It follows from $\hat{A} = \hat{B}^\perp$ that an edge set D lies in $\hat{A}$ if and only if D intersects every polygon in an even number of edges.

DEFINITION. *An edge set D of a graph $G(V,E)$ is called a* **bipartition** *if $D = \emptyset$ or if there is a decomposition $V = V_1' \dot{\bigcup} V_2$ such that $D = \{e \in E : e$ has one end vertex in V_1, and the other in $V_2\}$.*

As an example, consider the graph in Figure 4.14. The bold edges yield a bipartition.

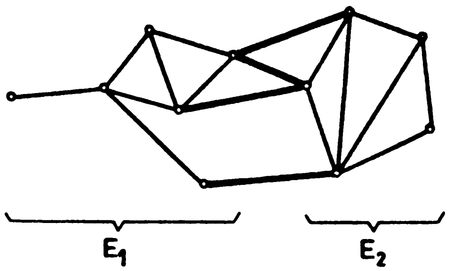


Figure 4.14

CLAIM: *The bipartitions are exactly the edge sets of $\hat{A}$.* Indeed if D is a bipartition, then $|D \cap C|$ must be even for each polygon C since C must "run to and fro" between V_1 and V_2. Conversely, let $\emptyset \neq D \in \hat{A}$. We define the relation $\sim$ on V by

$$u \sim v \Leftrightarrow \text{there is a } u,v\text{-path W that intersects } D \text{ in an even number of edges.}$$

The relation $\sim$ is an equivalence relation on V (reflexivity and symmetry are clear, transitivity may be verified by the reader) which decomposes each component G_i of G into at most two equivalence classes $V_{i,1}$ and $V_{i,2}$. It remains to show that D is precisely the bipartition induced by $V_1 = \bigcup_i V_{i,1}$, $V_2 = \bigcup_i V_{i,2}$. Each edge $e = uv$ with $u \in V_1$, $v \in V_2$ must lie in D since otherwise e would be a u,v-path that does not intersect D at all (i.e., in the even number of times 0). Conversely, if the edge $e = uv \in D$ had both end vertices in, say, V_1, then there would be a u,v-path W with $|W \bigcap D|$ even and thus the polygon $C = W \bigcup \{e\}$ would satisfy $|C \bigcap D|$ odd, in contradiction to $D \in \hat{A} = \hat{B}^{\perp}$.

Therefore $\hat{A}$ consists of exactly the bipartitions which in usual homology terminology are also called **cocycles**. Thus $\hat{A}$ is also called a **cocycle space** and $\hat{B}$ a **cycle space**. Obviously, each nonempty bipartition is a separating edge set and, conversely, each bond is a bipartition, and hence in $\hat{A}$. With these preliminaries, we can easily prove Mac Lane's Theorem.

4.13 THEOREM (Mac Lane). *A graph G is planar if and only if there is a set Z_0 of polygons such that*

(i) *each edge lies in at most two polygons of Z_0;*

(ii) *Z_0 generates the cycle space $\hat{B}$.*

Proof. Since loops have no influence whatsoever, we may omit them. If G is planar, we take the boundary edges of the countries of any planar realization, and the conditions are satisfied. Conversely, let (i) and (ii) be satisfied. We can assume that G is 2-connected since it was already noted in the proof of Kuratowski's theorem that we can combine the planar 2-connected pieces of G. Let $Z_0 = \{Z_1, \ldots, Z_m\}$. By 3.10, each edge e lies in a polygon. Thus by condition (ii), e must lie in at least one Z_i. If we consider the sum $Z' = \sum_{i=1}^{m} Z_i$ over $GF(2)$, then the Euler subgraph Z' consists of the edges that lie in exactly one polygon of Z_0. Z' decomposes into disjoint polygons. If we add these polygons to our set, we can then assume the following:

(i′) Each edge lies in exactly two polygons of Z_0.

We define a graph $G^{\star}(V^{\star}, E^{\star})$ as follows: The vertices $V^{\star} = \{v_1^{\star}, v_2^{\star}, \ldots, v_m^{\star}\}$ correspond to the polygons $Z_1, \ldots, Z_m$, the edges $E^{\star}$ to the edges E and the edge $e^{\star}$ has end points $v_i^{\star}$, $v_j^{\star}$ if and only if e lies on the two polygons Z_i, Z_j. Because of (i′), $G^{\star}$ is a graph without loops. Taking 4.11 into account, we need only show that $G^{\star}$ is a W-dual of G via the mapping $\varphi : e \to e^{\star}$.

Let A and B be the previously introduced incidence matrices for G and $A^\star$ and $B^\star$ the corresponding matrices for $G^\star$. For the sake of brevity, the i-th row of $A^\star$, corresponding to the edges incident with $v_i^\star$ shall be denoted by $V_i^\star$. Let $C^\star$ be a bond of $G^\star$, then $C^\star \in \hat{A}^\star$. Hence $C^\star = \sum_{i\in I} V_i^\star$ for a certain index set $I \subseteq \{1, \dots, m\}$. Since by the construction of $G^\star$ the matrix $A^\star$ is a submatrix of B, we obtain that $C = \sum_{i\in I} Z_i$. Therefore C is an Euler graph. Conversely, let C be a polygon in G. Then we have $C = \sum_{j\in J} Z_j$, and hence $C^\star = \sum_{j\in J} V_j^\star \in \hat{A}^\star$. That is, $C^\star$ is a bipartition. Since the bonds are precisely the minimal nonempty bipartitions and the polygons are precisely the minimal nonempty Euler subgraphs, it immediately follows: $C^\star$ is a bond in $G^\star \Leftrightarrow C$ is a polygon in G, and thus the theorem.
■

Just as the Heawood color theorem stimulated interest in the coloring problem on surfaces of higher genus, the diverse characterizations of planar graphs broaden our view with regard to other surfaces. A few years ago such a "Kuratowski" characterization was found for the projective plane by Bodendieck, Schumacher and Wagner, suggesting that maybe for all surfaces there are only a finite number of "forbidden" subgraphs (as K_5 and $K_{3,3}$ are for the plane). Very recently, this was indeed established in full generality by Robertson and Seymour. Their result constitutes without doubt one of the great advances in graph theory in recent years. We turn briefly to two natural problems that arise from the impossibility of a planar realization.

If a graph G, as say K_5, can not be embedded in the plane, then we can ask ourselves two questions: To what extent must we raise the genus h resp. k in order that G can be embedded in S_h, resp. N_k? That is we ask for the **orientable genus** $\gamma(G)$ resp. the **nonorientable genus** $\tilde{\gamma}(G)$. If we can not embed G in the plane without crossings, then how close can we come to an embedding? Here we ask for the **crossing number** $\mathrm{cr}(G)$, the minimal number of crossings that occur in a plane embedding.

We always consider connected graphs. Furthermore, we shall assume that G has neither loops nor multiple edges, without further mention. For our quantities $\gamma(G)$ and $\tilde{\gamma}(G)$ we naturally draw upon the Euler-Poincaré formula, in order to obtain a lower bound, as we have already done in Chapter 2 for the complete graph K_p. In this connection, let us make another remark: We proved Formula 2.2 for **maps** $\mathcal{L}$. In a map, each edge is homeomorphic to $\bar{\mathbb{R}}$ and each country (this is crucial) is homeomorphic to $\mathbb{R}^2$. One also calls such mappings **2-cell embeddings**. If a connected graph is embedded in another way then the formula no longer applies: Figure 4.15 shows an embedding of K_4 on the torus S_1 with 3 countries so that $|V| - |E| + |R| = 4 - 6 + 3 = 1 \neq 0 = e(S_1)$. The outer country is not homeomorphic to $\mathbb{R}^2$.

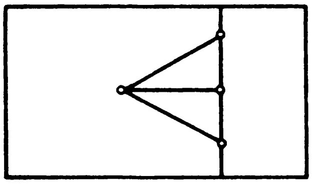


Figure 4.15

If the embedding of G in S is not a map, one can establish without too much difficulty that the embedding can be modified so that an embedding of G in S' is obtained with $e(S') > e(S)$. We can continue this until a map in fact arises, for $e(S)$ is bounded above by 2, and every embedding on the sphere is a map. From this follows immediately:

4.14 LEMMA. *Let $G(V, E)$ be a connected simple graph with p vertices and q edges. Let g be the length of the shortest circuit (in case a circuit exists). If G is embedded in the surface S, then*

$$q \leq \frac{g}{g-2}\Big(p - e(S)\Big)$$

holds. If G contains no circuits, and hence is a tree, we have the trivial inequality $q \leq p - e(S) + 1$.

Proof. By our preliminary remark, we first find a surface S' with $e(S') \geq e(S)$ on which G is embedded as a map with r' countries. Since each country has at least g boundary edges, we deduce that $gr' \leq 2q$ and by means of the Euler-Poincaré formula

$$e(S) \leq e(S') = p - q + r' \leq p - q + \frac{2q}{g}.$$

After rearrangement, the result follows. ∎

As a consequence we obtain formulas that we have already used in the plane. In this connection we shall always view the sphere also as a nonorientable surface of genus 0.

4.15 LEMMA. *Let G be a connected simple graph with $p \geq 3$ vertices and q edges.*

(i) *If G can be embedded in the orientable surface S_h $(h \geq 0)$ then $q \leq 3(p + 2h - 2)$. If G has no circuit of length 3, this can be sharpened to $q \leq 2(p + 2h - 2)$.*

(ii) *If G can be embedded in a nonorientable surface N_k, then $q \leq 3(p + k - 2)$. If G has no circuit of length 3, then this can be sharpened to $q \leq 2(p + k - 2)$.*

Very few general results are known about the parameters γ and $\bar{\gamma}$. Essentially, the results are restricted to three classes of graphs: The *complete graphs* K_p, the *complete bipartite graphs* $K_{m,n}$ and the *n-dimensional cube* Q_n. We have already discussed the first family of graphs in Chapter 2, during our report on the Heawood coloring problem, and the ultimate proof by Ringel and Youngs. Let us state the result once again.

4.16 THEOREM.

$$\gamma(K_p) = \left\lceil \frac{(p-3)(p-4)}{12} \right\rceil, \qquad p \geq 3,$$

$$\bar{\gamma}(K_p) = \left\lceil \frac{(p-3)(p-4)}{6} \right\rceil, \qquad p \geq 3, \quad p \neq 7,$$

$$\bar{\gamma}(K_7) = 3.$$

For the graphs $K_{m,n}$ we obtain from 4.15 and the fact that these graphs contain no circuits of length 3, that $mn \leq 2(m+n+2\gamma-2)$ and from this that

$$\gamma(K_{m,n}) \geq \left\lceil \frac{(m-2)(n-2)}{4} \right\rceil, \text{ and analogously,}$$

$$\bar{\gamma}(K_{m,n}) \geq \left\lceil \frac{(m-2)(n-2)}{2} \right\rceil.$$

Ringel showed in 1954 resp. 1965 that equality always holds in both of these inequalities.

4.17 THEOREM.

$$\gamma(K_{m,n}) = \left\lceil \frac{(m-2)(n-2)}{4} \right\rceil,$$

$$\bar{\gamma}(K_{m,n}) = \left\lceil \frac{(m-2)(n-2)}{2} \right\rceil, \qquad \text{for all } m,n \geq 2.$$

What are the n-dimensional cubes Q_n? Q_2 is the circuit of length 4, Q_3 is the skeleton of the 3-dimensional cube (see Figure 4.16).

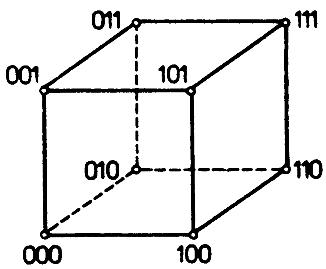


Figure 4.16

As already indicated in Figure 4.16, we take as the vertex set of Q_n all 2^n vectors of length n with entries of 0 and 1, and we join two vertices by an edge if and only if the corresponding vectors differ in exactly one coordinate. We see immediately that each vertex has degree n so that Q_n has $p = 2^n$ vertices and $q = n2^{n-1}$ edges. Furthermore, one sees immediately that Q_n has no circuit of length 3, from which, by means of 4.15, the inequality $n2^{n-1} \leq 2(2^n + 2\gamma - 2)$ and by rearrangement $\gamma(Q_n) \geq (n-4)2^{n-3} + 1$ results. Analogously, one obtains $\bar{\gamma}(Q_n) \geq (n-4)2^{n-2} + 2$. That equality also holds here with few exceptions was verified by several authors (chiefly Ringel and Jungerman).

4.18 THEOREM.

$$\gamma(Q_n) = (n-4)2^{n-3} + 1 \quad \textit{for} \quad n \geq 2,$$

$$\bar{\gamma}(Q_n) = (n-4)2^{n-2} + 2 \quad \textit{for} \quad n \geq 2,\ n \neq 4, 5$$

$$\bar{\gamma}(Q_4) = 3, \qquad \bar{\gamma}(Q_5) = 11.$$

In general, Lemma 4.15 gives a lower bound for the genus of a graph. The essential difficulty therefore consists in actually finding an embedding in the possible surfaces S_h or N_k by 4.15, or to verify its impossibility. For small graphs, resp. surfaces of small genus, one can still verify this directly, as in Figure 2.10 for K_7 on the torus. For more complicated graphs a systematic method, clearly must be found. The **rotation method** whose basic idea we briefly sketch, turns out to be especially successful (mainly in proving the Heawood theorem).

As an example we take the embedding $\mathcal{L}$ of K_7 on the torus in Figure 2.10. We can assign a scheme of 7 rows corresponding to the vertices 1,2,...,7 of $\mathcal{L}$ in the following way. We choose for each vertex i a sense of traversal of the vertices adjacent to i and call this a **rotation** at i. If, for example, we always choose the counterclockwise sense (Figure 4.17), then there results the following **rotation scheme** where the rows are understood to be ordered cyclically.

The rotation scheme gives only the neighborhoods of the individual vertices, each ordered in a certain cyclic way. But how do we obtain the countries from this? Let us consider

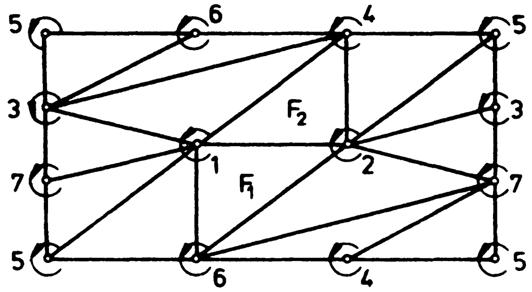


Figure 4.17

1:	2	4	3	7	5	6
2:	3	5	4	1	6	7
3:	4	6	5	2	7	1
4:	5	7	6	3	1	2
5:	6	1	7	4	2	3
6:	7	2	1	5	3	4
7:	1	3	2	6	4	5

in Figure 4.17 the country F_1 with the boundary vertices 1,2,6. If we begin with the edge (1,2) (thought of as an **edge directed from 1 to 2**) then the next edge must be $(2,i)$, where in row $2 : \ldots 1i \ldots$, i occurs after 1. If we look into our scheme, we see $2 : \ldots 16 \ldots$, therefore the next edge is (2,6). If we look at $6 : \ldots 2j \ldots$, there occurs $6 : \ldots 21 \ldots$, therefore the next edge is (6,1) and we have again arrived at 1. Conversely, if we begin with (2,1), we get (1,4) and then (4,2) and thus the country F_2.

Therefore we have constructed from the given map a rotation scheme and see that the countries can be read off from the scheme alone. The essential aspect of the idea of a rotation scheme is that the converse also works.

Let G be a connected graph that we assume to be without loops and multiple edges, with the vertex set $\{1, 2, \ldots, p\}$. For each vertex i we think of the incident edges (i, j) as being oriented from i to j, therefore each edge ij determines two oriented edges (i, j) and (j, i). We next give for each vertex i, a cyclic permutation π_i of the adjacent vertices of i, i.e., $\pi_i(j)$ is the successor of j in the row corresponding to i. We call the collection of permutations $(\pi_1, \pi_2, \ldots, \pi_p)$ a rotation ρ of G.

4.19 THEOREM. *Each rotation $\rho = (\pi_1, \ldots, \pi_p)$ of a simple graph G gives a map on a certain orientable surface S_h where S_h can be oriented so that for each vertex i the (oriented) edge (i, j) is succeeded by $(j, \pi_j(i))$. Conversely, if G is embedded as a map in S_h then there is a rotation of G which corresponds precisely to this embedding.*

Sketch of the proof. We have already indicated the converse in our example. Now suppose the rotation $\rho = (\pi_1, \ldots, \pi_p)$ is given. We denote the set of oriented edges by $\vec{E}$. We define a bijection $\Pi_\rho : \vec{E} \to \vec{E}$ by $\Pi_\rho(i, j) = (j, \pi_j(i))$. Each cycle of the permutation Π_ρ determines a country and these countries can be glued together via identification of (i, j) with (j, i) (Figure 4.18).

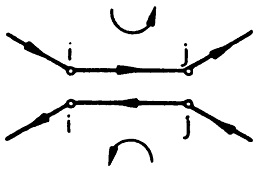


Figure 4.18

Since each permutation π_i is cyclic, all the countries are homeomorphic to $\mathbb{R}^2$. Since (i, j) is always glued to (j, i) the surface that arises is orientable. Now the genus h can be computed from the formula $|V| - |E| + r = 2 - 2h$, where r is the number of cycles in Π_ρ. ∎

In order for the genus h of an embedding of G to be as small as possible, one must therefore find rotations ρ whose corresponding permutation Π_ρ has as many cycles as possible. In the extreme case only 3-cycles occur. Such embeddings are then called **triangulations**: each country has exactly 3 boundaries. Ringel and Youngs and other authors have developed a number of interesting methods to generate suitable rotation schemes (also for the nonorientable case); the interested reader can learn more about this in Ringel's book. Ringel himself relates that in his joint work with Youngs he took over the topological part in the proof of the Heawood coloring theorem and Youngs the combinatorial part. He tried to incorporate a rotation scheme produced by Youngs. When it did not work, he returned it to Youngs who modified the scheme to take account of the new problems, whereupon it was again Ringel's turn, until finally, as we have seen, all difficulties were surmounted.

Now we turn to the problem of crossings. Here the results are still scattered and no general methods are known. Let us first make a couple of general remarks. Instead of an embedding of a graph G in the plane, we now speak of a **drawing** of G. Therefore a drawing is sought with the smallest possible number of crossing points. We can immediately exclude some types of crossing: No two edges with a common vertex cross, no two edges intersect in more than one point, no three edges intersect in a common point (see Figure 4.19). It is clear that every drawing can be modified in such a way that these three conditions are met.

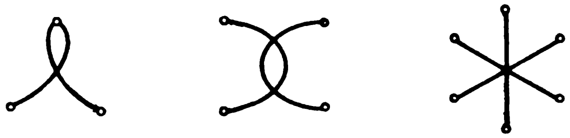


Figure 4.19

We begin with the complete bipartite graph $K_{m,n}$. Here, a drawing has been known for a long time, which has never been bested. We partition the m vertices of V_1 into two parts V_1', and V_1'' with $|V_1'| = \lfloor m/\rfloor$, and $|V_1''| = \lceil m/2 \rceil$ and similarly V_2 into V_2' and V_2'' with $|V_2'| = \lfloor n/2 \rfloor$ and $|V_2''| = \lceil n/2 \rceil$. We now draw in a Cartesian coordinate system the vertices of V_1 on the x-axis such that V_1' appears on the negative x-axis at the points $-1, -2, \ldots, -\lfloor m/2 \rfloor$, V_1'' on the positive x-axis at the points $1, 2, \ldots, \lceil m/2 \rceil$. Analogously , we draw the vertices from V_2' on the negative y-axis and V_2'' on the positive y-axis. We now join all the vertices of V_1 and V_2 by straight lines. Figure 4.20 shows the corresponding drawing for $K_{6,5}$.

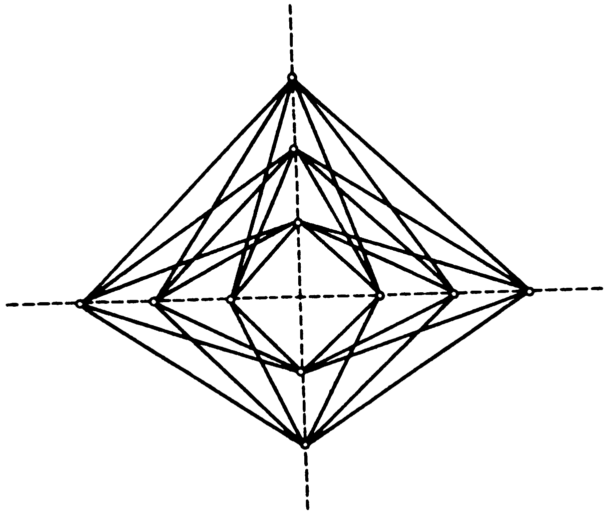


Figure 4.20

With this drawing, how many crossings occur? Consider the first quadrant. The 4 edges that join two vertices on the positive x-axis with two vertices on the positive y-axis give exactly one crossing; this clearly holds for all such vertex pairs and for all four quadrants. Therefore we obtain

$$\binom{\lfloor m/2\rfloor}{2}\binom{\lfloor n/2\rfloor}{2}+\binom{\lfloor m/2\rfloor}{2}\binom{\lceil n/2\rceil}{2}+\binom{\lceil m/2\rceil}{2}\binom{\lfloor n/2\rfloor}{2}+\binom{\lceil m/2\rceil}{2}\binom{\lceil n/2\rceil}{2}$$

crossings, and by a short calculation one sees that this number equals

$$\left\lfloor\frac{m}{2}\right\rfloor\left\lfloor\frac{(m-1)}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor\left\lfloor\frac{(n-1)}{2}\right\rfloor.$$

On several occasions proofs have appeared showing that this is the exact number, but so far all attempts have been faulty. Until now, equality has been proven for $\min(m,n)\le 6$.

4.20 THEOREM.

$$cr(K_{m,n})\le\left\lfloor\frac{m}{2}\right\rfloor\left\lfloor\frac{(m-1)}{2}\right\rfloor\left\lfloor\frac{n}{2}\right\rfloor\left\lfloor\frac{(n-1)}{2}\right\rfloor.$$

Even less is known about the complete graph K_p. Here also there is a simple construction presented in the example of K_6 in Figure 4.21. Perhaps the reader will derive some pleasure from finding the general drawing which yields the following formula with which we conclude this chapter.

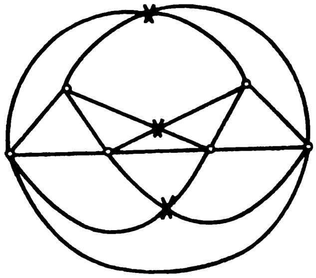


Figure 4.21

4.21 THEOREM.

$$cr(K_p) \le \frac{1}{4}\left\lfloor\frac{p}{2}\right\rfloor\left\lfloor\frac{(p-1)}{2}\right\rfloor\left\lfloor\frac{(p-2)}{2}\right\rfloor\left\lfloor\frac{(p-3)}{2}\right\rfloor.$$

Equality was proved for small p also in 4.21. Therefore, while for the classes $K_{m,n}$ and K_p there is something to be said in favor of the upper bounds in 4.20, resp., 4.21, being the right values, no drawing is known for the n-dimensional cube graph Q_n, which appears promising.

EXERCISES FOR CHAPTER 4

1. The edge connectivity number $\lambda(G)$ is the minimal number of edges whose removal separates the graph G with the additional definition $\lambda(K_1) = 0$. Show: $\kappa(G) \le \lambda(G) \le \delta(G)$, $\delta(G) = \min_{u \in V} d(u)$, and construct examples in which equality holds.

2°. Formulate and prove the edge analogue of Menger's Theorem 4.4.

3. Complete the proof of 4.5.

4. Verify 4.4, 4.5 and the edge version from exercise 2 for the Petersen graph and the Platonic graphs.

5. Suppose G has at least $2n$ vertices. Show: G is n-connected $\Leftrightarrow$ each two disjoint vertex sets V_1 and V_2 with $|V_1| = |V_2| = n$ are joined by n vertex-disjoint paths (note that here also the initial and terminal vertices are different). (Hint: Reduce to 4.4.)

6°. Let G be planar. Show: G is bipartite $\Leftrightarrow$ G^* is an Euler graph $\Leftrightarrow$ all circuits in G have even length.

7. Complete the proofs of 4.8 and 4.9.

8°. Show that $K_{3,3}$ has no W-dual.

9. Suppose $G(V,E)$ is given and that $E = E_R \bigcup E_B \bigcup \{e\}$ is a subdivision of the edge set into red edges E_R, blue edges E_B and one additional edge e. Show that exactly one of the two alternatives holds: a. There is a polygon C with $e \in C \subseteq E_R \bigcup \{e\}$. b. There is a bond D with $e \in D \subseteq E_B \bigcup \{e\}$.

10. Prove 4.12.

11. Construct a cycle basis and a cocycle basis: a. For the wheels W_n. b. For the Petersen graph.

12°. Try to find some "forbidden" Kuratowski graphs for the torus or for the projective plane.

13. Determine the Platonic graphs of the projective plane (i.e. all vertices have the same degree, all countries have the same number of boundaries).

14. Show the embeddability of $K_{4,4}$ in the torus by means of a suitable rotation scheme.

15°. Show equality in 4.20 for $K_{4,5}$.

16. Verify 4.21 by means of a suitable drawing.

17. Let G be a simple graph with p vertices and q edges. The **thickness** $\mathrm{t}(G)$ of G is the smallest number of planar subgraphs into which G can be decomposed. Therefore $\mathrm{t}(G) = 1$ if and only if G is planar. Show:

$$t(G) \geq \left\lceil q\frac{g-2}{g(p-2)} \right\rceil$$

where g is the length of the shortest circuit.

18°. Deduce from the preceding exercise that

$$t(K_p) \geq \left\lfloor \frac{p+7}{6} \right\rfloor$$

and show equality for $p \leq 8$. Does equality also hold for $p = 9$?

19. Find a lower bound for the thickness of the cube Q_n. and determine $\mathrm{t}(Q_4)$.

5. COLORING

After planarity, coloring is the second concept that occurs in the 4-color conjecture. As we have done in earlier chapters, we now discuss colorings of arbitrary graphs—naturally with the hope that in the process we also obtain information concerning the coloring of planar graphs.

We repeat the definition: A **coloring** of a graph $G(V, E)$ is a mapping $f : V \to M$ (M being the color set) with $f(u) \neq f(v)$ for $uv \in E$. If there is a coloring with n colors, then G is said to be **n-colorable** and the smallest number n for which G is n-colorable is called the **chromatic number**, $\chi(G)$. It follows from the definition that a graph with loops does not have any colorings. Hence in this chapter we shall assume that *all* graphs are without loops, without further mention. Moreover, it is not relevant for a coloring whether two vertices are joined by one, two, or arbitrarily many edges. This means that a graph G without loops and its underlying simple graph have the same chromatic number. Thus whenever it is convenient, we shall restrict ourselves to simple graphs. Along with (vertex-) colorings we have also considered edge-colorings. An **edge-coloring** of $G(V, E)$ is a mapping $\varphi : E \to M$ such that if e is incident with f then $\varphi(e) \neq \varphi(f)$. Similar to the above, we have the concept of **n-edge-colorable**. The smallest number n for which G is n-edge-colorable is called the **chromatic index**, $\chi'(G)$.

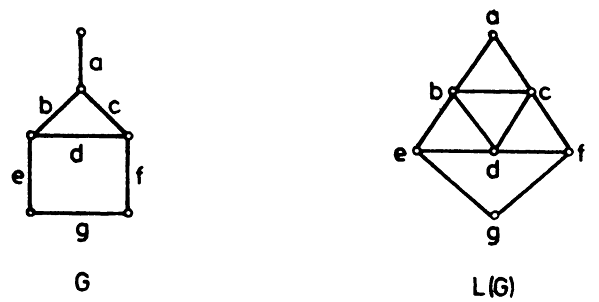


Figure 5.1

We can also consider edge-colorings as ordinary (vertex-) colorings if we associate with each graph $G(V, E)$ its **line graph** $L(G)$. The vertices of $L(G)$ correspond to the edges of G, and we join two vertices e, f in $L(G)$ if and only if e and f are incident edges in G. Clearly $\chi'(G) = \chi(L(G))$. Figure 5.1 displays an example of a graph G and its associated line graph.

We recall once more the three essential coloring theorems from Chapters 2 and 3.

a. The statement that every 3-regular bridgeless planar graph is 3-edge-colorable is equivalent to the 4-color conjecture, (1.7).
b. Every planar graph is 5-colorable, (2.1).
c. The number $p(\lambda)$ of λ-colorings is a polynomial in λ, (3.4).

This immediately suggests some questions: How is the vertex degree (or other graph parameters) related to the chromatic number, resp. the chromatic index? Can we describe the structure of minimal 5-chromatic, or in general minimal n-chromatic graphs? What else can we say about the chromatic polynomial?

We begin with several examples: The complete graph K_p obviously has $\chi(K_p) = p$. The complete bipartite graphs $K_{m,n}$ have $\chi(K_{m,n}) = 2$. What is the situation for circuits C_n of length n? For n even, we can color the vertices alternately, so $\chi(C_n) = 2$. However, if $n \geq 3$ is odd, then this (basically unique) alternating coloring assigns the same color to the first and last vertices, so we conclude that $\chi(C_n) = 3$. In summary:

5.1 THEOREM.
(i) $\chi(K_p) = p$.
(ii) $\chi(K_{m,n}) = 2$.
(iii) $\chi(C_n) = 2$ *when* $n \geq 2$ *is even, and* $\chi(C_n) = 3$ *when* $n \geq 3$ *is odd.*

The most ambitious program would, of course, be a characterization of n-colorable graphs for arbitrary n, since, by 1.7, the 4-color problem would be included for $n = 3$. Thus starting with $n = 3$ such a characterization is not to be expected. Now, what is the situation for $n = 1$ and 2? 1-colorable graphs are trivial to describe. They are exactly the ones without edges. The 2-colorable graphs are just those for which the vertex set V can be separated into two disjoint subsets V_1 and V_2, such that all edges run between the vertices of V_1 and V_2. That is, the 2-colorable graphs are exactly the subgraphs of the complete bipartite graphs. For that reason we shall usually call them **bipartite graphs**. The characterization of this class of graphs was the first general color theorem, contained in König's graph theory book of 1936.

5.2 THEOREM. *A graph is 2-colorable (bipartite) if and only if G contains no circuits of odd length.*

Proof. The condition is certainly necessary, since, as we have seen, circuits of odd length require 3 colors. Conversely, we now assume that the condition is satisfied. Since the connected components are colored independently from one another, we may assume that G is connected.

Let $u_0 \in V$ be an arbitrary vertex. We denote the length of a shortest path from u_0 to v by $d(u_0, v)$, and call it the **distance** between u_0 and v. Now we color u_0 red and extend the coloring to all of V by coloring v red when $d(u_0, v)$ is even, and blue when $d(u_0, v)$ is odd. It remains to show that adjacent vertices v and w always get different colors. Without loss of generality, let $d(u_0, v) = d \leq d(u_0, w)$. Since a shortest path W from u_0 to v together with the edge vw yields a u_0,w-path of length $d+1$, we have $d(u_0, w) \leq d+1$. Thus $d(u_0, w) = d$ or $d+1$. If $d(u_0, w) = d+1$, then v and w received different colors, which is what we wanted to prove. Hence we must exclude the case $d(u_0, w) = d$. Let W' be a shortest u_0,w-path and x the last common vertex of W and W' (see Figure 5.2). Then the subpaths $W(x, v)$ and $W'(x, w)$ have the same length $l \geq 1$, so that together with the edge vw we obtain a circuit of length $2l + 1$, which contradicts our assumption. ∎

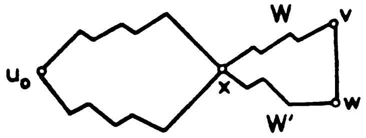


Figure 5.2

Since trees contain no circuits at all, we note the following:

5.3 COROLLARY. *Every tree is 2-colorable.*

With that, the characterizations are at an end, and we shall now turn, more modestly, to our first question, the representation of the chromatic number by other parameters.

Let us first make a couple of general remarks. If $G_1, \ldots, G_t$ are the components of G, then clearly $\chi(G) = \max \chi(G_i)$. Therefore we restrict ourselves mostly to connected graphs. If v is a cut vertex of G, then G decomposes into at least two subgraphs $G_1, \ldots, G_t$ which intersect precisely at v. If we begin the coloring at v, then again we see that $\chi(G) = \max \chi(G_i)$. The same consideration applies if G has a separating vertex set $\{w_1, \ldots, w_l\}$ which spans a complete subgraph H. Then G again decomposes into several subgraphs $G_1, \ldots, G_t$ with $\bigcap_{i=1}^{t} V(G_i) = \{w_1, \ldots, w_l\}$. By initially coloring the w_i's, one again sees that $\chi(G) = \max \chi(G_i)$. Figure 5.3 shows a graph G with a complete subgraph K_3 as a separating vertex set.

How would we go about coloring a given graph G. The most obvious method is to arbitrarily list the vertices $v_1, v_2, \ldots, v_p$, and then to color the vertices sequentially. We take $M =$

$\{1,2,3,\ldots\}$ as the color set. We begin by coloring v_1 with the color 1. If v_2 is not joined

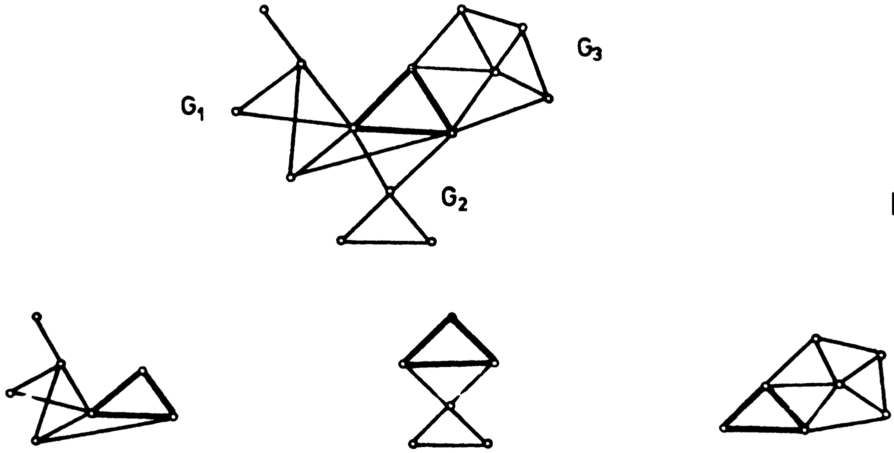


Figure 5.3

$\chi(G_1) = 3$ $\chi(G_2) = 3$ $\chi(G_3) = 4$

to v_1, then we choose 1 again for v_2, but if v_2 is joined with v_1, then we choose the next color 2. In general, we choose for v_i the smallest color that is possible, taking into account the already assigned coloring of $v_1, \ldots, v_{i-1}$. This naive method will usually use more colors than necessary. Figure 5.4 shows a graph G that uses 3 colors for the given indexing, while $\chi(G) = 2$. However, it is easy to show that it is always possible to order the vertices so that this algorithm actually does not use more than $\chi(G)$ colors. Therefore we will keep this coloring method in mind and use the symbol $\mathcal{A}$ for this "greedy" algorithm with respect to a given indexing of the vertices.

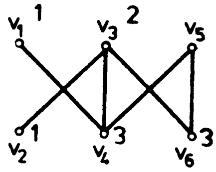


Figure 5.4

Tait's Theorem, 1.7, suggests the degree of a vertex as the first useful parameter. For this purpose we need some notation: For a graph $G(V,E)$ we denote by $\delta(G)$, the $\min_{v\in V} d(v)$, and by $\Delta(G)$, the $\max_{v\in V} d(v)$, and call them the **minimal** resp. **maximal vertex degree.** We call a subgraph $H(V',E')$ of $G(V,E)$ an **induced subgraph** when H contains all the edges joining vertices of the vertex set V' which also occur in G. Naturally, $\chi(H) \le \chi(G)$ for every subgraph H.

If we enumerate the vertices arbitrarily, $V = \{v_1, v_2, \ldots, v_p\}$, then each vertex v_i is adjacent to at most $\Delta(G)$ predecessors. Therefore the algorithm $\mathcal{A}$ at each step uses at most $\Delta(G)+1$ colors, and we conclude that $\chi(G) \leq \Delta(G)+1$. We can immediately sharpen this.

5.4 THEOREM. $\chi(G) \leq \max_H \delta(H) + 1$, *where the maximum is taken over all induced subgraphs H of G.*

Proof. Let $n = \max \delta(H)$. Then there is a vertex in G having degree $\leq n$. We take such a vertex and call it v_p. By assumption there again is a vertex with degree $\leq n$ in the induced subgraph $H_{p-1} = G - v_p$. We select one and call it v_{p-1}. Now we consider the induced subgraph $H_{p-2} = G - \{v_{p-1}, v_p\}$, etc. In this way we enumerate all vertices backwards until we arrive at v_1. If we consider the list $v_1, v_2, \ldots, v_p$, then we recognize that every vertex v_i is adjacent to at most n predecessors. Hence our algorithm $\mathcal{A}$ never uses more than $n+1$ colors. Therefore $\chi(G) \leq n+1$. ∎

We have already seen that $\chi(G) \leq \Delta(G)+1$. Let us see whether this bound can be improved. According to our initial remarks, it is no restriction if we assume that G is connected. If G is not $\Delta(G)$-regular, then every subgraph contains a vertex v with $d(v) < \Delta(G)$. Hence $\chi(G) \leq \Delta(G)$ follows from 5.4. Now let G be Δ-regular. Then we immediately find examples where $\chi(G) \leq \Delta(G)+1$ can not be improved, for instance $G = K_p$, and $G = C_n$ for n odd. In the first case $\chi = p = \Delta + 1$ and in the second case $\chi = 3 = \Delta + 1$. Brooks' Color Theorem of 1941 asserts that these are the only exceptions to $\chi \leq \Delta$.

5.5 THEOREM (Brooks). *Let G be a connected graph with $p \geq 2$ vertices and maximal degree Δ. Then $\chi(G) \leq \Delta$ holds except when G is complete or a circuit of odd length ≥ 3; in these cases $\chi(G) = \Delta + 1$.*

Proof. The assertion is trivial for $p = 2$. Hence, let $p \geq 3$. By our previous remarks, we may assume that G is Δ-regular and indeed has $\Delta \geq 3$, since a 2-regular connected graph is a circuit for which the situation has already been settled. Furthermore, we may assume that G is 2-connected, since our considerations at the beginning yield $\chi(G) = \max \chi(G_i)$, where G_i ranges over all the subgraphs attached to a cut vertex v. Since the degree of v is less than Δ in each of the graphs G_i, the subgraphs G_i are not Δ-regular. Thus $\chi(G_i) \leq \Delta$ holds for all i. Finally, we can assume that G is simple since upon omitting a multiple edge there results a G' with the same chromatic number which is no longer Δ-regular.

Now let v_p be any vertex. Since $G \neq K_p$ (i.e. $3 \leq \Delta < p-1$), there are vertices v_1 and v_2 in the neighborhood of v_p with $v_1v_2 \notin E$. If G is 3-connected, then the subgraph $H = G - \{v_1, v_2\}$ is connected. If G is 2-connected, then let $\{v_p, v\}$ be a separating vertex set. As usual, the graph G decomposes into connected subgraphs $G_1, G_2, \ldots, G_t$ with $\bigcap_{i=1}^{t} V(G_i) = \{v_p, v\}$. The vertex v_p has neighbors in all G_i (otherwise v would be a cut vertex). Let v_1 be a neighbor in G_1, and v_2 a neighbor in G_2, then $v_1v_2 \notin E$ and $G - \{v_1, v_2\}$ is connected. Thus in every case we have found vertices v_p, v_1, v_2 with $v_1v_p \in E$, $v_2v_p \in E$, $v_1v_2 \notin E$ and $G - \{v_1, v_2\}$ connected. In order to apply our algorithm $\mathcal{A}$, we enumerate the vertices as follows: let $v_{p-1} \in V - \{v_1, v_2, v_p\}$ be adjacent to v_p (v_{p-1} exists, since $\Delta \geq 3$), let $v_{p-2} \in V - \{v_1, v_2, v_p, v_{p-1}\}$ be adjacent to v_p or to v_{p-1}, etc. For each i, $3 \leq i < p$, there is a vertex $v_i \in V - \{v_1, v_2, v_p, \ldots, v_{i+1}\}$ which is adjacent to at least one of the vertices $v_{i+1}, \ldots, v_p$, since otherwise $G - \{v_1, v_2\}$ would not be connected. Starting with v_1, the algorithm $\mathcal{A}$ assigns v_1 and v_2 the color 1. It never needs more than Δ colors, since each vertex v_i, $3 \leq i < p$, is adjacent to at most $\Delta - 1$ predecessors, and v_p is adjacent to v_1 and v_2. ■

Let us apply our results to planar graphs. By 2.9 we know that every simple planar graph G must have a vertex with degree at most 5. Since every subgraph H is also planar, it follows that $\delta(H) \leq 5$ and thus, by 5.4, $\chi(G) \leq 6$. Since we already know that $\chi(G) \leq 5$, this result is not remarkable. The connection with Tait's Theorem is more interesting. Let G be a 3-regular bridgeless planar graph. The line graph $L(G)$ is then obviously 4-regular and it is surely not a circuit of odd length, nor is it complete. Hence, by Brooks' Theorem, $\chi(L(G)) \leq 4$, i.e., $\chi'(G) \leq 4$. Since each triple of edges that is incident with a vertex must receive different colors, we infer that $\chi'(G) = 3$ or $=4$ and it all comes down to excluding the case $\chi'(G) = 4$.

Accordingly, it is of benefit to study **edge colorings**. Suppose the graph G (as usual, assumed to be without loops) has maximal degree Δ. All the edges incident to a vertex v with $d(v) = \Delta$ must be colored differently. Hence, $\chi'(G) \geq \Delta$. Since an edge e is adjacent to at most $2\Delta - 2$ other edges (see Figure 5.5), it follows from 5.5 applied to $L(G)$ that $\chi'(G) \leq 2\Delta - 1$. This upper bound was improved by Shannon in 1949 who verified that $\chi'(G) \leq \lfloor 3\Delta/2 \rfloor$. Finally, in a remarkable article, Vizing showed in 1964 that for a simple graph G the chromatic index can differ by at most 1 from the trivial lower bound $\Delta(G)$.

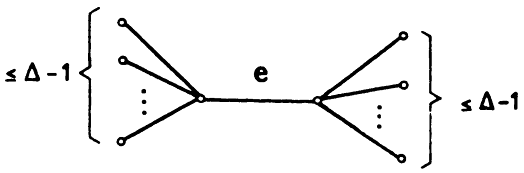


Figure 5.5

5.6 THEOREM (Vizing). *Let* $\Delta(G)$ *be the maximal degree of the simple graph* $G(V, E)$. *Then* $\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1$.

Proof. We use induction on the number of edges. The assertion is trivial for $|E| = 0$. We now assume that all edges apart from $e_1 = uv_1$ were colored properly with at most $\Delta + 1$ colors. We will show that there is also a $(\Delta + 1)$-coloring for all of E. We say that the color α is **missing at the vertex** w if none of the edges incident with w was colored by α. If α, β are two colors, then we denote by $G(\alpha, \beta)$, the subgraph spanned by the α- and β-colored edges. It is clear that each vertex in $G(\alpha, \beta)$ has degree at most 2. Therefore the components are either paths or circuits (of even length). A color must be missing at u as well as at v_1 (because $d(v)$, $d(v_1) \leq \Delta$). If the same color is missing at both vertices, we can use this color for the edge e_1 and are finished. Therefore let α be a missing color at u and $\beta \neq \alpha$ a missing color at v_1. The proof now proceeds in 3 steps.

(i) Let $e_2 = uv_2$ be the edge incident with u which was colored with β_1. We remove the coloring of e_2 and instead color e_1 with β_1. If α is missing at v_2 then we can take α for e_2 and are finished. Hence we assume that α appears at v_2. If the vertices u, v_1, v_2 are not in the same component of $G(\alpha, \beta_1)$, then we can interchange the colors α and β_1 in the $G(\alpha, \beta_1)$-component containing v_2 without changing the β_1-coloring of e_1 whereby α would be available for the edge e_2. We thus have the situation of Figure 5.6.

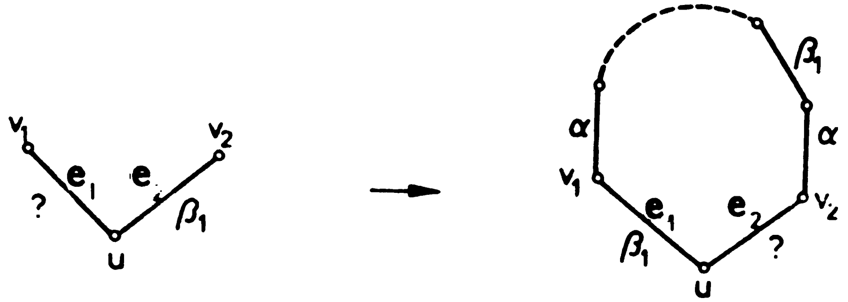


Figure 5.6

(ii) Now let $\beta_2 \neq \beta_1$ be a color missing at v_2. We can again assume that β_2 appears at u, since otherwise we could use β_2 for coloring e_2. Therefore let $e_3 = uv_3$ be colored with

β_2. We strip the coloring from e_3 and instead color e_2 with β_2. The same line of reasoning as in (i) shows that we can assume that α appears at v_3 and that u, v_2, v_3 lie in the same component of $G(\alpha, \beta_2)$ (see Figure 5.7).

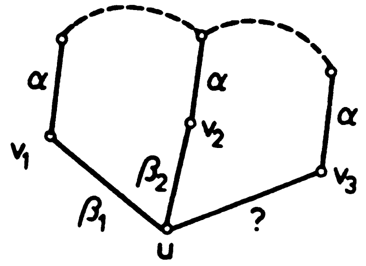


Figure 5.7

(iii) If we continue this recoloring procedure then sooner or later we come to a vertex v_j which is adjacent to u, for which the edge $e_j = uv_j$ is not colored, and where a color $\neq \beta_{j-1}$ missing at v_j is also missing at u (we then color e_j with it and are finished) or a color β_i is missing at v_j with $i < j - 1$. As in step (i), u, v_i, v_{i+1} lie in the same $G(\alpha, \beta_i)$-component, H. Since α is missing at u and β_i at v_{i+1}, this component must be a u, v_{i+1}-path whose edges are alternately colored β_i and α (Figure 5.8). This path does not contain the vertex v_j since β_i is missing at v_j. From this it follows that the $G(\alpha, \beta_i)$-component H' containing v_j is disjoint from H. We can thus interchange the colors in H' and, finally, apply the free color α to color $e_j = uv_j$. ∎

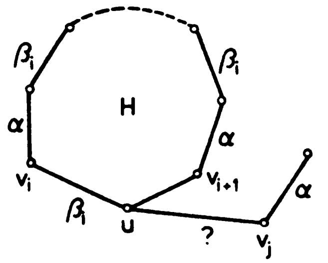


Figure 5.8

Note that the proof yields an algorithm for an edge-coloring with $\Delta + 1$ colors. The following classification is suggested by Vizing's Theorem. A graph G lies in the class Γ_0 if $\chi'(G) = \Delta$ and in the class Γ_1 if $\chi'(G) = \Delta + 1$. If we had a decision method to decide whether a given graph is in Γ_0 or Γ_1, then in light of Tait's Theorem, we would have a solution of the 4-color problem also.

It is just because of this theorem that regular graphs are of primary interest. Assume that an r-regular graph G is in Γ_0. The individual color classes of an r-edge coloring consist solely of nonincident edges and can thus contain at most $\lfloor p/2 \rfloor$ edges. From this we obtain

$r\lfloor p/2 \rfloor \geq q = rp/2$ (by 1.2). Therefore if p is odd, then G can not lie in Γ_0. But if p is even, then it follows from $G \in \Gamma_0$ that the individual color classes are 1-factors of G and thus the mentioned edge coloring is a 1-factorization of G in the sense of Chapter 3. Therefore we see the close connection with the concept of factorization that we will go into more deeply in the next chapter.

5.7 THEOREM. *Let G be a regular graph. If the number of vertices p is odd, then $G \in \Gamma_1$; if p is even then G lies in Γ_0 if and only if G is 1-factorable.*

Let us look once again at the planar graphs. If $\Delta(G) \leq 5$ then G can lie in Γ_0 as well as in Γ_1. Examples for Γ_0 are the even circuits ($\Delta = 2$), the tetrahedron graph ($\Delta = 3$), the octahedron graph ($\Delta = 4$) and the icosahedron graph ($\Delta = 5$). The last three are depicted in Figure 5.9. The reader can furnish a corresponding edge coloring by himself.

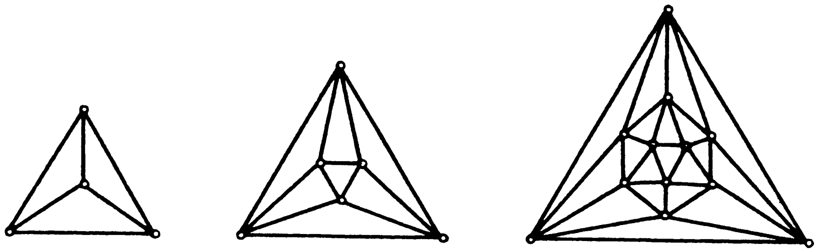


Figure 5.9

The odd circuits are in Γ_1 ($\Delta = 2$) and if we add a vertex in any edge of the graphs depicted in Figure 5.9 then we obtain graphs in Γ_1. We verify this for the tetrahedron. If the new graph G' is in Γ_0 then the edges can be decomposed into 3 color classes. Since G' has 5 vertices, each such class contains at most $\lfloor 5/2 \rfloor = 2$ edges, so that we could color at most 6 edges altogether. But G' contains 7 edges. The proof in the two other cases proceeds in exactly the same way.

The result of Vizing (1965) that every simple planar graph G with $\Delta(G) \geq 8$ is in Γ_0 is thus quite surprising. Vizing, in fact, conjectured that $\Delta \geq 6$ already implies $G \in \Gamma_0$.

We return to the chromatic number $\chi(G)$. After we have seen that Δ provides an interesting **upper** bound, we ask which parameters occur as **lower** bounds. A first trivial, but useful, bound results immediately from the following observation. If we color G with n colors, then vertices of the same color have the property that there are no edges between them. In general, we call a subset $A \subseteq V$ **independent** if no two vertices of A are joined by an edge. If we define the **independence number** $\alpha(G)$ to be the size of a largest independent set, then the next lemma follows immediately:

5.8 LEMMA. *For a graph G with p vertices, $\chi(G) \geq p/\alpha(G)$.*

For circuits C_p of odd length, it again follows that $\chi(C_p) \geq 3$, since $\alpha(C_p) = (p-1)/2$. On the other hand, it is easy to give examples for which 5.8 yields little.

Another fact is immediate. If G contains a complete subgraph with n vertices, then $\chi(G) \geq n$ must hold trivially. This suggests the following definition: Let the **clique number** $\omega(G)$ be the number of vertices of the largest complete subgraph of G. The following is clear:

5.9 LEMMA. $\chi(G) \geq \omega(G)$.

How good is this bound? Does a graph with a small clique number also have a small chromatic number? Unfortunately no. Blanche Descartes (the incomparable Tutte hides behind this name), Mycielski and Zykov have independently and almost simultaneously shown that there are graphs G with arbitrarily large chromatic number $\chi(G)$ that have no triangles, therefore for which $\omega(G) = 2$.

Let us look at Mycielski's construction. We begin with $G_3 = C_5$, the circuit of length 5 for which $\chi(G_3) = 3$ and $\omega(G_3) = 2$. Assume we have already constructed G_n with $\chi(G_n) = n$, and $\omega(G_n) = 2$. To each vertex v of G_n we introduce a new vertex v' and join v' with precisely the neighbors of v. No adjacencies between the new vertices v' are introduced. Finally we add another vertex x that we join with all the vertices v'. This is our graph G_{n+1}. Figure 5.10 shows the step from G_3 to G_4.

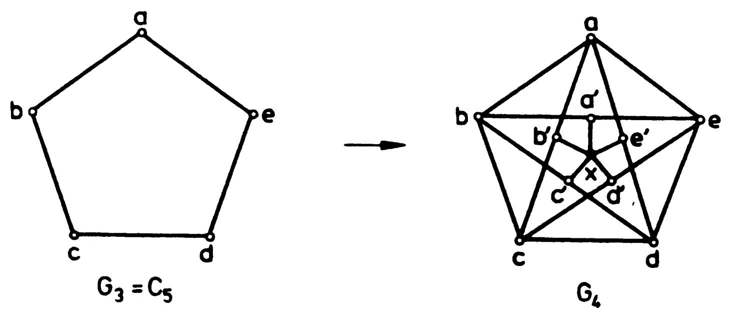


Figure 5.10

It is immediately clear that again $\omega(G_{n+1}) = 2$. If we assume that G_{n+1} is n-colorable then we consider the n-coloring induced on G_n. There must be a vertex v_i in each color class i of G_n that is adjacent to at least one vertex from every other color class. Otherwise we could distribute the vertices from the color class i to the other classes and G_n would be $(n-1)$-colorable, in contradiction to $\chi(G_n) = n$. Since v_i' has exactly the same neighbors as v_i in G_n, therefore v_i' must receive the same color as v_i. Thus all n colors appear among the vertices v' and we could no longer color x. Therefore G_{n+1} is not n-colorable. If we transfer the coloring of the vertex v from G_n to the vertex v' and color x with the $(n+1)$-st color, then we obtain an $(n+1)$-coloring of G_{n+1}, and consequently $\chi(G_{n+1}) = n + 1$.

Therefore, in general, the clique number $\omega(G)$ does not affect the number of required colors. But a very interesting class of graphs results from the equality $\chi = \omega$. We call a simple graph G **perfect** if $\chi(H) = \omega(H)$ holds for every induced subgraph H (hence also for G). The smallest non-perfect graph is C_5 since $\omega(C_5) = 2 < 3 = \chi(C_5)$ and naturally all other circuits of odd length ≥ 5 are also non-perfect. A remarkable theorem by Lovász asserts that a simple graph G is perfect if and only if this also occurs for the complement $\overline{G}$. By the **complement** $\overline{G}$ we mean the graph that arises by removing all edges from G and adding precisely the previously missing edges. Figure 5.11 shows a graph G and its complement $\overline{G}$.

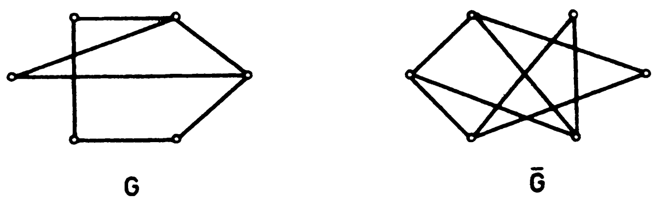


Figure 5.11

Therefore by Lovász's result, the complement of a circuit of odd length is once again non-perfect, and one of the most interesting unsolved problems of graph theory states that these graphs determine non-perfectness. Berge posed the conjecture in 1961: *A simple graph G is perfect if and only if G neither contains a circuit of odd length ≥ 5 nor does it contain the complement of such a circuit as induced subgraph.*

We turn to the second group of problems—the study of the **structure** of n-chromatic graphs. We already encountered the determination of a hypothetical minimal 5-chromatic map in the treatment of the 4-color problem. In the dual formulation this is a planar graph G with $\chi(G) = 5$ but for which $\chi(G - v) < 5$ holds for each vertex v. We will now pursue this idea.

DEFINITION. *A simple graph G is called* **critical** *if $\chi(G - v) < \chi(G)$ for every vertex v. If $\chi(G) = n$, then G is called* **n-critical**.

If G is critical, then naturally $\chi(G - v) = \chi(G) - 1$ holds for every vertex v. The importance of this concept consists in the fact that every n-chromatic graph G contains an n-critical graph, for obviously every smallest induced subgraph H of G with $\chi(H) = \chi(G)$ is n-critical. Critical graphs exhibit more structure—and it is of course our goal to understand as much as possible concerning the structure of n-chromatic graphs. Most of the results that now follow were proven by Dirac in the 1950's.

Trivially, there are no 1-critical graphs, and the only 2-critical graph is K_2. The 3-critical graphs are also easily determined. By 5.2, they are exactly the circuits of odd length. No characterization is known for n-critical graphs with $n \geq 4$.

5.10 LEMMA. *In an n-critical graph G, the minimal degree $\delta(G) \geq n-1$. If G has p vertices and q edges then $q \geq \frac{1}{2}(n-1)p$.*

Proof. We assume the contrary, that G has a vertex v with $d(v) \leq n-2$. Since G is n-critical, we can color $G-v$ with $n-1$ colors and extend this $(n-1)$-coloring to v since v has at most $n-2$ neighbors. Contradiction. The last assertion follows from 1.2. ∎

Since an n-critical graph trivially has at least n vertices, this implies:

5.11 COROLLARY. *An n-chromatic graph has at least n vertices of degree $\geq n-1$.*

A critical graph is clearly connected. But we can say more:

5.12 LEMMA. *A critical graph G can not be separated by a complete subgraph, i.e., if S is a separating vertex set, then the subgraph induced by S is never complete. In particular, an n-critical graph with $n \geq 3$ must be 2-connected.*

Proof. Let G_S be the subgraph spanned by S. Then G decomposes into subgraphs G_i with $\bigcap V(G_i) = S$ and, as we have seen earlier, $\chi(G) = \max \chi(G_i)$ when G_S is complete. Since all G_i are proper subgraphs, we then have $\chi(G_i) < \chi(G)$ for all i, and thus the contradiction $\chi(G) < \chi(G)$. ∎

Let us consider, once more, the estimate $\chi(G) \geq \omega(G)$. We have noted that this inequality can be arbitrarily bad. And yet one has the feeling that an n-chromatic graph must contain, if not a subgraph K_n, at least a subgraph that reflects in some sense the structure of the complete graph K_n. This is the content of the famous conjecture stated by Hadwiger in 1943.

HADWIGER'S CONJECTURE. *The following holds for all n: If $\chi(G) \geq n$ then G contains a subgraph that can be contracted to K_n.*

Let us abbreviate the conjecture for n by the symbol (H_n). We recall that the contraction of a graph G onto H means that H arises from G by a sequence of edge contractions (i.e., identification of the end vertices).

5.13 THEOREM. *(H_n) holds for $n \leq 4$.*

Proof. The assertion is trivial for n=1 and 2. If G has chromatic number $\chi(G) \geq 3$, then by 5.2 G must contain a circuit of odd length ≥ 3. Such a circuit can clearly be contracted to a complete graph K_3. Now let $\chi(G) \geq 4$. We can assume G to be simple. We use induction on the number of vertices p. If $p = 4$, then $G = K_4$ must hold. Suppose the assertion is correct for all graphs with fewer than p vertices. If G is not critical, we choose a critical subgraph H for which $\chi(H) = \chi(G) \geq 4$, and we are finished with the induction. Therefore assume that G is critical. By 5.12, G is 2-connected.

Let us assume that G is also 3-connected. If $G = K_p$, then we are finished. Otherwise we choose two vertices u and v that are not joined. By Menger's Theorem, 4.4, there are three disjoint u,v-paths W_1,W_2,W_3. Let w_1 and w_2 be vertices on W_1, resp. W_2 different from u,v. Again by 4.4 there is at least one path W' from w_1 to w_2 that contains neither u nor v (see Figure 5.12). But now we can immediately see that $W_1 \bigcup W_2 \bigcup W_3 \bigcup W'$ contains a subdivision H of K_4. Therefore H can be contracted to K_4. As the last possibility assume that G is 2-connected and $\{u, v\}$ is a separating vertex set. By 5.12, u and v are not adjacent. We add the edge $e = uv$ and decompose $G \bigcup e$ as usual into two subgraphs G_1 and G_2 with $V(G_1) \bigcap V(G_2) = \{u, v\}$. We already know that one of the two graphs G_1 or G_2 must have chromatic number ≥ 4. We assume that this occurs for G_1. By the induction assumption, G_1 contains a subgraph H that can be contracted to K_4. Since u and v are joined in $G_2 - k$ by a path W (G is 2-connected!), we can replace the edge e in H (in case it occurs) by W, and in any case can then contract to K_4. ■

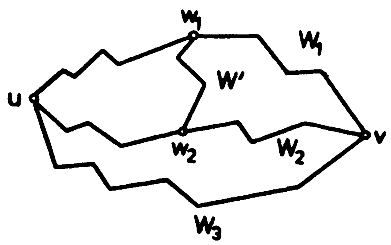


Figure 5.12

What is the situation for $n = 5$? We know from 4.12 that a planar graph contains no subgraphs that can be contracted to K_5. Accordingly, if (H_5) is correct, then the 4-color conjecture must also be correct. In 1960 Wagner showed the surprising situation that conversely the correctness of the 4-color conjecture implies the correctness of (H_5). Therefore: 4-CC $\Leftrightarrow$ (H_5). Since one can easily see that (H_{n+1}) always implies (H_n), we have before us two possibilities: Either (H_n) is correct for all n, or there is a smallest number n_0 such that (H_n) is false for all $n \geq n_0$.

In the chapter on planar graphs, besides the concept of contraction, we also became acquainted with the concept of a subdivision. In light of Kuratowski's Theorem, 4.6, the following conjecture is also plausible:

(U_n) *If* $\chi(G) \geq n$, *then* G *contains a subdivision of* K_n.

Since each subdivision H of K_n can naturally be contracted to K_n, (U_n) is stronger than (H_n) for every n. In our proof of 5.13 we, indeed, always found subdivisions, hence (U_n) is also true for $n \leq 4$. Furthermore, on the basis of Kuratowski's Theorem, the correctness of the 4-color conjecture would follow from (U_5). And just as for the Hadwiger conjecture, it is also true here that $(U_{n+1}) \Rightarrow (U_n)$. For a long time it was believed that (U_n) is in fact correct for all n, until Catlin (1979) gave the simple counterexample G in Figure 5.13. Figure 5.13 is to be understood as follows: we replace the vertices of a circuit of length 5 by a graph K_3 and join two vertices from different K_3's if and only if the original vertices of the 5-circuit were joined. Therefore G has 15 vertices and it is regular of degree 8. Since G clearly contains no independent set with more than 2 vertices, it follows from 5.8 that $\chi(G) \geq 15/2$, i.e., $\chi(G) \geq 8$. Moreover, because of Brooks' Theorem 5.5, $\chi(G) = 8$. But the reader can easily verify that G contains no subdivision of a K_8. Thus (U_8) is false and therefore (U_n) is also false for $n \geq 8$. Whether (U_n) holds for n between 5 and 7 is an open question. Observe, however, that G is not a counterexample to (H_8). The reader can take the trouble to discover a corresponding subgraph, indeed there are even subgraphs that can be contracted to K_9.

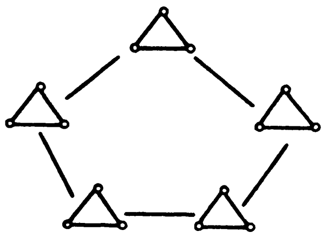


Figure 5.13

We now come to the third group of problems: How many colorings are there? In Chapter 3 we discussed the work of Birkhoff and mainly his result 3.4 that the number $p(G; \lambda)$ of the λ-colorings of G is a polynomial in λ. We have seen in 3.4 that the coefficients of the chromatic polynomial give some information about the graph itself. The ideal case would naturally be if the graph could be uniquely determined by its chromatic polynomial. But this is not even close to being true, as we shall immediately demonstrate using trees. First a useful lemma.

5.14 LEMMA. *If G is the union of two subgraphs G_1, G_2 whose intersection is the complete graph K_n, then*

$$p(G;\lambda) = \frac{p(G_1;\lambda)p(G_2;\lambda)}{p(K_n;\lambda)}.$$

Proof. Each coloring of G corresponds exactly to a pair (f_1, f_2) of colorings of G_1 resp. G_2 that coincide on K_n. Now if f_1 is a λ-coloring of G_1 then there clearly are exactly $p(G_2;\lambda)/p(K_n;\lambda)$ λ-colorings f_2 of G_2 that coincide with f_1 on K_n, and the assertion follows immediately. ∎

EXAMPLE. *Let us compute the chromatic polynomial of the graph G in Figure 5.14. The two K_3's are attached along the edge e, therefore we obtain*

$$p(G;\lambda) = \frac{(\lambda(\lambda-1)(\lambda-2))^2}{\lambda(\lambda-1)} = \lambda(\lambda-1)(\lambda-2)^2.$$

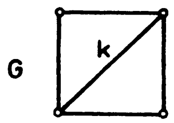


Figure 5.14

If we omit e, then there results from the usual recursion (consult the proof of 3.4) the chromatic polynomial of the circuit of length 4,

$$p(C_4;\lambda) = p(G;\lambda) + p(G/e;\lambda) = \lambda(\lambda-1)(\lambda-2)^2 + \lambda(\lambda-1)^2 = \lambda(\lambda-1)(\lambda^2-3\lambda+3),$$

where G/e is the graph that arises by contraction of e.

5.15 LEMMA. *Every tree T with p vertices has the chromatic polynomial*

$$p(T;\lambda) = \lambda(\lambda-1)^{p-1} = \sum_{i=0}^{p-1}(-1)^i\binom{p-1}{i}\lambda^{p-i}.$$

Proof. We use induction on p. For $p = 1$ there is nothing to prove. Now let $p > 1$. Since $2q = 2p - 2 = \sum_{i=1}^{p} d(v_i)$, there is a vertex u with $d(u) = 1$. For the tree T' arising upon deletion of u, $p(T';\lambda) = \lambda(\lambda-1)^{p-2}$ holds by induction. The edge uv intersects T' at v so that with 5.14 we immediately obtain

$$p(T;\lambda) = \frac{p(T';\lambda)p(K_2;\lambda)}{p(K_1;\lambda)} = \frac{\lambda(\lambda-1)^{p-2}\lambda(\lambda-1)}{\lambda} = \lambda(\lambda-1)^{p-1}. \quad ∎$$

Moreover, the converse also holds in the preceding: If $p(G;\lambda) = \lambda(\lambda-1)^{p-1}$, then G must be a tree as follows immediately from Theorem 3.4(i)-(iii).

An interesting question, open up to now, is suggested by this result: Assume the simple graphs G and H have identical chromatic polynomials, $\chi(G) = \chi(H)$. What can one say about the common structure? From 3.4 we know that G and H must have the same number of vertices, edges and components; what else holds? In other words, we must be able to interpret the coefficients a_i of $p(G;\lambda)$ graph-theoretically, i.e., be able to state what they really count in the graph. Such an interpretation was found by Whitney in 1932, whose proof uses one of the most important combinatorial counting principles.

Let us first discuss the combinatorial principle. It is the **inclusion-exclusion** method. To stick to our coloring theme, assume that 10 boxes are placed on a table. Person A comes and marks r of them red, then B comes and marks b of the 10 boxes blue, and finally C comes who marks y of the boxes yellow. Question: How many remain without any marking? In order to determine this number $E(0)$ we first subtract from 10, the number r of the boxes marked red, the number b of them marked blue, and finally the number y of them marked yellow. In general $10-r-b-y$ will not be the correct result since we have, e.g., subtracted a box marked both red and blue twice. Therefore we again add the number $rb+ry+by$ where rb is the number of boxes marked both red and blue and similarly for ry and by. Now we have already counted almost correctly, except for those boxes that are marked red, blue and yellow, which we have subtracted and added three times so that they are still not accounted for. The correct result thus is: $E(0) = 10-r-b-y+rb+ry+by-rby$.

5.16 INCLUSION-EXCLUSION PRINCIPLE. *Let S be a finite set and $P = \{P_1, \ldots, P_t\}$ a list of properties that the elements of S may or may not possess. For $A \subseteq P$ we denote by $S(A) \subseteq S$ that subset of elements that exhibit all the properties of A (and possibly also further properties). Thus, in particular, $S(\emptyset) = S$. Let $N_i = \sum_{A\subseteq P,\ |A|=i} |S(A)|$, $i = 0, \ldots, t$. Then for the number, $E(0)$, of elements of S that do not possess a single property:*

$$E(0) = N_0 - N_1 + N_2 - N_3 + \ldots + (-1)^t N_t.$$

Proof. If we replace N_i by its defining expression, then the assertion is equivalent with $E(0) = \sum_{A\subseteq P}(-1)^{|A|}|S(A)|$. Assume that the element $s \in S$ has precisely the properties $B \subseteq P$. Then, on the right side, s is counted exactly in the expressions $(-1)^{|A|}$ with $A \subseteq B$. Thus, if $|B| = k$, then the contribution of s is exactly

$$1 - \binom{k}{1} + \binom{k}{2} - \ldots + (-1)^k.$$

By the binomial theorem this is 0 when $k \geq 1$ and 1 when $k = 0$. ∎

Although this principle appears simple, it is one of the most fruitful methods in combinatorial enumeration. Let us go back to Whitney. If $G(V,E)$ is given, then we call a subgraph H a **spanning subgraph** when H is defined on the entire vertex set. We denote by $c(H)$ the number of components of H. We next consider *all* mappings $f : V \to \{1, 2, \dots, \lambda\}$. Let this be our set S, for which then $|S| = \lambda^p$. For each edge $e = uv \in E$ we define a property P_e by stating $f \in S$ has the property P_e if $f(u) = f(v)$. Therefore, by definition, our chromatic polynomial $p(G;\lambda)$ is precisely equal to the above number $E(0)$. It remains to find an expression for $|S(A)|$, $A \subseteq E$. Let $H_A = H(V, A)$ be the spanning subgraph with edge set A. Then clearly a mapping $f \in S(A)$ must assign the same value in $\{1, 2, \dots, \lambda\}$ to all vertices of a component of H_A. But the converse is also true: If a mapping f is constant on the vertices of each component (vertices in different components can obtain different values), then $f \in S(A)$. But the number of these mappings f is obviously $\lambda^{c(H_A)}$, and we obtain

$$p(G;\lambda) = \sum_{A \subseteq E} (-1)^{|A|} \lambda^{c(H_A)}.$$

Hence by combining the edge sets A with equal $c(H_A)$, we obtain the following result:

5.17 THEOREM (Whitney). *Let $p(G;\lambda) = \sum_{i=0}^{p} a_i \lambda^{p-i}$ be the chromatic polynomial of the graph G. Then $a_i = \sum_{l=0}^{q} (-1)^l N(l, p-i)$, where $N(l,j)$ is the number of spanning subgraphs with l edges and j components.*

Note that it again follows immediately from the formula that $a_0 = 1$ and $a_1 = -q$ for simple graphs. The linear coefficient a_{p-1} is also interesting when G is connected. We observe that a_{p-1} is equal to the difference between the number of connected spanning subgraphs with an even number of edges and those with an odd number of edges—a remarkable result. For our decision problem whether two graphs have equal chromatic polynomial, 5.17 is naturally applicable only to a very restricted extent. We would have to test all spanning subgraphs. Indeed, Whitney showed in the same work, that we can restrict ourselves to a subset of the subgraphs, however, his result is still far from being a workable algorithm.

Despite that, it is desirable to obtain more precise information about the coefficients of chromatic polynomials. To do this, the recursion in the proof of Theorem 3.4 is most useful. We have shown there: If e is an edge of G and if $G - e$ and G/e are the graphs that arise upon removing the edge e, resp. by contracting e, then $p(G;\lambda) = p(G - e;\lambda) - p(G/e;\lambda)$. In order to simplify the induction proof, it is advantageous to consider instead of the coefficients a_i in $p(G;\lambda) = \sum_{i=0}^{p} a_i \lambda^{p-i}$ the absolute values $\alpha_i = |a_i|$. We write $p(G;\lambda) = \sum_{i=0}^{p} (-1)^i \alpha_i \lambda^{p-i}$; then all $\alpha_i \geq 0$, since by 3.4(ii) the coefficients $a_0, a_1, a_2, \dots$ have alternating signs.

5.18 LEMMA. *Let $p(G;\lambda) = \sum_{i=0}^{p}(-1)^i\alpha_i\lambda^{p-i}$ be the chromatic polynomial of G and $\psi(G;\lambda) = \sum_{i=0}^{p}\alpha_i\lambda^{p-i}$. Then the following hold:*

(i) $\psi(G;\lambda) = (-1)^p p(G;-\lambda)$ *for all* λ.

(ii) $\psi(G;\lambda) = \psi(G-e;\lambda) + \psi(G/e;\lambda)$.

(iii) $\psi(G;\lambda) = \lambda^r\sum_{j=0}^{p-r}t_j(\lambda+1)^{p-r-j}$ *with integer coefficients $t_j \geq 0$, where r is the number of components in G.*

Proof. We have

$$\begin{aligned}(-1)^p p(G;-\lambda) &= \sum(-1)^{p+i}\alpha_i(-1)^{p-i}\lambda^{p-i}\\ &= \sum\alpha_i\lambda^{p-i} = \psi(G;\lambda).\end{aligned}$$

The recursion (ii) now follows directly by substitution from the corresponding recursion for the chromatic polynomials. Because of 3.4(iv), it suffices in (iii) to consider connected graphs G, and to verify that here $\psi(G;\lambda) = \lambda\sum_{j-0}^{p-1}t_j(\lambda+1)^{p-1-j}$. By 5.15,

$$\psi(T;\lambda) = (-1)^p p(T;-\lambda) = (-1)^p(-\lambda)(-\lambda-1)^{p-1} = \lambda(\lambda+1)^{p-1}$$

holds for trees T. Therefore (iii) is satisfied for trees. The rest follows, by induction on the number of edges , from the recursion (ii), where an edge e is chosen that is not a bridge. ■

Now a couple of remarkable conclusions can easily be drawn from 5.16(iii). Because of 3.4(iv), we may again restrict ourselves to connected graphs. The corresponding results clearly hold in general.

5.19 THEOREM. *Let G be a connected simple graph with p vertices and q edges and let $p(G;\lambda) = \sum_{i-0}^{p-1}(-1)^i\alpha_i\lambda^{p-i}$ be the chromatic polynomial. (Note: by 3.4(iii), $a_p = 0$.) Then*

$$\binom{p-1}{i} + (q-p+1)\binom{p-2}{i-1} \leq \alpha_i \leq \binom{q}{i}$$

holds for all $i \geq 1$.

Proof. By 5.15, the upper bound is valid for trees. If α_i', α_i'' are the coefficients of $G-e$, resp. G/e, then by 5.18(ii) $\alpha_i = \alpha_i' + \alpha_{i-1}''$ holds for $i \geq 1$, $\alpha_0 = \alpha_0' = 1$. By induction we deduce from

$$\alpha_i' \leq \binom{q-1}{i}, \quad \alpha_{i-1}'' \leq \binom{q-1}{i-1},$$

using the recursion for the binomial coefficients, that

$$\alpha_i = \alpha_i' + \alpha_{i-1}'' \leq \binom{q-1}{i} + \binom{q-1}{i-1} = \binom{q}{i}.$$

If we set $r = 1$ in 5.18(iii) then we obtain

$$\sum_{i=0}^{p-1} \alpha_i \lambda^{p-i} = \lambda \sum_{j=0}^{p-1} t_j (\lambda+1)^{p-1-j},$$

and from this, by comparison of coefficients, that

$$(\star) \qquad \alpha_i = \sum_{j=0}^{i} \binom{p-1-j}{i-j} t_j.$$

Since all $t_j \geq 0$, it follows in particular that $\alpha_i \geq \binom{p-1}{i} t_0 + \binom{p-2}{i-1} t_1$. But with $\alpha_0 = 1$, $\alpha_1 = q$ substituted into $(\star)$, the values $t_0 = 1$, $t_1 = q - p + 1$ immediately result, and from this the lower bound. ∎

Note that it follows from the proof of the upper bound that equality for all i can hold only for trees. For equality in the lower bound, there are other easily determined examples besides the trees.

In addition to bounds, 5.18 also permits a statement about the total behavior of the number sequence $(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{p-1})$. For trees, we obtain by Lemma 5.15, the sequence of binomial coefficients

$$\binom{p-1}{0}, \binom{p-1}{1}, \ldots, \binom{p-1}{p-1}.$$

It is a well-known fact that this sequence first increases strictly to the maximum $\binom{p-1}{\lfloor \frac{p-1}{2} \rfloor}$ (which is assumed twice when p is even) and then strictly decreases. In general, one calls sequences with this property **unimodal** and it is conjectured that the sequence $(\alpha_0, \alpha_1, \alpha_2, \ldots)$ is unimodal for *every* chromatic polynomial. The following theorem proves half of this conjecture.

5.20 THEOREM. *Let G be a connected simple graph with $p \geq 3$ vertices and chromatic polynomial $p(G; \lambda) = \sum_{i=0}^{p-1} (-1)^i \alpha_i \lambda^{p-i}$. Then the following hold:*

(i) $\alpha_k < \alpha_l$ *for* $0 \leq k < (p-1)/2$, *and* $k < l < p - 1 - k$,

(ii) $\alpha_k \leq \alpha_{p-1-k}$ *for* $0 \leq k < (p-1)/2$, *and in particular,*

(iii) $\alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{\lfloor p/2 \rfloor - 1} \leq \alpha_{\lfloor p/2 \rfloor}$,

where in (ii) *and* (iii) *we have strict inequality when G is not a tree.*

Proof. We have already verified the assertion for trees. Therefore, assume G is not a tree. Then the coefficient $t_1 = q - p + 1 > 0$. We apply the expression $(\star)$ from the proof of the preceding theorem to the α_i's. For $0 \le k < (p-1)/2$ and $k < l \le p-1-k$ we have $\binom{p-1-j}{k-j} \le \binom{p-1-j}{l-j}$ for $j = 0, 1, \ldots, k$, and thus $\binom{p-1-j}{k-j}t_j \le \binom{p-1-j}{l-j}t_j$. Since $t_1 > 0$, we have strict inequality for $j = 1$, from which it follows from $(\star)$ that $\alpha_k < \alpha_l$. Now (ii) and (iii) follow directly. ∎

In 3.4, 5.19, and 5.20 we deduced a number of properties that the coefficients of a chromatic polynomial must satisfy. However in no way do these properties suffice to characterize a given polynomial as being chromatic.

The polynomial $\lambda^4 - 4\lambda^3 + 5\lambda^2 - 3\lambda$ serves as an example. The reader can easily verify that all the given conditions are satisfied, but that there does not exist a corresponding graph. Therefore, here is another open problem: Characterize the chromatic polynomials!

After dealing with the coefficients, we turn to our last problem, that of the **roots** of the chromatic polynomial. The prime motivation is, of course, again the 4-color problem in its equivalent form: Does $p(G;4) > 0$ hold for every plane graph? Or, formulated differently: Is four never a root of $p(G;\lambda)$ when G is a plane graph?

Where do the real roots of an arbitrary chromatic polynomial lie? By 3.4(iv), we can again restrict our attention to connected graphs. A first result is contained in the following lemma.

5.21 LEMMA. *Let G be a connected graph with $p \ge 3$ vertices. Then the following hold:*

(i) $p(G;\lambda) \ne 0$ *for* $\lambda < 0$.

(ii) $p(G;\lambda) \ne 0$ *for* $0 < \lambda < 1$.

(iii) $p(G;0) = p(G;1) = 0$; 0 *is a simple root,* 1 *is a root of multiplicity m where m is the number of blocks.*

(iv) *If λ_0 is the largest real root, then* $\lambda_0 \le p - 1$.

Proof. If we write $p(G;\lambda)$ in the form

$$p(G;\lambda) = \lambda^p - \alpha_1\lambda^{p-1} + \alpha_2\lambda^{p-2} - \cdots + (-1)^{p-1}\alpha_{p-1}\lambda,$$

then we immediately see that $p(G;\lambda) > 0$ for p even and $p(G;\lambda) < 0$ for p odd, when $\lambda < 0$, since $\alpha_i > 0$ for all i. Thus (i) holds. For (ii) we show more precisely that $(-1)^{p-1}p(G;\lambda) > 0$ for $0 < \lambda < 1$. By 5.15, this is surely true for trees. As before, we use induction on the number of edges. If G is not a tree and e is not a bridge, then $G-e$ and G/e are connected and we obtain

$$(-1)^{p-1}p(G;\lambda)=(-1)^{p-1}p(G-e;\lambda)+(-1)^{p-2}p(G/e;\lambda)>0.$$

0 and 1 are roots since G is not 1-colorable and since G is connected. Thus, by 3.4(ii), 0 is a simple root. If two blocks B and B' are joined at the cut vertex v, then by 5.14 we have

$$p(B\bigcup B';\lambda)=p(B;\lambda)p(B';\lambda)/\lambda,$$

and by induction we immediately deduce from this that

$$p(G;\lambda)=\Big(\prod_{i=1}^{m}p(B_i;\lambda)\Big)\Big/\lambda^{m-1},$$

where $B_1,\ldots,B_m$ are the blocks of G. Clearly, $p(B_i;1)=0$ holds for all these blocks, i.e., it remains to show that 1 is a simple root for every 2-connected graph. We shall show more precisely that in this case

$$(-1)^p\frac{p(G;\lambda)}{(\lambda-1)}\Big|_{\lambda=1}>0$$

holds. As always, we can assume that G is simple. For $p=3$, we must have $G=K_3$, and for K_3 we obtain

$$(-1)^3\frac{\lambda(\lambda-1)(\lambda-2)}{(\lambda-1)}\Big|_{\lambda=1}>0.$$

Therefore, let $p\geq 4$. By induction it suffices to verify that either $G-e$ or G/e is 2-connected. If we assume that G/e has a cut vertex, then this must be the vertex arising from the contraction of $e=uv$, i.e., $\{u,v\}$ must be a separating vertex set of G. Let W_1 and W_2 be two u,v-paths in different components of $G-\{u,v\}$. $W_1\bigcup W_2$ yields a circuit and it follows from that without great effort that $G-e$ is 2-connected.

To prove (iv) we consider an alternate expression for $p(G;\lambda)$. Each coloring of G corresponds to a partition P of the vertex set into classes of independent sets. For each such partition P there are obviously exactly $\lambda(\lambda-1)\cdots(\lambda-|P|+1)$ λ-colorings ($|P|$ =the number of classes), whose color classes are precisely the classes of P. From this it follows that

$$p(G;\lambda)=\sum_P\lambda(\lambda-1)\cdots(\lambda-|P|+1)$$

summed over all partitions P. But the right-hand expression is always positive for $\lambda>p-1$ since $|P|\leq p$ holds for all partitions, and thus also $p(G;\lambda)>0$ for all $\lambda>p-1$. ∎

By 5.21, the curve of a chromatic polynomial looks roughly like that in Figure 5.15.

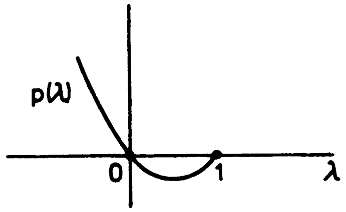


Figure 5.15

Where are the other roots? Obviously all natural numbers $0, 1, 2, \ldots, \chi(G)-1$ are roots, but, of course, we want to infer precisely the converse. That is, we want to deduce from the structure of the chromatic polynomial the smallest natural number $\chi(G)$ for which $p(G; \chi(G)) > 0$. Of particular interest are the plane graphs. From Chapter 1 we know that it suffices to prove the 4-CC for normal cubic maps. Their dual graphs have the property that all countries are bounded by 3 edges. Accordingly, they are called **triangulations** of the plane and our problem thus takes on the form: Where are the roots of $p(G; \lambda)$ when G is a triangulation?

With a line of reasoning similar to that used in 5.21, it can be shown that the chromatic polynomial of a triangulation also has no roots between 1 and 2. We proceed to the next interval, $2 < \lambda < 3$. Here Berman and Tutte discovered in 1969 an interesting phenomenon. Let $\tau = (1+\sqrt{5})/2$ be the positive root of the equation $x^2 - x - 1 = 0$. Then all the chromatic polynomials calculated until now have a root close to $\tau + 1 = 2.61803\ldots$, while $\tau + 1$ itself can not be a root, as we are going to prove in a moment. A year later, Tutte was able to formulate quantitatively the explanation for this phenomenon: For each triangulation of G with p vertices, $|p(G; \tau + 1)| \leq \tau^{5-p}$. From this it follows in particular: If the number of vertices of a triangulation is sufficiently large, then $|p(G; \tau + 1)|$ is arbitrarily small and one can expect that a root lies close to $\tau + 1$.

5.22 LEMMA. $p(G; \tau + 1) \neq 0$ *for every graph* G.

Proof. We can assume G to be connected. The polynomial $\lambda^2 - 3\lambda + 1$ has the roots $\tau + 1 = (3 + \sqrt{5})/2$ and $\eta = (3 - \sqrt{5})/2$. Since $\lambda^2 - 3\lambda + 1$ is an irreducible polynomial over the field Q of rational numbers, i.e., it is not factorable in Q, it follows from theorems in algebra that $p(G; \lambda)$ would also have to have η as a root together with $\tau + 1$. But, because of $0 < \eta < 1$ and 5.21(ii), this is impossible. ∎

The number τ is one of the fundamental numbers of all mathematics. The Greeks called it the "golden ratio". The reason for this is the following: If we take a rectangle with side lengths a and b, $a \leq b$, then this rectangle is found to be particularly esthetically pleasing (and this has been verified empirically) if the ratio of b to a is roughly equal to the ratio of a to the difference $b-a$ (see Figure 5.16). If we denote the ratio sought by $\tau = b/a$, then we obtain

$$\tau = \frac{b}{a} = \frac{a}{b-a} = \frac{1}{\frac{b}{a}-1} = \frac{1}{\tau - 1},$$

and thus $\tau^2 - \tau - 1 = 0$. This means that τ is the previously determined number $\tau = (1+\sqrt{5})/2 = 1.61803\cdots$. τ arises in almost all areas of mathematics, especially in combinatorics. As the best known example we mention the Fibonacci numbers, defined by the recursion $F_0 = 0$, $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$: hence $F_2 = 1$, $F_3 = 2$, $F_4 = 3$, $F_5 = 5$, $F_6 = 8$ etc. The sequence of fractions $(F_{n+1}/F_n) = \{1, 2, 3/2, 5/3, 8/5, \ldots\}$ tends to the golden ratio.

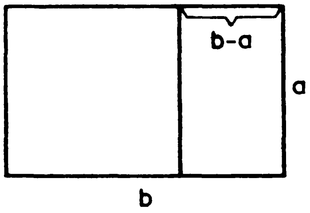


Figure 5.16

We conclude this chapter with a theorem due to Tutte (1968) which is often called the Golden Ratio Theorem in the theory of chromatic polynomials.

5.23 THEOREM (Tutte). *Let G be a triangulation with p vertices. Then* $p(G;\tau+2) = (\tau+2)\tau^{3p-10}p^2(G;\tau+1)$.

Proof. Using the defining equation $\tau^2 - \tau - 1 = 0$, we deduce the relations

$$\text{(i)}\quad \tau^{-1} = \tau - 1, \qquad \text{(ii)}\quad \tau^2 = \tau + 1, \qquad \text{(iii)}\quad \tau^3 = 2\tau + 1.$$

We use induction on the number p of vertices. For $p = 3$, we have $G = K_3$, and by substitution into $p(K_3;\lambda) = \lambda(\lambda-1)(\lambda-2)$ one sees that the identity holds. To prepare the induction step, we first establish two equations. Let G be a plane graph with a country F that is bounded by 4 edges. Let the vertices of the boundary be a, b, c, and d. We insert initially the edge $e = ac$, and then the edge $f = bd$ and denote the resulting graphs by G_1 resp. G_2 (see Figure 5.17).

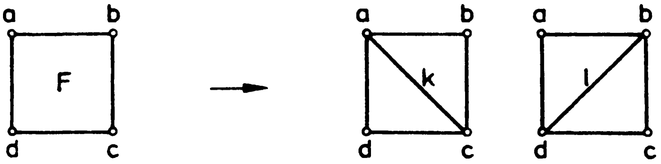


Figure 5.17

Claim: (iv) $p(G;\lambda) = p(G_1;\lambda) + p(G_1/e;\lambda) = p(G_2;\lambda) + p(G_2/f;\lambda)$

(v) $p(G;\tau+1) = (\tau+1)(p(G_1;\tau+1) + p(G_2;\tau+1))$.

Since $G = G_1 - e = G_2 - f$, (iv) is nothing other than the usual recursion equation for chromatic polynomials. To prove (v) we apply induction on the number of edges. Assume that G has only the vertices of F. Then G is either a circuit of length 4 or a and c, resp. b and d, (but not both) are joined by an edge outside of F. In the first case we have $p(G;\lambda) = \lambda(\lambda-1)(\lambda^2 - 3\lambda + 3)$, in the second case $p(G;\lambda) = \lambda(\lambda-1)(\lambda-2)^2$ (see the example following 5.14). The chromatic polynomials of G_1 and G_2 are just as easy to calculate and by substitution of $\lambda = \tau + 1$ we see, taking (i) through (iii) into account, the validity of (v) in all cases. We now assume that G contains other vertices. If none of these vertices are adjacent to $\{a,b,c,d\}$, then, by use of 3.4(iv), we are finished. Now let g be an edge that has an end point in F. We set $G' = G - g$, $G'_1 = G_1 - g$, $G'_2 = G_2 - g$ and $G'' = G/g$, $G''_1 = G_1/g$, $G''_2 = G_2/g$. Then by the induction hypothesis we have

$$p(G';\tau+1) = (\tau+1)(p(G'_1;\tau+1) + p(G'_2;\tau+1))$$

$$p(G'';\tau+1) = (\tau+1)(p(G''_1;\tau+1) + p(G''_2;\tau+1))$$

and (v) follows by subtraction of these two equations.

We return to the proof of the theorem. Let G be a triangulation with a minimal number of vertices p for which the identity is not valid. Among all counterexamples with p vertices we choose a graph G_1 for which the maximal degree $\Delta(G_1)$ is as large as possible. Let u be a vertex in G_1 with $d(u) = \Delta(G_1)$ and $\{u,v,w\}$ a triangle that contains u. Furthermore, let $\{u',v,w\}$ be another triangle with the edge $e = uv$. We then assert that u necessarily equals u'. Let us assume the contrary, $u \neq u'$. We set $G = G_1 - e$, $G_2 = G \bigcup f$, with $f = uu'$ as in Figure 5.18.

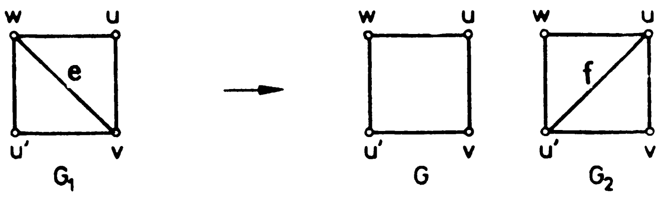


Figure 5.18

After striking out the multiple edges (which have no effect on the chromatic polynomial), the graphs G_1/e and G_2/f are again triangulations with fewer than p vertices. The triangulation G_2 has p vertices with $\Delta(G_2) \geq \Delta(G_1)+1$; therefore the asserted identity holds for all three graphs G_1/e, G_2/f, and G_2. By (iv) and (v) we have

$$\begin{aligned}p(G_1;\tau+1)+p(G_1/e;\tau+1)+p(G_2;\tau+1)+p(G_2/f;\tau+1) &= 2p(G;\tau+1)\\ &= 2(\tau+1)(p(G_1;\tau+1)+p(G_2;\tau+1)),\end{aligned}$$

and thus, by (iii),

$$\begin{aligned}p(G_1/e;\tau+1)+p(G_2/f;\tau+1) &= (2\tau+1)(p(G_1;\tau+1)+p(G_2;\tau+1))\\ (\star)\qquad &= \tau^3(p(G_1;\tau+1)+p(G_2;\tau+1)).\end{aligned}$$

Furthermore, by (iv), we have

$$(\star\star)\qquad p(G_2/f;\tau+1)-p(G_1/e;\tau+1) = p(G_1;\tau+1)-p(G_2;\tau+1).$$

If we multiply $(\star)$ by $(\star\star)$, then we obtain

$$p^2(G_2/f;\tau+1)-p^2(G_1/e;\tau+1) = \tau^3(p(G_1;\tau+1)-p^2(G_2;\tau+1)).$$

If we now apply the identities to the graphs G_1/e, G_2/f and G_2, then, taking (iv) into account, there results

$$\begin{aligned}p(G_1;\tau+2) &= p(G_2;\tau+2)+p(G_2/f;\tau+2)-p(G_1/e;\tau+2)\\ &= (\tau+2)\tau^{3p-10}[p^2(G_2;\tau+1)+\tau^{-3}(p^2(G_2/f;\tau+1)-p^2(G_1/e;\tau+1))]\\ &= (\tau+2)\tau^{3p-10}[p^2(G_2;\tau+1)+p^2(G_1;\tau+1)-p^2(G_2;\tau+1)]\\ &= (\tau+2)^{3p-10}p^2(G_1;\tau+1),\end{aligned}$$

hence the identity is indeed satisfied by G_1.

Now the proof can be concluded quickly. We have just verified that if a country F has the vertex u as a boundary point then all countries adjacent to F must also have u as a boundary point. But this means that u is a boundary point of all countries. In particular, it follows that G_1-u contains no circuit and that u is adjacent to all other vertices (G_1 is a triangulation!). Since a triangulation is clearly 2-connected, G_1-u is connected, i.e., it must be a tree. If we want to color G_1 with λ colors, then we can give u one of the colors λ and color the rest of G_1-u arbitrarily with $\lambda-1$ colors. With 5.15, we then obtain $p(G_1;\lambda)=\lambda p(G_1-u;\lambda-1)=\lambda(\lambda-1)(\lambda-2)^{p-2}$ and, since $\tau^2=\tau+1$,

$$p(G_1;\tau+2) = (\tau+2)(\tau+1)\tau^{p-2} = (\tau+2)\tau^p.$$

On the other hand, because of (i) and (ii), we also have

$$\begin{aligned}(\tau+2)\tau^{3p-10}p^2(G_1;\tau+1) &= (\tau+2)\tau^{3p-10}(\tau+1)^2\tau^2(\tau-1)^{2(p-2)}\\ &= (\tau+2)\tau^{3p-10}\tau^4\tau^2\tau^{-2(p-2)} = (\tau+2)\tau^p.\end{aligned}$$

Therefore G_1 would not be a counterexample and the proof is complete. ∎

5.24 COROLLARY. *For every triangulation G of the plane, $p(G; \tau + 2) > 0$.*

The numerical value of $\tau + 2$ is $\tau + 2 = 3.618033\cdots$. With 5.24 we have come somewhat closer to our goal of verifying $p(G; 4) > 0$ for all triangulations. Are the numbers $\tau + 1$ and $\tau + 2$ special cases for which we can verify $p(G; \lambda) \neq 0$? Perhaps. But it is just as likely that there is a special sequence of such numbers b_n that converges to 4 and which could yield information about $p(G; 4)$ from the behavior of $p(G; b_n)$. Beraha proposed such a sequence in 1974: $b_n = 2 + 2\cos(2\pi/n)$. The first values are $b_1 = 4$, $b_2 = 0$, $b_3 = 1$, $b_4 = 2$. Our numbers $\tau + 1$ and $\tau + 2$ likewise appear: $\tau + 1 = b_5$, $\tau + 2 = b_{10}$, and clearly (b_n) converges to 4. Now whether the numbers b_n in fact touch upon the core of the 4-color problem (and further work by Tutte suggests that they do) or not, in any case it is clear that the potential of the quantitative method is far from being exhausted.

EXERCISES FOR CHAPTER 5

1°. Let G be a simple graph with p vertices and $\overline{G}$ its complement. Show: a. $2\sqrt{p} \leq \chi(G) + \chi(\overline{G}) \leq p + 1$. b. $p \leq \chi(G)\chi(\overline{G}) \leq ((p+1)/2)^2$. Give examples for which equality holds.

2. Let G be a simple graph with p vertices and q edges. Show: a. $q \geq \binom{\chi(G)}{2}$. b. $(p^2/(p^2 - 2q)) \leq \chi(G) \leq 1 + \sqrt{2q(p-1))/p}$. Give examples for equality.

3. Let n, k be natural numbers with $2k \leq n$. The Kneser graph $K(n, k)$ is defined as follows: The vertices are all k-subsets of an n-set and two such k-sets are joined if and only if they are disjoint. Show: $\chi(K(n, k)) \leq n - 2k + 2$. What does $K(5, 2)$ look like?

4. For what graphs G is the line graph $L(G)$ isomorphic to G?

5°*. Show: G is n-colorable if and only if the edges can be oriented so that there are at least $\ell(C)/n$ edges oriented in each of the two directions on each polygon C.

6. Verify the preceding exercise for the wheels W_n and the Petersen graph.

7°. Let G be a graph. Order the vertices in such a way that the algorithm $\mathcal{A}$ (see the proof of 5.4) requires exactly $\chi(G)$ colors.

8. Determine the chromatic number of the cube Q_n.

9. A graph is said to be uniquely n-colorable if $\chi(G) = n$ and each n-coloring yields the same partition of the vertex set. Determine the smallest uniquely 3-colorable graphs that are different from K_3.

10°. Prove: No plane graph is uniquely 5-colorable.

11. Let G be a k-regular graph. Show that $\chi'(G) = k + 1$ holds for the chromatic index if G has a cut vertex.

12. Show: $\chi'(K_{m,n}) = \max(m, n)$.

13°. The graph G_4 in Figure 5.10 satisfies $\omega(G_4) = 2$ and $\chi(G_4) = 4$. Show that G_4 is the smallest such graph.

14. Show: For a bipartite graph G, $\omega(\overline{G}) = \chi(\overline{G})$.

15. Let G be a critical graph and let $A \subseteq V$ be an independent set of vertices. Show that $\chi(G - A) = \chi(G) - 1$.

16. The total chromatic number $\chi_t(G)$ is the minimal number of colors that are needed to color the vertices *and* the edges, so that adjacent vertices, incident edges and incident vertices-edges obtain different colors. Estimate $\chi_t(K_p)$ and $\chi_t(K_{m,n})$.

17°. Use the inclusion-exclusion principle to determine the number of fixed-point free bijections of an n-set.

18. Determine all graphs for which equality holds for the lower bound in 5.19, for all i.

19. Prove the statement concerning the Fibonacci numbers, that $F_{n+1}/F_n \to \tau$.

20*. A little more difficult. Show that the chromatic polynomial of a triangulation has no roots between 1 and 2.

6. FACTORIZATION

We recall the definition of an r-factor of a graph $G(V, E)$. An r-factor of G is an r-regular subgraph $H(V, E')$, i.e., each vertex v has degree r in H. The point of departure for interest in factorization was Tait's Theorem, 1.7, which reads as follows in the terminology of factorizations: The 4-color conjecture holds if and only if the skeleton of each bridgeless cubic map $\mathcal{L}$ decomposes into three disjoint 1-factors. Petersen's result, 3.6, showed the existence of at least *one* 1-factor in each *bridgeless* 3-*regular* graph (plane or not). We now pose the general question: ***Which graphs have a* 1*-factor*?**

As a trivial condition we note that the graph must exhibit an even number of vertices. For the complete graphs, K_p, we have already seen that this is also sufficient (in that case, even a 1-factorization occurs) but in general this does not suffice. For example, the complete bipartite graph $K_{2,4}$ has 6 vertices, however it clearly contains no 1-factor (see Figure 6.1).

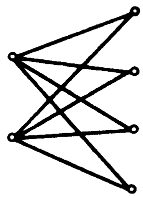


Figure 6.1

A preliminary remark: Since 1-factors contain no loops and at most one multiple edge, we can restrict ourselves in the question of the existence of a 1-factor to **simple** graphs. We shall do this without further mention.

The first class of graphs in which the problem of the existence of 1-factors was studied and completely solved were the bipartite graphs. It was the Hungarian mathematician D. König around 1910, who took up this concept and introduced it to graph theory (on which he made a lasting impression with his book of 1936). Let $G(S \bigcup T, E)$ be a bipartite graph with defining vertex sets S and T, where $|S| = m \leq n = |T|$. Then each edge has one end point in S and the other in T. Clearly G can have a 1-factor only if $m = n$. Now we shall tackle the general question, what is the maximal number of edges such that no two have a common vertex. For such edge sets (and indeed in arbitrary graphs) we today use the term **matching**. Hence a matching is an edge set $M \subseteq E$ such that every vertex in the subgraph $H(V, M)$ has degree 1 or 0. Thus we ask about the **matching number**, $m(G) = \max_M |M|$, where M ranges over all matchings. Figure 6.2 shows a bipartite graph G in which the 4 bold edges form a matching. The reader can easily convince himself that a matching with 5 edges does not exist, so that $m(G) = 4$.

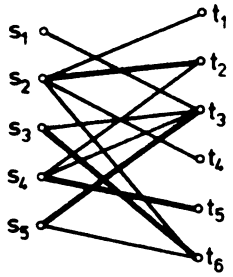


Figure 6.2

The word matching stems from an amusing interpretation of the problem. Let S be a set of ladies and T a set of gentlemen. We join the lady s_i and the gentleman t_j by an edge if the two are not opposed to marriage. Then how many weddings can occur (without practicing bigamy)? This is exactly our matching number $m(G)$. In 1931 König expressed the number $m(G)$ in terms of another graph parameter. His work at first was considered unimportant (there even developed a strangely sharp controversy between him and the famous algebraist Frobenius), but the significance of his theorem was soon recognized and today it appears as one of the fundamental results of all of combinatorics. It opened the door to one of the most fruitful areas, that of **transversal theory**.

We still need another definition. A subset D of vertices in an arbitrary graph G is called a **support**, if each edge has at least one endpoint in D. If M is a matching, then every support must contain at least one endpoint of every edge of M. Since all of these vertices are distinct, we conclude that $|M| \leq |D|$ for each matching M and each support D. Reformulated, this means that $\max |M| \leq \min |D|$ must hold, and indeed for an arbitrary graph G. In general, this inequality will be strict. As an example, in a circuit C_5 of length 5, $\max |M| = 2 < 3 = \min |D|$. However, for bipartite graphs we always have equality, which is the content of the following theorem.

6.1 THEOREM (König). *If $G(S \bigcup T, E)$ is a bipartite graph, then* $\max |M| = \min |D|$, *where M ranges over all matchings and D ranges over all supports.*

1. **Proof.** We use Menger's Theorem, 4.4, which was proven in 1927, some time before Theorem 6.1 was known (even though König's Theorem was not proven in this way). We

add two new vertices u and v to G and join u with all vertices of S and v with all vertices of T. We denote the new graph by $\widetilde{G}$. Figure 6.3 shows the new graph $\widetilde{G}$ corresponding to the graph G in Figure 6.2.

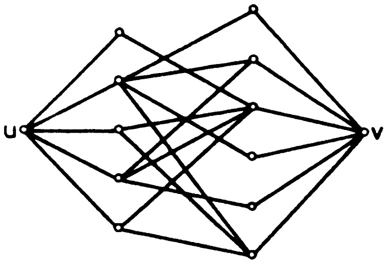


Figure 6.3

Which vertex sets separate u from v in $\widetilde{G}$? Obviously every edge of G must be involved, i.e., every separating vertex set must contain at least one endpoint of each edge of G. But these are exactly the supports of G, and we conclude that $\kappa_{\widetilde{G}}(u,v) = \min|D|$. It is also as clear, however, that disjoint paths from u to v in $\widetilde{G}$ yield a matching in G, so that $\mu_{\widetilde{G}}(u,v) = \max|M|$. Thus the equation $\kappa_{\widetilde{G}}(u,v) = \mu_{\widetilde{G}}G(u,v)$ in **4.4** implies the desired relation $\max|M| = \min|D|$. ∎

2. Proof. We give another proof that uses the inherent bipartite structure of the graph. For brevity, we write $m(G) = \max|M|$, $d(G) = \min|D|$. Since we know that $m(G) \leq d(G)$ always holds, we only need to show that $d(G) \leq m(G)$. We do this by induction on the number of edges. If G has no edges at all, then $m(G) = d(G) = 0$ holds trivially. Now let $G(S \bigcup T, E)$ be a bipartite graph with $q \geq 1$ edges and $|S| \leq |T|$. Since isolated vertices appear neither in matchings nor in supports, we can assume that none occur. Every support D is of the form $D = X \bigcup Y$ with $X \subseteq S$ and $Y \subseteq T$. Assume G has a minimum support $D = X \bigcup Y$ with $X \neq \emptyset$, $Y \neq \emptyset$. Then the two induced subgraphs H_X, resp. H_Y, have fewer edges on the vertex set $X \bigcup (T - Y)$, resp. $(S - X) \bigcup Y$, than G, and no edges run between $S - X$ and $T - Y$. (See Figure 6.4.)

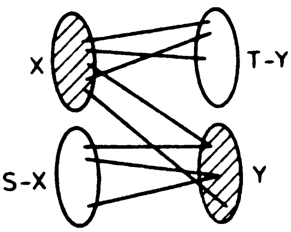


Figure 6.4

For H_X we must have $d(H_X) = |X|$, since every support of H_X together with Y gives a support of G, and analogously $d(H_Y) = |Y|$. By our induction assumption, there is a matching M_X in H_X with $|M_X| = |X|$, and in H_Y with $|M_Y| = |Y|$. But $M = M_X \bigcup M_Y$ is clearly a matching in G with $|M| = |X| + |Y| = d(G)$, and we are finished. The case remains where the only minimum supports are S, or possibly T (in case $|T| = |S|$). If each vertex in T has degree 1, then we are again finished, since it is obvious then that $m(G) = |S| = d(G)$. (See Figure 6.5.) Hence we can assume that there are vertices $u,v \in S$ and $w \in T$, with $uw, vw \in E$. By the induction assumption, for the graph $G_1 = G - vw$ we have $m(G_1) = d(G_1)$. Since clearly $m(G_1) \leq m(G)$ and $d(G_1) \leq d(G)$, we can assume that $d(G_1) < d(G) = |S|$. Let $D_1 = X_1 \bigcup Y_1$ be a minimum support of G_1. Since D_1 is not a support of G, we conclude that $u \in S - X_1$, $w \in T - Y_1$ and furthermore $v \in X_1$ (since otherwise the edge vw in D_1 would not be involved). Thus $X_1 \neq \emptyset$. We have $|Y_1| < |S - X_1|$, since $d(G_1) = |X_1| + |Y_1| < |S| = d(G)$. For the support $D = X_1 \bigcup (Y_1 \bigcup \{w\})$ of G, there now holds $|D| = |X_1| + |Y_1| + 1 \leq |X_1| + |S - X_1| = |S|$. Therefore D is minimal, and, since $X_1 \neq \emptyset$, and $Y_1 \bigcup \{w\} \neq \emptyset$, we have again arrived at the preceding case. ∎

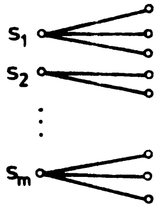


Figure 6.5

Let us look at minimal supports more closely. For $A \subseteq S$, let $R(A)$ be that subset of the vertices of T which one joined to at least one vertex from A. If the minimal support D from S contains exactly the vertices $S - A$, then it must also contain at least the vertices of $R(A)$. However, since every set $(S - A) \bigcup R(A)$ is a support (no edges run between A and $T - R(A)$), we see that each minimal support is of the form $D = (S - A) \bigcup R(A)$. Thus we obtain the following version of König's Theorem, which is due to Ore.

6.2 THEOREM. *If $G(S \bigcup T, E)$ is a bipartite graph, then $m(G) = |S| - \max_{A \subseteq S} \delta(A)$, where $\delta(A)$ is defined as $|A| - |R(A)|$. $\delta(A)$ is called the* **defect** *of the set A.*

Proof. By the line of reasoning above, we have

$$\begin{aligned} m(G) &= \min |D| = \min_{A \subseteq S} (|S - A| + |R(A)|) \\ &= \min_{A \subseteq S} (|S| - |A| + |R(A)|) \\ &= |S| - \max_{A \subseteq S} (|A| - |R(A)|). \end{aligned}$$ ∎

The formulation 6.2 pertains to the set S and its subsets. But naturally T and the defects $\delta(B)$, with $B \subseteq T$, can just as well be used, so that $m(G) = |T| - \max_{B \subseteq T} \delta(B)$ results. This "duality" will be useful in the following.

If $\delta(A) > 0$, i.e., $|A| > |R(A)|$, then no matching can contain all vertices of A. To be precise: A matching can contain at most $|A| - \delta(A)$ vertices from A, thus the name defect. From 6.2 we can immediately read off the famous theorem of Hall (1935) that was the starting point for transversal theory, which we will sketch briefly in the following.

6.3 THEOREM (Hall). *If $G(S \bigcup T, E)$ is a bipartite graph, then $m(G) = |S|$ if and only if $|A| \leq |R(A)|$ holds for all $A \subseteq S$.*

In the literature, Theorem 6.3 is often referred to as the "marriage theorem", since, in the interpretation that we gave earlier, it gives the exact conditions under which *all* ladies can be married. Theorem 6.3 says that this is possible exactly when for each k ladies, there are always at least k gentlemen that propose marriage to at least one of them. That this condition is necessary is obvious—therefore the importance of the theorem rests on the converse, that this condition is also sufficient. In general, it is a characteristic of maximum-minimum theorems that one direction, namely, max $\leq$ min is trivial and the significance lies in the verification of the opposite inequality.

Now, as a consequence we state the solution of the 1-factor problem for bipartite graphs.

6.4 COROLLARY. *A bipartite graph $G(S \bigcup T, E)$ has a 1-factor if and only if $|S| = |T|$ and $|A| \leq |R(A)|$ holds for all $A \subseteq S$.*

Now, what is this transversal theory that has been mentioned a couple of times? It is based on an interpretation of bipartite graphs as set systems. If on a basic set $N = \{t_1, \ldots, t_n\}$ there is given a family $\mathcal{A} = \{A_1, \ldots, A_m\}$ of subsets $A_i \subseteq N$ (which need not be distinct), then we call $(N; \mathcal{A})$, resp. $(N; A_1, \ldots, A_m)$, a **set system**. We assign to $(N; \mathcal{A})$, in an obvious way, a bipartite graph G. The vertex sets of G are $\mathcal{A}$ and N and we join A_i with t_j when $t_j \in A_i$.

EXAMPLE. $N = \{1,2, \ldots, 6\}$, $A_1 = \{3\}$, $A_2 = \{1,2,4,6\}$, $A_3 = \{3,6\}$ $A_4 = \{2,3,5\}$, $A_5 = \{1,3\}$. *The corresponding bipartite graph is depicted in Figure 6.6.*

It is immediately clear that this construction can also be reversed, if we take T as the underlying set in a bipartite graph $G(S \bigcup T, E)$ and identify $s_i \in S$ with the set $A_i = R(s_i)$. Bipartite graphs and incidence structures of set systems are thus combinatorially equivalent. This is not remarkable in itself, but what makes the situation interesting is the fact that we can now study set-theoretical problems by means of graph concepts, and conversely.

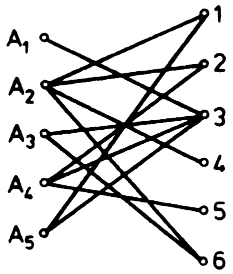


Figure 6.6

Suppose the set system $(N; \mathcal{A})$ and the corresponding bipartite graph G are given. To a matching in G there corresponds a set of pairs $(A_{i_1}, t_{i_1}), \dots, (A_{i_l}, t_{i_l})$ where all indices i_j and all elements t_{i_j} are distinct, and $t_{i_j} \in A_{i_j}$ holds for all j. Then, we say that the subfamily $\mathcal{A}' = (A_{i_j} : j = 1, \dots, l)$ has a **system of distinct representatives** or a **transversal** $(t_{i_1}, \dots, t_{i_l})$. For $I \subseteq \{1, \dots, m\}$, let $\mathcal{A}_I = (A_i : i \in I)$, then $R(\mathcal{A}_I) = \bigcup_{i \in I} A_i$ clearly holds, and this leads directly to the fundamental theorem of transversal theory.

6.5 THEOREM (Hall). *Let $(N; A_1, \dots, A_m)$ be a set system. $(A_1, \dots, A_m)$ has a transversal precisely when $|\bigcup_{i \in I} A_i| \geq |I|$ for all $I \subseteq \{1, \dots, m\}$. In words: $\mathcal{A}$ has a transversal exactly when each k sets from $\mathcal{A}$ together contain at least k elements.*

The reader will now be able to easily derive the general formula for the size $m(N; \mathcal{A})$ of a largest transversal from 6.2. The dual form is also interesting. For $B \subseteq N$, we have $R(B) = \{A_i : A_i \bigcap B \neq \emptyset\}$ from which another formula for $m(N; \mathcal{A})$ results.

As stated, 6.3 was the point of departure for an enormous number of results, and variations on the theme: transversals of set systems. Here, we will only briefly consider two aspects that are connected with graph theory proper.

We return once again to the marriage problem. Assume that every lady s_i can imagine marriage with any one of exactly k gentlemen ($k \geq 1$) and a similar situation holds for each gentleman t_j. The number of edges in our marriage graph $G(S \bigcup T, E)$ is $k|S| = k|T|$. Therefore we have $|S| = |T|$. Let $A \subseteq S$. Then $k|A|$ edges run between A and $R(A)$. On the other hand, since exactly $k|R(A)|$ edges go out from $R(A)$, we conclude that $k|A| \leq k|R(A)|$ and thus $|A| \leq |R(A)|$. Hall's condition 6.3 is thus satisfied for every A—i.e., all ladies and gentlemen can be married. In more prosaic wording, our result states: Every k-regular bipartite graph ($k \geq 1$) has a 1-factor. But we can say even more. Namely, if we remove the edges of a 1-factor, there results a $(k-1)$-regular graph to which we can apply the same conclusion. If we continue in this manner, we obtain:

6.6 THEOREM. *Every k-regular bipartite graph ($k \geq 1$) is 1-factorable.*

Earlier, we have seen that the 1-factorizations of k-regular graphs correspond exactly to the different edge-colorings with k colors. With the help of Hall's Theorem, we can now easily make a claim about the *number* of these edge-colorings. For brevity, we let $\mathcal{G}(n,k)$ denote the class of k-regular bipartite graphs $G(S \bigcup T, E)$ with $|S| = |T| = n$.

6.7 THEOREM. *Let $G \in \mathcal{G}(n,k)$. Then G has at least $k!$ different 1-factors and at least $k!(k-1)! \cdots 2!1!$ different edge-colorings with k colors. Conversely, the number of these edge-colorings is bounded above by $(k!)^n$.*

Proof. We prove the first assertion, more generally, for all bipartite graphs $G(S \bigcup T, E)$, $|S| = |T| = n$, that have a 1-factor and in which the vertex degree $d(u) \geq k$ for all $u \in S$. We use induction on n. For $n = k$, we obtain the complete bipartite graph $K_{k,k}$. We enumerate the vertex sets S and T by $1, 2, \ldots, k$. Then it is clear that the 1-factors correspond exactly to the permutations of $\{1, 2, \ldots, k\}$ of which there are $k!$. Now let $n > k$. Let us call $A \subseteq S$ a **critical set** if $|A| = |R(A)|$. $\emptyset$ and S are naturally always critical.

Case a. S and $\emptyset$ are the only critical sets. Let uv be an arbitrary edge and let G' be the subgraph induced on $(S-u) \bigcup (T-v)$. Since $|A| < |R(A)|$ for $A \neq \emptyset, \neq S$, then $|A| \leq |R'(A)|$ always holds in G'. Thus G' satisfies Hall's condition, and by the induction assumption G' has at least $(k-1)!$ 1-factors. Hence, each of the edges $uv_1, \ldots, uv_d$, $d = d(u)$, is contained in $(k-1)!$ different 1-factors. Thus, since $d(u) \geq k$, we obtain altogether at least $k!$ 1-factors.

Case b. $A \neq \emptyset, \neq S$ is a critical set. Then, each 1-factor of G must join precisely the vertices of A with those of $R(A)$ and thus the vertices of $S - A$ precisely with those of $T - R(A)$. Let E_A be the edges incident with A. By the induction assumption, there are at least $k!$ 1-factors in $G(A \bigcup R(A), E_A)$, all of which can be extended to 1-factors in G.

The second assertion is now clear. We can choose the first color (i.e., the first 1-factor) in at least $k!$ ways. If we have chosen and fixed the first color (i.e., we have removed the 1-factor), then we can choose the second color in at least $(k-1)!$ ways, etc. The upper bound follows immediately by noting that the edges incident with a vertex $u \in S$ can be colored in $k!$ ways and by combining these colorings. ∎

When applied to the marriage problem, our result says that for example with 6 marriage proposals we can already celebrate $6! = 720$ different weddings, which is a comforting thought.

We now turn to the second aspect. We can think of a matching as a subgraph of $G(S \bigcup T, E)$ in which all vertices have degree ≤ 1. Now generalizing this we ask under what conditions does a subgraph H with prescribed vertex degrees exist. We will take up this question in full generality in Theorem 6.16. For the moment, let us derive an interesting special case directly from 6.3. Assume in the marriage situation that each lady s_i has exactly d_i marriageable gentlemen (where each gentleman must be chosen at most once). When is this possible?

We again formulate the solution of this "harem problem" in graph-theoretic language.

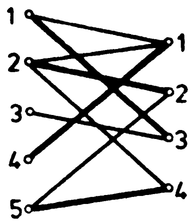
6.8 THEOREM. *Let $G(S \bigcup T, E)$ be a bipartite graph. There is a subgraph $H(S \bigcup T, E')$ such that $d_H(s_i) = d_i$ for each i and $0 \leq d_H(t_j) \leq 1$ for each j if and only if*

$$|R(A)| \geq \sum_{s_i \in A} d_i$$

holds for every $A \subseteq S$.

Proof. We replace each vertex s_i by d_i new vertices, which we join to each vertex from $R(s_i)$. Then there exists a desired subgraph H if and only if the new graph has a matching that contains all the new vertices. Now the result follows without further ado from 6.3. ∎

Beside the set systems, bipartite graphs have still another interesting and useful interpretation. If $G(S \bigcup T, E)$ is a (as always, simple) bipartite graph, then, in a natural way, we can assign a 0,1-matrix $M = (a_{ij})$ to G. The rows and columns of M correspond to the vertex sets S and T, and we set $a_{ij} = 1$ or 0 depending on whether s_i and t_j are adjacent or not. Figure 6.7 shows a graph and its corresponding matrix.



G

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

M

Figure 6.7

Obviously this is a one-to-one correspondence, so that we can identify bipartite graphs and 0,1-matrices, whenever it is advantageous. A matching in G corresponds to a set of 1's in M of which no two appear in the same row or column (see the bold edges of the graph in Figure 6.7 and the corresponding 1's in M). We call such a set of 1's a **diagonal** in M and call the number of 1's the **length** of the diagonal. A support in G corresponds to a set of rows and columns that include all the 1's of M. In the matrix of Figure 6.7, these are, for example, the rows 2,5 and the columns 1,3. König's Thereom, 6.1, reads as follows in this matrix formulation: The largest possible length of a diagonal is equal to the minimal number of rows and columns which include all 1's.

We can now translate all the matching theorems we know to statements about matrices, and thus obtain interesting matrix results. However, it is much more important to move in the other direction: We will elicit inferences about bipartite graphs from algebraic facts concerning matrices, in particular, about the number of 1-factors and edge-colorings.

Let $G(S \bigcup T, E)$ be a bipartite graph with $|S| = |T| = n$. The corresponding matrix $M = (a_{ij})$ is a square $n \times n$ matrix whose rows and columns we number from 1 to n. Clearly, to each diagonal D of M of length n there corresponds a permutation σ with

$$D = \{a_{1\sigma(i)}, a_{2\sigma(2)}, \dots, a_{n\sigma(n)}\},$$

where all $a_{i\sigma(i)} = 1$. Therefore, if we take the sum over all products $a_{1\sigma(1)} \cdots a_{n\sigma(n)}$ (σ ranges over all $n!$ permutations), then we obtain exactly the number of such diagonals of M, since only the diagonals yield the value 1. But this sum is a well-known concept in matrix theory, namely, the **permanent** per(M) of a matrix. Thus from our previous considerations we recognize that

$$\text{per}(M) = \sum_{\sigma} \prod_{i=1}^{n} a_{i\sigma(i)}$$

gives exactly the number of 1-factors of the corresponding graph. Thus:

6.9 LEMMA. *If $G(S \bigcup T, E)$ is a bipartite graph with $|S| = |T| = n$, then the number of 1-factors in G is given by*

$$\text{per}(M) = \sum_{\sigma} \prod_{i=1}^{n} a_{i\sigma(i)},$$

where M is the $0,1$-matrix corresponding to G.

Well and good, until now we have only reformulated the counting problem, this will become interesting provided that we can give formulas or at least estimates for per(M). But that is exactly the case—and the discussion of this leads us to another famous conjecture—and its solution in 1980.

For arbitrary 0,1-matrices the calculation (even approximate) of the permanent is a hopeless exercise. However we are interested mainly in special matrices, namely the matrices of graphs from the class $\mathcal{G}(n,k)$. That G is k-regular means that in the corresponding matrix all row- and column-sums equal k. We use the symbol $\mathcal{M}(n,k)$ for this class of matrices. Let $M \in \mathcal{M}(n,k)$. If we factor k from each row, then it follows that $M = kA$, where A is a matrix whose row- and column-sums all add up to 1. Furthermore, $\text{per}(M) = k^n\text{per}(A)$ clearly holds. Therefore, if we have control of the permanents of all these matrices A then we also know the permanents of all matrices from $\mathcal{M}(n,k)$. The matrices A, on which we now want to concentrate, are examples of a class of matrices that are of fundamental significance in analysis, information theory and many other areas.

DEFINITION. *An $n \times n$ matrix $A = (a_{ij})$ is said to be* **doubly stochastic** *if all entries a_{ij} are nonnegative real numbers and each row and column adds up to* 1.

Let A be an arbitrary doubly stochastic matrix. It is not clear to begin with that $\text{per}(A) > 0$, or what is the same thing, that A has a diagonal of length n with only positive entries. König showed that this, indeed, is always the case by applying his matching theorem.

6.10 THEOREM. *For every doubly stochastic matrix A, per$(A) > 0$.*

Proof. We replace each $a_{ij} \neq 0$ in A by 1 and leave the 0's unchanged. Then the new matrix A' has a diagonal of length n if and only if $\text{per}(A) > 0$. Now if A' had no such diagonal, then by the matrix form of König's Theorem (introduced above) there would be e rows and f columns in A' (and thus in A) with $e + f \leq n - 1$ that include all nonzero elements. Since each row and column in A adds up to 1, it follows from this that

$$n = \sum_{i,j=1}^{n} a_{ij} \leq e + f \leq n - 1,$$

which is absurd. ∎

One derives from analytical considerations that for fixed n the function per(A), ranging over all doubly stochastic $n \times n$, matrices has a minimum, and intuition leads one to the conjecture that this minimum is attained if and only if the entries a_{ij} are distributed as equally as possible, i.e., if and only if $a_{ij} = 1/n$ for all i,j. The matrix $J_n = (a_{ij} = 1/n)$ has $\text{per}(J_n) = n!/n^n$ and this is the basis for the famous conjecture stated by van der Waerden in 1926.

CONJECTURE. *For every doubly stochastic $n \times n$ matrix, per$(A) \geq n!/n^n$.*

This conjecture attracted many great mathematicians and the efforts toward its solution constitute another fascinating mathematical story that found its crownimg conclusion in 1980 when the Soviet mathematicians Egorychev and Falikman independently gave a proof. The consequences from this theorem for counting results are immense. Let us see what consequences this result has for the enumeration of 1-factors.

For a matrix $M \in \mathcal{M}(n,k)$ we have already established that $\mathrm{per}(M) = k^n \mathrm{per}(A)$ where A is a doubly stochastic matrix. Thus it follows from the theorem of Egorychev and Falikman that $\mathrm{per}(M) \geq k^n n!/n^n$ and this is substantially better than the bound $k!$ we found in 6.7, for fixed k and $n \to \infty$. Summarizing, we have the following theorem.

6.11 THEOREM. *A graph $G \in \mathcal{G}(n,k)$ has at least $k^n n!/n^n$ different 1-factors, and*

$$(k!)^n \left(\frac{n!}{n^n}\right)^k \leq f(G) \leq (k!)^n$$

holds for the number $f(G)$ of edge-colorings with k colors.

Of special interest is the case $n = k$, i.e., $G = K_{n,n}$ is the complete bipartite graph. We take as color set the numbers 1 to n. If we consider the $n \times n$ matrix M (consisting only of 1's) corresponding to $K_{n,n}$, then we see that an n-edge-coloring of G corresponds to a covering of M with the numbers 1 to n such that all numbers appear in each row and column and, conversely, these coverings give precisely the edge-colorings of $K_{n,n}$ with n colors. These matrices whose investigation goes back to Euler, are called Latin squares. Therefore: **A Latin square** of order n is an $n \times n$ matrix $A = (a_{ij})$ all of whose rows and columns are permutations of $\{1, \ldots, n\}$. Figure 6.8 shows Latin squares of order 4 and 6.

Latin squares are intimately connected with important structures such as, e.g., finite projective planes. They are an indispensable tool in the theory of combinatorial designs. A first hint of their significance is obtained by noticing that the multiplication table of each finite algebraic system with the cancellation rules $ax = ay \Rightarrow x = y$, and $xa = ya \Rightarrow x = y$, is a Latin square.

$$\begin{matrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{matrix} \qquad \begin{matrix} 4 & 1 & 3 & 5 & 2 & 6 \\ 3 & 2 & 5 & 6 & 1 & 4 \\ 1 & 4 & 2 & 3 & 6 & 5 \\ 2 & 6 & 1 & 4 & 5 & 3 \\ 6 & 5 & 4 & 2 & 3 & 1 \\ 5 & 6 & 3 & 1 & 4 & 2 \end{matrix}$$

Figure 6.8

Euler in 1779 was apparently the first to discuss the counting of these "magic" squares, as he called them. He called a Latin square **reduced** if the numbers of the first row and first column appear in their natural order. The first square in Figure 6.8 is reduced, the second is not. If we let $L(n)$ denote the number of all Latin squares and $R(n)$ the number of reduced ones, then clearly $L(n) = n!(n-1)!R(n)$. Euler found the values $R(2) = R(3) = 1$, $R(4) = 4$, $R(5) = 56$, but could not compute $R(6)$. The correct value $R(6) = 9408$ was found in 1890 by Frolov who in turn calculated the wrong value 221276160 for $R(7)$. Sade, who wrote the most important papers in the beginnings of this topic, gave the correct value $R(7) = 16942080$ in 1948. Today we know the exact value up to $n = 10$. It is obvious that the numbers $L(n)$ and $R(n)$ grow enormously fast, so that at best the asymptotic order of magnitude is to be expected.

First, we have the trivial upper bound $L(n) \leq (n!)^n$. A lower bound is yielded by 6.7: $L(n) \geq n!(n-1)!\cdots 2!1!$, and the situation remained there—apart from some small improvements—up to the proof of the permanent inequality which at one blow yields

$$L(n) \geq \frac{(n!)^{2n}}{n^{n^2}}.$$

6.12 THEOREM. *For the number $L(n)$ of Latin squares,*

$$\frac{(n!)^{2n}}{n^{n^2}} \leq L(n) \leq (n!)^n.$$

The upper bound goes to ∞ at a higher order than the lower bound, it is however conjectured that the lower bound yields the correct asymptotic relation. This guess is strenghtened by the following heuristic reasoning: What is the probability that we obtain a Latin square by a random covering with the numbers 1 to n? Altogether there are n^{n^2} coverings. The probability that in a covering all rows are permutations is $(n!)^n/n^{n^2}$. Analogously, the probability that all columns are permutations is likewise $(n!)^n/n^{n^2}$, and under the (admittedly rash) assumption that these two events are independent, we obtain for $L(n)$ the value

$$n^{n^2}\left(\frac{(n!)^n}{n^{n^2}}\right)^2 = \frac{(n!)^{2n}}{n^{n^2}},$$

which is just our lower bound.

After this excursion into combinatorics we will turn again to our main theme, the factorization of graphs and pose the general problem: When does an arbitrary graph G have a 1-factor? We know that G must contain an even number $p = 2n$ of vertices. It is also easy to give sufficient conditions. An example is the following: If the simple graph G has an even number $p = 2n$ of vertices and $d(u) + d(v) \geq p - 1$ holds for any two vertices of G, then G has a 1-factor. (Proof?) The following statement completely solves the problem of the existence of a 1-factor.

6.13 THEOREM (Tutte). *A graph $G(V,E)$ has a 1-factor if and only if $c_u(G-A) \le |A|$ is satisfied for all $A \subseteq V$, where $c_u(G-A)$ denotes the number of odd components (i.e., with an odd number of vertices) of the subgraph $G-A$.*

Proof. Again the necessity of the condition can be seen immediately. Namely, let G have a 1-factor M. In $G-A$, let $C_1, \dots, C_m$ be the odd and $D_1, \dots, D_n$ the even components. Since $|C_i|$ is odd, from each C_i there must emanate at least one edge $e_i \in M$ to A, say to $a_i \in A$. (See Figure 6.9.) But since all vertices a_i must be distinct, we deduce that $|A| \ge m = c_u(G-A)$.

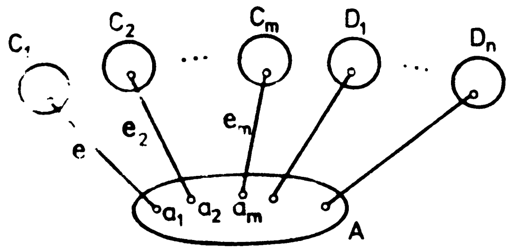


Figure 6.9

Conversely, we now assume the hypothesis of the theorem and use induction on the number p of vertices to prove the existence of a 1-factor. For $p = 0$ there is nothing to show. Hence, let $p \ge 1$. Let us consider the hypothesis when $A = \emptyset$ and $A = \{v\}$. It follows from $c_u(G-\emptyset) \le |\emptyset| = 0$ that G has no odd components, in particular that p is even. But then $G-v$ has an odd number of vertices from which it follows that $1 \le c_u(G-v) \le |\{v\}| = 1$. Therefore $c_u(G-v) = |\{v\}|$. Let $A_0 \subseteq V$ be a maximal sized set which satisfies $c_u(G-A_0) = |A_0|$. Since equality, as we have just seen, is satisfied for all one-element sets, A_0 is not empty. As in Figure 6.9, let $C_1, \dots, C_m$ be the odd and $D_1, \dots, D_n$ the even components, thus $m = |A_0|$. The existence of a 1-factor follows immediately from the following three observations.

a. Every component D_j has a 1-factor. Namely, if $S \subseteq V(D_j)$, then clearly $c_u(G-(A_0 \bigcup S)) = c_u(G-A_0) + c_u(D_j - S) \le |A_0 \bigcup S| = |A_0| + |S|$. Thus $c_u(D_j - S) \le |S|$, since $c_u(G-A_0) = |A_0|$. But A_0 is nonempty, so our claim follows by induction.

b. If $v \in C_i$, then $C_i - v$ has a 1-factor. Let us assume the contrary, then by the induction hypothesis there is an $S \subseteq V(C_i) - v$ with $c_u(C_i - (v \bigcup S)) > |S|$. Since $C_i - v$ contains an even number of vertices, $c_u(C_i - (v \bigcup S)) - |S|$ must also be even. Therefore $c_u(C_i - (v \bigcup S)) \ge |S| + 2$. It now follows from this that

$$\begin{aligned}|A_0 \cup \{v\} \cup S| = |A_0| + 1 + |S| &\ge c_u(G - (A_0 \cup \{v\} \cup S)) \\ &= c_u(G-A_0) - 1 + c_u(C_i - (v \cup S)) \\ &\ge |A_0| - 1 + |S| + 2 = |A_0| + 1 + |S|.\end{aligned}$$

Thus we would also have equality for the set $A_0 \bigcup \{v\} \bigcup S$, which contradicts the maximality of A_0.

c. There are m edges e_i, $i = 1, \dots, m$, where e_i joins the component C_i with A_0, and whose endpoints in A_0 are all distinct. This claim looks like a matching problem, and hence we will prove it accordingly. We construct a bipartite graph $\widetilde{G}$ on the vertex set $C = \{C_1, \dots, C_m\}$ and A_0, where we join C_i with $a_j \in A_0$ when in G at least one edge goes from a_j to C_i. Then our claim is equivalent to the existence of a matching $\widetilde{M}$ in $\widetilde{G}$ with $|\widetilde{M}| = m$. We verify Hall's condition in 6.3. For $X \subseteq C$, $B = R(X) \subseteq A_0$ is the set of vertices of A_0 that are joined to at least one component from X. Therefore $G - B$ also contains all these components from X. But this means that $|X| \le c_u(G-B)$ and thus, because $c_u(G-B) \le |B|$, also $|X| \le |R(X)|$ for all $X \subseteq C$.

With the help of c) we now construct m edges e_i from the C_i's to A_0 and complete the matching to a 1-factor by a) and b). ■

Tutte's Theorem gives a condition when the matching number $m(G)$ of a graph G with p vertices attains the maximal possible value, $p/2$. Generalizing, we obtain an exact formula for $m(G)$ in the following result.

6.14 THEOREM. *For the matching number $m(G)$ of an arbitrary graph G,*

$$m(G) = \frac{1}{2}\Big(p - \max_A (c_u(G-A) - |A|)\Big).$$

Proof. Let $d = p - 2m(G) \ge 0$. We construct the graph $\widetilde{G} = G + K_d$, i.e., we add to G a disjoint complete graph K_d on d vertices and join all vertices of G with all vertices of K_d. Clearly $\widetilde{G}$ then has a 1-factor and now the formula follows easily from 6.13. ■

Undoubtedly Hall's and Tutte's theorems are among the deepest combinatorial results about graphs. Moreover from Tutte's result we can immediately derive Petersen's Theorem 3.6. Let G be bridgeless and 3-regular, $A \subseteq V$, and $C_1, \dots, C_m$ the odd components of $G - A$. Since G contains no bridges, at least two edges lead from each C_i to A. But there can not be exactly two, since otherwise the sum of the vertex degrees in C_i would be $3|V(C_i)| - 2$. But that is an odd number, in contradiction to 1.2. Therefore, from the C_i's there are altogether at least $3c_u(G-A)$ edges that lead to A. On the other hand, by the 3-regularity this number is at most $3|A|$, whence $3c_u(G-A) \le 3|A|$. Thus it follows that $c_u(G-A) \le |A|$.

The problem of the *existence* of 1-factors is thus completely solved. Much more difficult is the question: When is a regular graph 1-*factorable*? (This is no wonder, since, by our remarks on edge-colorings, this would settle the 4-color problem.) Here except for the graph class K_p (that we have already discussed in Chapter 3) and the bipartite graphs, only partial results are known.

What about 2-factors? The existence of a 2-factor will be treated below as a special case of Tutte's Theorem 6.16; but we can immediately dispose of the 2-factorization problem in a very simple way. Every 2-factorable graph is obviously regular with an even degree—and this is also sufficient.

6.15 THEOREM. *Every* $2d$*-regular graph with* $d \geq 1$ *is* 2*-factorable.*

Proof. For $d = 1$, the graph consists of disjoint circuits and is itself a 2-factor. Now suppose $d > 1$; then it suffices to show by induction that $G(V, E)$ has *one* 2-factor. In Chapter 3 we have seen that every graph each of whose vertices has even degree (we have called such a graph a cycle), decomposes into edge-disjoint circuits. Therefore, let $E = E_1 \bigcup E_2 \bigcup \cdots \bigcup E_t$ be a decomposition into circuits. In every circuit we orient the edges cyclically around the circuit. Since every vertex of a circuit is exactly once an initial point and once a terminal point of an oriented edge, then each vertex is in toto an initial point exactly d times and a terminal point exactly d times. If we write (u, v) for the oriented edge, then u will be the initial point and v the terminal point of the edge. We associate a bipartite graph $\widetilde{G}(V \bigcup V', \widetilde{E})$ to G in the following way: V is the vertex set of G, V' is a copy of V whose vertices we denote by $u', v', \cdots$. Finally, let $uv' \in \widetilde{E}$ if and only if (u, v) is an oriented edge of G. Figure 6.10 shows a 4-regular graph G. The circuits of the decomposition are indicated by the same number of arrows, and $\widetilde{G}$ is the corresponding bipartite graph. $\widetilde{G}$ is d-regular by construction, and hence it has a 1-factor by 6.6. If we identify all the vertices v and v' of this 1-factor, and ignore the orientation, then each vertex in the edge set of G that we have constructed has degree 2, i.e., we obtain a 2-factor. In Figure 6.10, for example,

$$\{12', 24', 35', 46', 57', 61', 78', 83'\}$$

is a 1-factor of $\widetilde{G}$, whence we obtain the 2-factor $\{(1, 2, 4, 6), (3, 5, 7, 8)\}$ of G. ∎

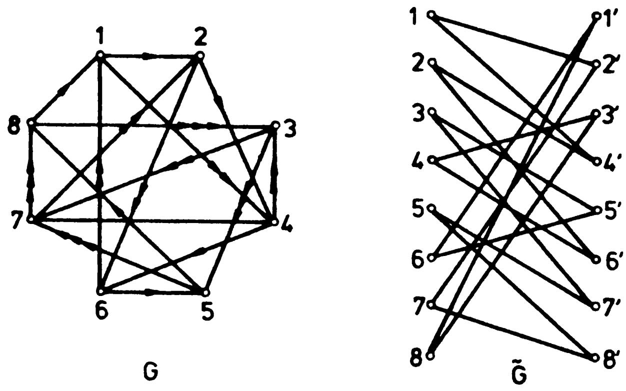


Figure 6.10

For 3-factors, and in general for r-factors with $r \geq 3$, the problem of r-factorization is much more difficult, and is unsolved. But the problem of when an r-factor *exists* was completely answered by Tutte. He even solved the following general problem: An r-factor, as we know, is nothing more than an r-regular (spanning) subgraph of G. Now let $f : V \to \mathbb{N}_0$ be an arbitrary function. Question: When does there exist a subgraph H of G with $d_H(v) = f(v)$ for all $v \in V$. We call such a subgraph an **f-factor**. We now associate with G a new graph $\widetilde{G}$ and show that G has an f-factor if and only if $\widetilde{G}$ has a 1-factor. First we subdivide each edge $e = uv$ by adding two vertices e', e'', and then extend f by setting $f(e') = f(e'') = 1$ for all of these new vertices (see Figure 6.11). Denote the graph so constructed by G'. Clearly, G has an f-factor if and only if this also holds true for G'. We now replace each vertex v in G' by $f(v)$ pairwise non-adjacent vertices $V(v)$. If $uv \in E'$, then we join all vertices in $V(u)$ with all vertices in $V(v)$. In the other case, $uv \notin E'$, we let $V(u)$ and $V(v)$ remain completely unjoined. The reader can immediately convince himself that the graph $\widetilde{G}$ obtained in this way satisfies precisely the condition given above. Accordingly, we can apply Theorem 6.13 to $\widetilde{G}$ and then carry it over to G. This last correspondence needs some care, so let us just note the result.

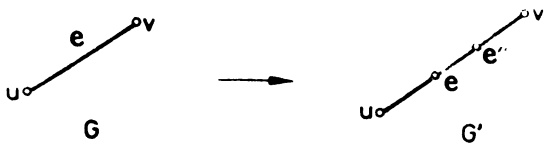


Figure 6.11

6.16 THEOREM (Tutte). *Let $G(V,E)$ and $f : V \to \mathbb{N}_0$ be given. There exists a subgraph $H(V,E')$ with $d_H(v) = f(v)$ for all v if and only if there are no disjoint vertex sets S and T*

with

$$\sum_{u \in S} f(u) < k_0(S,T) + \sum_{v \in T} (f(v) - d_{G-S}(v)),$$

where $k_0(S,T)$ *is the number of components* A *of* $G - (S \bigcup T)$ *for which the sum:* (number of edges between A and T) + $\sum_{u \in A} f(u)$ is odd.

We return once again to bipartite graphs. 6.16 provides an answer to the following question: When does a bipartite graph $G(S \bigcup T, E)$ with $|S| = m$, $|T| = n$ and prescribed vertex degrees $r_1, \ldots, r_m$ resp. $s_1, \ldots, s_n$ exist? Or, expressed in the language of matrices: When does a 0,1-matrix of the form $m \times n$ with prescribed row sums $r_1, \ldots, r_m$ and column sums $s_1, \ldots, s_n$ exist? Clearly, we must have $\sum r_i = \sum s_j$, but that is not enough as the following example shows. Let the row sums be 3,3,2,1 and the column sums be 4,4,1. Then a 4×3 matrix would have to have all 1's in the first two rows and columns, hence altogether at least 10.

We take $G = K_{m,n}$ on the vertex sets S and T as the initial graph, and construct the graph $\widetilde{G}$ as before 6.16. In this case, $\widetilde{G}$ is again bipartite, so that we can use Hall's condition 6.3 instead of 6.13 for the existence of a 1-factor in $\widetilde{G}$.

6.17 THEOREM (Gale-Ryser). *Let nonnegative integers* $r_1, \ldots, r_m$, *resp.* $s_1, \ldots, s_n$ *be given. Then there exists a* 0,1*-matrix with row sums* $r_1, \ldots, r_m$ *and column sums* $s_1, \ldots, s_n$ *if and only if*

$$\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} s_j \text{ and } \sum_{i \in I} r_i \le |I||J| + \sum_{i \notin J} s_j$$

holds for all $I \subseteq \{1, \ldots, m\}$, $J \subseteq \{1, \ldots, n\}$.

Proof. The necessity of the conditions is immediately seen. Now, conversely, suppose the conditions are satisfied. Assume the vertex set of $K_{m,n}$ is $A \bigcup B$ where $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$. As described in the construction of $\widetilde{G}$, two vertices e_{ij}, f_{ij} are added to each edge $a_i b_j$ and a_i is replaced by r_i vertices $V(a_i)$ and b_j by s_j vertices $V(b_j)$ which are joined as given there. The resulting graph $\widetilde{G}$ is bipartite with the vertex sets $\widetilde{A} = \bigcup_{i=1}^{m} V(a_i) \cup \bigcup_{i,j} f_{ij}$, $\widetilde{B} = \bigcup_{j=1}^{n} V(b_j) \cup \bigcup_{i,j} e_{ij}$. We have $|\widetilde{A}| = \sum_{i=1}^{m} r_i + mn = \sum_{j=1}^{n} s_j + mn = |\widetilde{B}|$. Now let $X \subseteq \widetilde{A}$, and let $I \subseteq \{1, \ldots, m\}$ be the set of those indices for which $|X \bigcap V(a_i)| = p_i > 0$ and $J' = \{1, \ldots, n\} - J$ be those indices for which $|X \bigcap \{f_{1j}, \ldots, f_{mj}\}| = q_j > 0$. Now we must show that $|X| \le |R(X)|$. We first have $|X| = \sum_{i \in I} p_i + \sum_{j \in J'} q_j$. Each of the p_i vertices from $V(a_i)$ is adjacent to $\{e_{i1}, \ldots, e_{in}\}$ and since $R(V(a_i)) \bigcap R(V(a_j)) = \emptyset$ for $i \neq j$, these vertices together contribute $|I|n$ to $|R(X)|$. A vertex f_{ij} is adjacent to $V(b_j)$ and e_{ij}, therefore the $\sum_{j\in J'} q_j$ vertices f_{ij} contribute together $\sum_{j\in J'} s_j + \sum_{j\in J'} q_j$. In these sums, some vertices e_{ij} that can be reached from $V(a_i)$ as well as from f_{ij} are possibly counted twice. Since this is possible only for $i \in I$ and $j \in J'$, at most $|I||J'| = |I|(n - |J|)$ vertices are counted twice. Thus we obtain

$$|R(X)| \geq |I|n + \sum_{j\notin J} s_j + \sum_{j\notin J} q_j - |I|(n - |J|) = |I||J| + \sum_{j\notin J} s_j + \sum_{j\notin J} q_j.$$

Since $p_i \leq r_i$ for all i, it follows from the hypothesis that:

$$|X| = \sum_{i\in I} p_i + \sum_{j\notin J} q_j \leq \sum_{i\in I} r_i + \sum_{j\notin J} q_j \leq |I||J| + \sum_{j\notin J} s_j + \sum_{j\notin J} q_j \leq |R(X)|. \quad \blacksquare$$

Let us simplify the conditions of 6.17. By permuting the rows and columns we can attain the situation that $r_1 \geq r_2 \geq \ldots \geq r_m$ and $s_1 \geq s_2 \geq \ldots \geq s_n$. If $|I| = k$, and $|J| = l$, then we obtain the sharpest assertion in 6.17 when the r_i are as large as possible and the s_j are as small as possible, i.e., $\sum_{i=1}^{k} r_i \leq kl + (s_{l+1} + \cdots + s_n)$. But this is equivalent to $\sum_{i=1}^{k} r_i \leq \sum_{i=1}^{k} s_i^\star$ with $s_i^\star = |\{j : s_j \geq i\}|$, i.e., as the reader can easily convince himself, $s_i^\star$ is the number of columns which sum to at least i, $i = 1, \ldots, m$. Note that

$$\sum_{i=1}^{m} s_i^\star = \sum_{i=1}^{m} \sum_{\substack{j=1 \\ s_j \geq i}}^{n} 1 = \sum_{j=1}^{n} \sum_{i=1}^{s_j} 1 = \sum_{j=1}^{n} s_j.$$

6.18 COROLLARY. *Suppose that $r_1, \ldots, r_m$ and $s_1, \ldots, s_n$ are nonnegative integers. We order each sequence by size:*

$$\bar{r}_1 \geq \bar{r}_2 \geq \cdots \geq \bar{r}_m, \text{ and } \bar{s}_1 \geq \bar{s}_2 \geq \cdots \geq \bar{s}_n.$$

Then there exists a 0,1-matrix with row sums $r_1, \ldots, r_m$ and column sums $s_1, \ldots, s_n$ if and only if $\sum_{i=1}^{k} \bar{r}_i \leq \sum_{i=1}^{k} \bar{s}_i^\star$ holds for $k = 1, \ldots, m$, with equality when $k = m$.

In our example we have $r_1 = r_2 = 3$, $r_3 = 2$, $r_4 = 1$, resp. $s_1 = s_2 = 4$, $s_3 = 1$, and hence $s_1^\star = 3$, $s_2^\star = s_3^\star = s_4^\star = 2$ with $r_1 + r_2 = 6 > 5 = s_1^\star + s_2^\star$.

A generalization of our result immediately comes to mind: When does there exist an $m \times n \times p$ cube with 0,1 entries, and prescribed sums in every direction? Here essentially nothing is known; likewise the (naturally defined) 3-dimensional matching problem (triangles instead of edges) is completely open. One can content oneself only with the insight based on experience that in combinatorics the step from 2 to 3 dimensions is usually either very easy or hopeless.

In addition to the study of 1-factorizations and the applications to tranversal theory, 0,1-matrices and the existence of graphs with prescribed degree, other "factorizations" were also investigated. For example, one can try to factor a graph into the smallest possible mumber of plane subgraphs, in order to take this minimal number as a measure of nonplanarity. In general, one wishes to decompose an arbitrary graph most economically into graphs from a prescribed class. An interesting class are the trees, or more generally, the forests. A graph is called a **forest** if each of its components is a tree (or equivalently if G has no circuits). By the **arboricity** $a(G)$ we understand the minimal number of edge-disjoint forests into which G can be decomposed. Naturally, G must be assumed to be without loops. Clearly, such a decomposition always exists, since we may take every individual edge together with the other isolated vertices as a forest. Therefore, in particular $a(G) \leq |E|$.

As an example, Figure 6.12 shows a decomposition of K_5 into 3 forests. We can easily derive a lower bound for $a(G)$. It is first of all clear that $a(G) \geq a(H)$ if H is a subgraph of G. If we assume that W is a forest on l vertices with h components, then because of 1.3, W has exactly $l - h$ edges, and therefore at most $l - 1$ edges.

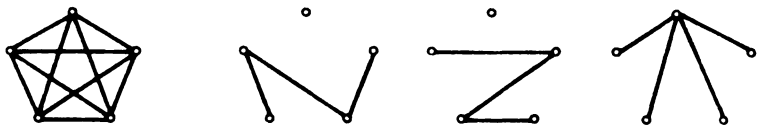


Figure 6.12

Hence if H is a subgraph on l vertices with q edges, then for the decomposition of H we need at least $\lceil q/(l-1) \rceil$ forests. If we define q_l as the maximal number of edges among all subgraphs of G with l vertices, then $a(G) \geq \max_l \lceil q_l/(l-1) \rceil$ must hold. In our example, K_5, $a(K_5) \geq \lceil 10/4 \rceil = 3$ must hold. Therefore $a(K_5) = 3$. That in fact equality always holds in $a(G) \geq \max_l \lceil q_l/(l-1) \rceil$ is one of the most remarkable general factorization formulas of graph theory. We postpone the proof until Chapter 8, where the formula will appear as a special case of a decomposition theorem for matroids.

6.19 THEOREM (Nash-Williams). *The arboricity of a graph G without loops is given by* $a(G) = \max_{l \geq 2} \lceil q_l/(l-1) \rceil$, *where* q_l *is the maximal number of edges among all subgraphs of G on* l *vertices.*

EXAMPLE. *For* K_p, $q_l = \binom{l}{2}$ *and thus* $\lceil q_l/(l-1) \rceil = \lceil l/2 \rceil$, *whence* $a(K_p) = \lceil p/2 \rceil$. *For the graphs* $K_{m,n}$ *the expression* $\lceil q_l/(l-1) \rceil$ *increases with increasing* l, *so that* $a(K_{m,n}) = \lceil mn/(m+n-1) \rceil$ *results.*

In summary, we can assert that the factorization theorems are among the most important *existence results* of graph theory, considering the variety of applications in practically all areas of discrete mathematics. Perhaps this was also a reason why the original problem: 'Is a 3-regular bridgeless plane graph 1-factorable?' was lost from sight. Here, no advances were achieved. And thus we will turn to the next theme.

EXERCISES FOR CHAPTER 6

1. Show that K_4 has a unique 1-factorization. How many 1-factorizations does K_6 have?

2°. Calculate the number of 1-factors of K_{2n}.

3. Computer problem. Design an algorithm that constructs a transversal on a set system $(N; \mathcal{A})$, resp. yields a subsystem that violates Hall's condition 6.5.

4°. Show: A bipartite graph with maximal degree Δ is the union of Δ matchings, or in other words: it has chromatic index $\chi' = \Delta$.

5. Let $(N; A_1, \dots, A_m)$ be a set system. Show: There is a partial transversal of size t (i.e., a transversal of some t sets A_j) if and only if $|\bigcup_{i \in I} A_i| \geq |I| + t - m$ holds for all $I \subseteq \{1, \dots, m\}$.

6. Let $N = \{1, \dots, n\}$. How many distinct transversals does the system $\{\{1,2\},\{2,3\},\{3,4\},\dots,\{n,1\}\}$ have?

7°. A square matrix with 0,1-entries is called a permutation matrix if each row and each column contains exactly one 1. Show: Every doubly stochastic matrix A is a convex combination of permutation matrices P_i, i.e., A can be represented in the form $A = \sum_{i=1}^{t} c_i P_i$ with $\sum_{i=1}^{t} c_i = 1$, $c_i \in \mathbb{R}$, $c_i \geq 0$.

8. Let A be an $n \times n$ matrix with row and column sums equal to k. Show that $A = \sum_{i=1}^{k} P_i$, where the P_i are permutation matrices.

9. By a Latin rectangle M of the form $(r, s; n)$, we understand an $r \times s$ matrix over $\{1, \dots, n\}$ in which each element occurs at most once in each row and column. Show that a Latin rectangle of the form $(m, n; n)$ with $m < n$ can always be completed to a Latin square of order n.

10. Complete to a Latin square.

$$\begin{array}{ccccc} 2 & 1 & 5 & 3 & 4 \\ 5 & 2 & 4 & 1 & 3 \end{array}$$

11°. Prove that an $(r, s; n)$ Latin rectangle M can be completed to a Latin square of order n if and only if each element appears at least $r + s - n$ times in M.

12. Let $G(S \bigcup T, E)$ be a simple bipartite graph. Suppose that $a_{ij} \neq 0$ whenever $s_i t_j \in E$, and that the a_{ij}'s are algebraically independent transcendental variables over Q. If $s_i t_j \notin E$ then we set $a_{ij} = 0$. Show that G has a 1-factor if and only if $\det(a_{ij}) \neq 0$ and deduce Hall's Theorem from this.

13°. Show: If the simple graph G has an even number $p = 2n$ of vertices, and $d(u) + d(v) \geq p - 1$ for any two vertices u, v, then G has a 1-factor.

14. Describe the maximal graphs with $p = 2n$ vertices that contain no 1-factors.

15. Show that for each n, K_{6n-2} has a 3-factorization.

16*. Prove 6.16.

17. Let $0 < d_1 \leq d_2 \leq \cdots \leq d_p$ be natural numbers. Show that (d_i) is the degree sequence of a tree if and only if $d_1 + \cdots + d_p = 2p - 2$.

18°. Show: $(d_1 \leq d_2 \leq \cdots \leq d_p)$ is the degree sequence of a simple graph if and only if this also holds for the sequence $(d_i')_1^{p-1}$ where

$$d_i' = \begin{cases} d_i & i = 1, \ldots, p - d_p - 1, \\ d_i - 1 & i = p - d_p, \ldots, p - 1. \end{cases}$$

19. Which sequences are graphical: 1,2,2,4,4,5; 2,2,2,3,3,4,4; 1,2,2,3,4,5,6,7,8?

20°. Prove: The arboricity $a(G)$ of every plane graph G satisfies $a(G) \leq 3$.

7. HAMILTONIAN CIRCUITS

If we turn back the pages to the end of Chapter 3, we read that the starting point for the study of Hamiltonian circuits was Tait's conjecture, 3.9: Every 3-regular polytopal graph has a Hamiltonian circuit. It was also mentioned there that in the light of Steinitz's Theorem, 3.11, Whitney showed that Tait's conjecture would even imply the 4-CC. We will begin with this result of Whitney.

His first observation (which was also already familiar to earlier authors) was that we can further restrict the class of normal cubic maps. Namely, to those maps in which any two distinct countries have at most *one* common boundary—we say that such a map has no **multiple neighborhoods**. For, if the countries F and G have common boundaries a and b, then we replace one of them, say a, by a new country A as shown in Figure 7.1. If the new map (which is again cubic and normal) is 4-colorable, then surely so is the old one.

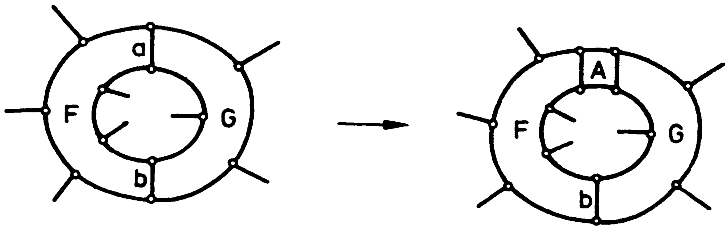


Figure 7.1

7.1 THEOREM. *The skeleton $G(\mathcal{L})$ of a normal cubic map with at least four countries and without multiple neighborhoods, is 3-connected.*

Proof. We first note that $G = G(\mathcal{L})$ has at least 4 vertices since for two vertices the map would be induced by 3 multiple edges, and would only have 3 countries, whereas a map with 3 vertices is impossible by 1.2. Clearly G is 2-connected, since $d(v) = 3$ implies that a bridge would have to occur at a cut vertex v. Now let $\{u, v\}$ be a separating vertex set. We first assume that u and v are not adjacent. Suppose the neighbors of u and v are u_1, u_2, u_3, resp. v_1, v_2, v_3 (which need not be distinct). By Menger's Theorem, 4.4, two of the neighbors of u, say u_1 and u_2, lie in different components G_1 and G_2 of $G - \{u, v\}$. Let F be the country (unique by assumption) with both $e_1 = uu_1$ and $e_2 = uu_2$ as boundaries. Then F must also contain v as a boundary vertex, and we choose the numbering so that v_1 lies in G_1 and v_2 in

G_2 (Figure 7.2). Without loss of generality, suppose the vertex u_3 is not in G_2. The country F' incident with $e_3 = uu_3$ and $e_2 = uu_2$ must then also contain v as boundary vertex, and thus have f_1 or f_2 as boundary (see Figure 7.2). But then F and F' have the boundaries e_2 and f_1 or f_2 in common, in contradiction to the hypothesis.

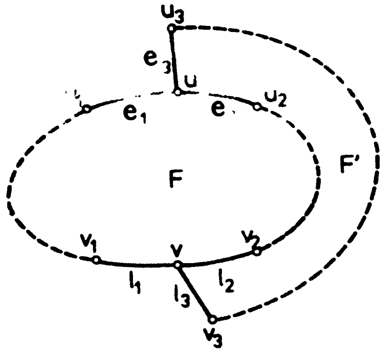


Figure 7.2

Now we assume that u and v are joined by an edge. As above, let u_1 and u_2 be adjacent to u in distinct components G_1 resp. G_2 of $G - \{u, v\}$, and analogously for v_1 and v_2. If $u_1 = v_1$, then the third edge, besides u_1u and u_1v, that shares u_1 must be a bridge. But if $u_1 \neq v_1$, then u_1 and v form a nonadjacent separating vertex set, and we are again back to the former case. ∎

Therefore, on the basis of 3.11, Tait's conjecture is in fact stronger than the 4-CC. But it is false, as we saw from Tutte's example, Figure 3.21. Not every 3-connected 3-regular plane graph G has a Hamiltonian circuit. Now Whitney in 1931 pointed out the remarkable fact that under additional assumptions, if not G then in any case the **dual graph** $G^\star$ is Hamiltonian. Or in other words: We can make a trip through $\mathcal{L}$ in which we visit the interior of each country exactly once, and then return to the starting point. The reader can easily find such a path in Figure 3.21 or on the dodecahedron in Figure 3.16.

Before we sketch the proof of this fundamental result and its extension by Tutte, we change from maps to graphs as in the preceding chapters. The dual graph $G^\star$ does not have any loops, bridges or multiple edges (corresponding to multiple neighborhoods). Moreover each country is bounded by exactly 3 edges. Stated briefly: $G^\star$ is a triangulation of the plane, more precisely, a **simple triangulation**.

It is easy to describe these graphs further. A graph is called **maximal plane** if G is simple and it is impossible to add any additional edges between nonadjacent vertices without destroying planarity. Obviously, every simple triangulation is maximal plane. Conversely, let G be a simple plane graph and let F be a country with the boundary vertices $v_1, v_2, \dots, v_t$, $t \geq 3$. If $t \geq 5$, then there must be a pair v_i, v_j of nonadjacent vertices since K_5 is not planar. But also for $t = 4$, the edges v_1v_3 and v_2v_4 can not both exist outside F. Hence, for $t \geq 4$ there is always a pair of unjoined vertices v_i and v_j, so that we can insert v_iv_j into the interior of F. Accordingly, a maximal plane graph must be a simple triangulation. Our initial remarks in the dual form assume the following self-evident formulation: *To prove the 4-CC it suffices to verify that all maximal plane graphs are 4-colorable.*

The most important properties of a maximal plane graph are summarized in the next theorem.

7.2 THEOREM. *Let G be a maximal plane graph with $p \geq 4$ vertices and q edges. Then the following hold:*

i) *G is a triangulation without multiple neighborhoods.*

ii) $d(v) \geq 3$ *for all* $v \in V$.

iii) $\sum_{v \in V} (6 - d(v)) = 12$.

iv) $q = 3p - 6$.

v) *If T is a minimal separating vertex set, then the subgraph spanned by T is a circuit.*

vi) *G is 3-connected.*

Proof. We have already seen that G is a triangulation. If two such triangular countries had two common boundaries, uv and uw, then the vertices v and w would be joined by two edges (since $p \geq 4$). ii) follows directly from the maximal plane property. iii) is the dual statement to 1.6, from which iv) immediately follows, since $\sum_{v \in V} d(v) = 2q$. Let T be an arbitrary separating vertex set. We claim the subgraph $G(T)$ spanned by T must contain a circuit. Let v_1 and v_2 be vertices in different components of $G - T$. If $G(T)$ contains no circuits, then there is a Jordan curve C from v_1 to v_2 that intersects neither the vertices nor the edges of $G(T)$. Since G is a triangulation, one infers from this (proof!), that C can be deformed to an edge trail in $G - T$. That, however, contradicts the choice of v_1 and v_2 (see Figure 7.3). Now v) follows from this, and since G contains no multiple edges, so does vi). ■

In Figure 7.3, the edges $G(T)$ are printed bold, the trail from v_1 to v_2 belonging to the Jordan curve C is drawn with dashes.

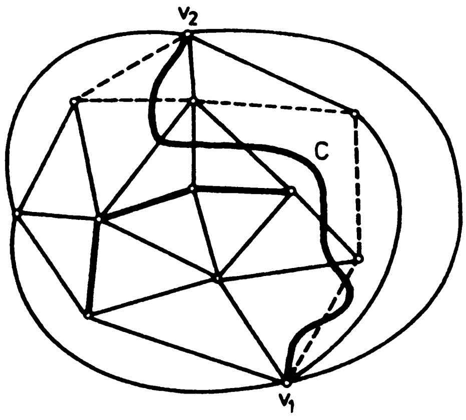


Figure 7.3

Back to Hamiltonian circuits. The tour, mentioned after 7.1, of the countries of a normal cubic map when dualized is a Hamiltonian circuit in a maximal plane graph. A 3-connected maximal plane graph G need not have a Hamiltonian circuit—the smallest example for this is depicted in Figure 7.4 (a separating circuit of length 3 is printed in heavy type). However, if we assume that G is 4-connected, then there must always be such a circuit, and this is Whitney's result.

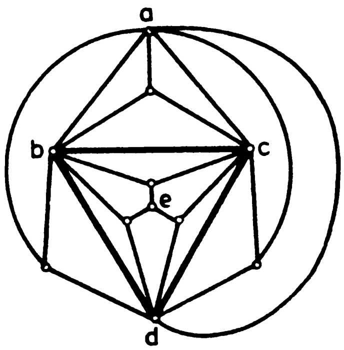


Figure 7.4

7.3 THEOREM (Whitney). *A maximal plane graph that is 4-connected (or, equivalently, which has no separating circuit of length 3) contains a Hamiltonian circuit.*

7.4 COROLLARY. *To prove the 4-CC, it suffices to verify that all plane Hamiltonian graphs are 4-colorable.*

To prove the corollary we assume that G_0 is a simple 5-chromatic plane graph with a minimal number of vertices. By adding edges we obtain a maximal plane graph G with the same number of vertices. If G had a separating circuit $C = (v_1, v_2, v_3)$, then because of the minimality of G we could color the subgraph in the interior of C together with $\{v_1, v_2, v_3\}$ with 4 colors, and analogously the exterior graph (together with $\{v_1, v_2, v_3\}$). By a permutation of the colors on C, we can attain the situation that the colorings coincide on C, from which $\chi(G) \leq 4$ would result. On the other hand, if G contains no separating 3-circuit, then by 7.3, G is Hamiltonian and the proof is finished.

In 1956 Tutte extended Whitney's result to arbitrary 4-connected plane graphs. We shall briefly sketch his proof. The fundamental concept is that of a C-bridge. Let G be a 2-connected plane graph. We identify each subgraph H with its edge set and say that v is an **on-vertex** of H if v is incident with both an edge of H and an edge from $G-H$. Now let C be a circuit. By a **C-bridge**, B, we mean a subgraph of $G-C$ with at least one edge, all of whose on-vertices lie on C, and which is minimal with respect to this property. Clearly, each edge in $G-C$ that joins two vertices of C is itself a bridge. In Figure 7.5, the C-bridges are B_1, B_2, B_3. It follows

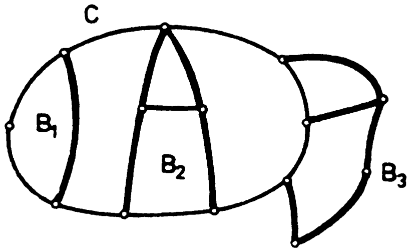


Figure 7.5

from the definition that a C-bridge containing a vertex v that does not lie on C must contain all edges incident with v. From this one sees immediately that the C-bridges are exactly the equivalence classes of the following relation defined on the edges of $G-C$: $e \sim e'$ if and only if there is a path that joins e and e' and has at most the end points in common with C. It is clear that C is a Hamiltonian circuit if and only if all C-bridges consist of individual edges. Now the decisive lemma in Tutte's proof reads as follows: Let G be a 2-connected plane graph and let C_0, C_1 be the boundary circuits of two neighboring countries F_0 and F_1 that are adjacent along the edge e. Let e_1 be another edge of C_1. Then there is a circuit J through e and e_1 such that all J-bridges have at most 3 on-vertices and all those J-bridges that contain edges of C_0 or C_1 have two on-vertices. As one can imagine, the proof of the lemma is complicated so that the reader is referred to Tutte's original paper or to the version in Ore's book. Here we will only briefly explain how the general result 7.5 follows from the lemma.

Suppose G is 4-connected and let e and e_1 be adjacent edges on C_1. Without loss of generality we can assume that F_1 lies in the interior of J and hence F_0 in the exterior. We have to show that J contains all vertices of G. Suppose the situation is as in Figure 7.6. By Menger's Theorem, 4.4, a vertex x, in the interior of J, is joined with b by 4 vertex-disjoint paths.

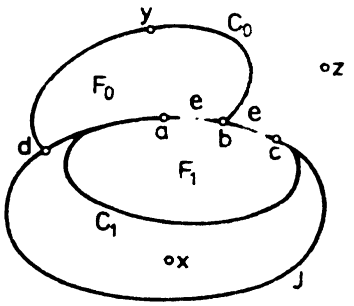


Figure 7.6

All these paths lie on one and the same bridge, which hence must have at least 4 on-vertices, which by the lemma is not possible. Let y be on C_0 and outside J. The 4 vertex-disjoint paths that join y with c lie on the same bridge B. B contains edges from C_0, but at least 3 on-vertices. Therefore, we can assume that C_0 has no vertices outside J. For an arbitrary vertex z outside J, there are 4 disjoint paths to a, and we can repeat the preceding argument for x and b.

7.5 THEOREM (Tutte). *Every 4-connected plane graph has a Hamiltonian circuit.*

Without doubt, this result of Tutte is one of the deepest theorems of graph theory. An indication of this is the fact that 4-connected plane graphs of arbitrary size have been constructed that do not contain two edge-disjoint Hamiltonian circuits. Furthermore: As the example $K_{4,5}$ shows, 7.5 becomes false if the planarity of the graph is omitted. A related problem, that remains open, is due to Barnette: Is every 3-regular 3-connected plane bipartite graph Hamiltonian?

Before we go on to arbitrary graphs, let us mention two other results that belong to this topic. One of them, proven in 1891 by Eberhard, goes back to the origins of our subject. Let us consider formula 7.2(iii). If we write p_i for the number of vertices of degree i, then it reads: $3p_3 + 2p_4 + p_5 - p_7 - 2p_8 - \cdots = 12$. One immediately notes that p_6 does not occur at all on the left-hand side. Question: Assume $(p_3, p_4, p_5, p_7, p_8, \ldots)$ is a sequence of integers ≥ 0 that satisfies the formula. Does there always exist at least one p_6 such that $(p_3, p_4, p_5, p_6, p_7, \ldots)$ is the degree sequence of a maximal plane graph? The answer (which is not easy) is yes, and that is Eberhard's Theorem. Naturally, the smallest possible value of p_6 is of particular interest, and it is conjectured that a realization with $p_6 < \max(k : p_k \neq 0)$ is possible. The reader can already see from the example $p_3 = 3$, $p_4 = p_5 = 1$, $p_k = 0$ for $k \geq 7$, that this minimum $(p_6 = 3)$ is not easy to find.

The second result again concerns Tait's conjecture about polytopal graphs. Starting from Tutte's counterexample in Figure 3.21, many subsequent authors were occupied with finding smaller examples of non-Hamiltonian polytopal graphs. Grinberg in 1968 gave a simple necessary condition for the existence of Hamiltonian circuits in a plane graph.

7.6 THEOREM (Grinberg). *Let G be a Hamiltonian plane graph without loops and with p vertices, and let C be a Hamiltonian circuit. We denote by f_i resp. g_i the number of countries with i boundary edges in the interior, resp. exterior of C. Then*

$$\sum_i (i-2)f_i = \sum_i (i-2)g_i = p-2.$$

Proof. It suffices to verify the equation for the f_i's. Let m be the number of edges lying in the interior of C. Since C is a Hamiltonian circuit, the interior of C contains $m+1$ countries, i.e., we have $\sum_i f_i = m+1$. Each of these m edges is incident with two interior countries and each of the p edges of C is incident with one of these countries. From this it follows that $2m+p = \sum_i if_i$ and thus $p = \sum_i (if_i - 2f_i) + 2 = \sum_i (i-2)f_i + 2$. ∎

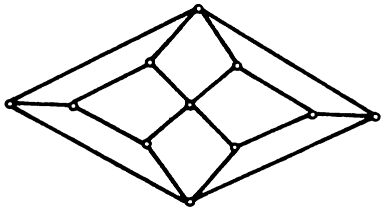


Figure 7.7

As an application we can immediately verify that the polytopal graph in Figure 7.7 has no Hamiltonian circuit. In this graph, $p = 11$, and all 9 countries have 4 boundaries. Therefore $f_4 + g_4 = 9$ and $f_i = g_i = 0$ for $i \neq 4$, whence the contradiction $2f_4 = 9$ results. The graph in Figure 7.7 is called the Herschel graph. It is the smallest example of a non-Hamiltonian polytopal graph. To be sure, it is not 3-regular. The smallest known examples of such

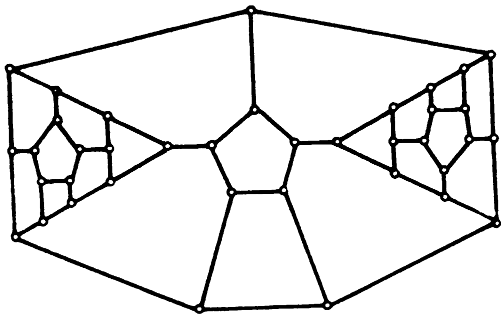


Figure 7.8

3-regular graphs have 38 vertices, as was mentioned in Chapter 3. One of them is shown in Figure 7.8. Here, the impossibility of a Hamiltonian circuit is also shown by means of the Grinberg formula. Conversely, it is known that every 3-regular polytopal graph with at most 34 vertices is Hamiltonian—the case $p = 36$ is not resolved.

Let us now pose the general question: When is a graph Hamiltonian? To spill the beans: A handy criterion for this is not known, and the characterization of Hamiltonian graphs is one of the biggest open problems of graph theory. For the reader who knows the concept of the complexity of algorithms, we add that the decision whether or not a graph is Hamiltonian is an **NP-complete problem**. Indeed, it still remains NP-complete if we restrict ourselves to plane 3-connected graphs. Naturally, this does not exclude finding a criterion. But any such criterion will be difficult to describe.

The general results concerning Hamiltonian circuits fall essentially into two classes, corresponding to necessary, resp. sufficient conditions. We begin with the *necessary* conditions. One fact is trivial: A Hamilton graph must be 2-connected. A generalization of this is:

7.7 THEOREM. *Let $G(V, E)$ be a Hamiltonian graph and $A \subseteq V$, $A \neq \emptyset$. Then $c(G - A) \leq |A|$ where $c(G - A)$ denotes the number of components of $G - A$.*

Proof. If C is a Hamiltonian circuit in G, then $c(C - A) \leq |A|$. Since $C - A$ is a spanning subgraph of $G - A$, it follows that $c(G - A) \leq c(C - A) \leq |A|$. ∎

Using this theorem, we can immediately verify that the graph in Figure 7.4 is not Hamiltonian. Namely, if we remove the vertices a, b, c, d, e, then the remaining graph has 6 components.

The condition of the theorem has a great similarity with Tutte's criterion, 6.13, for the existence of a 1-factor—a further reference to the intimate connection between factorization and Hamiltonian circuits. The introduction of the parameter $\tau(G) = \min(|A|/c(G - A))$ is suggested by 7.7, where the minimum ranges over all $A \subseteq V$ with $c(G - A) > 1$. In this connection, G is assumed not to be complete—complete graphs are always Hamiltonian. Now 7.7 states: If G is Hamiltonian, then $\tau(G) \geq 1$. Chvátal, who suggested the parameter $\tau(G)$, conjectures that there is a number t_0 such that $\tau(G_0) \geq t_0$ implies that G is Hamiltonian. $t_0 = 1$ would be the converse of 7.7, however the Petersen graph, P, is a counterexample ($\tau(P) = 4/3$). $\tau(G)$ is closely connected with two other graph parameters, the connectivity number $\kappa(G)$ and the independence number $\alpha(G)$ (=maximal number of pairwise nonadjacent vertices). If A is a separating vertex set (hence $c(G - A) > 1$), then we know from 4.4 that we must have $|A| \geq \kappa$. Since obviously $c(G - A) \leq \alpha$, we immediately obtain:

7.8 LEMMA. *For every non-complete graph G, $\tau(G) \geq \kappa(G)/\alpha(G)$.*

More is known about *sufficient* conditions. All of them are with respect to certain parameters of the graph, e.g., in the form: If the degrees of the individual vertices are large enough, then there is a Hamiltonian circuit. Let us note first a result that is complementary to 7.8.

7.9 THEOREM (Erdős-Chvátal). *Let G be a graph with at least 3 vertices, connectivity number κ, and independence number α. If $\kappa \geq \alpha$, then G is Hamiltonian.*

Proof. Assume G has connectivity number κ and has no Hamiltonian circuit. We will show that $\alpha \geq \kappa + 1$. For $\kappa = 0$ or 1 this is trivial. Therefore let $\kappa \geq 2$; then G has a circuit. Let C be a circuit of maximal length. Let G_1 be a component of $G - C$ and let $v_1, \ldots, v_l$ be the vertices on C which are adjacent to vertices of G_1. No pair of vertices v_i, v_j is adjacent on C since otherwise we could construct along G_1 a circuit whose length is greater than that of C. Therefore $\{v_1, \ldots, v_l\}$ is a separating vertex set, and thus $l \geq \kappa$. We choose a sense of traversal on C and denote by $w_1, \ldots, w_l$ those vertices on C that come directly after each of the v_i's.

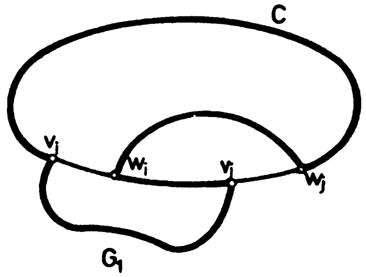


Figure 7.9

CLAIM: $\{w_1, \ldots, w_l\}$ *is an independent set in G.* Assume w_i and w_j are joined. We strike out the edges v_iw_i and v_jw_j and add w_iw_j and a v_i,v_j-path in G_1. Then the resulting circuit has greater length than C—a contradiction. (See Figure 7.9.) Finally, by the maximality of C, no w_i is adjacent to any vertex of G_1. Now if w_0 is a vertex of G_1, then accordingly $\{w_0, w_1, \ldots, w_l\}$ forms an independent set, and we have $\alpha \geq l + 1 \geq \kappa + 1$. ∎

7.9 can not be improved further, as we again learn from the Petersen graph, for which $\kappa = 3$ and $\alpha = 4$. If one combines 7.7 and 7.9, then one sees that $\kappa(G)/\alpha(G) < 1 \leq \tau(G)$ are the conditions under which no claim can be made about the existence or nonexistence of Hamiltonian circuits.

The most interesting *sufficient* conditions are obtained from the consideration of the vertex degrees. The first two results in this direction originated with Ore (1960) and Dirac (1952). They form the first and second part of the following theorem. First a definition: A **Hamiltonian path** is a path that contains all vertices exactly once. If the end points of the path are also joined then we obtain a Hamiltonian circuit.

7.10 THEOREM (Ore-Dirac). *Let G be a simple graph with $p \geq 3$ vertices. If $d(u)+d(v) \geq p$ holds for each pair of nonadjacent vertices, then G is Hamiltonian. In particular, this happens if $d(u) \geq p/2$ for all vertices u.*

Proof. Assume the theorem is false. From among all counterexamples with p vertices, we choose a graph $G(V,E)$ with the maximal number of edges. Let u and v be vertices that are not joined in G; then $G \bigcup uv$ is Hamiltonian. Therefore, there is a Hamiltonian path $(u = u_1, u_2, \ldots, u_p = v)$ in G. Let $A = \{u_i : uu_{i+1} \in E\}$ and $B = \{u_i : u_i v \in E\}$. By hypothesis, $|A| + |B| = d(u) + d(v) \geq p$. Since $v \notin A \bigcup B$, there must be a vertex u_k with $u_k \in A \bigcap B$. But this means that $u_k v \in E$ and $uu_{k+1} \in E$, so that we obtain the Hamiltonian circuit $(u = u_1, \ldots, u_k, v = u_p, u_{p-1}, \ldots, u_{k+1}, u)$. Hence G was Hamiltonian after all (Figure 7.10). ∎

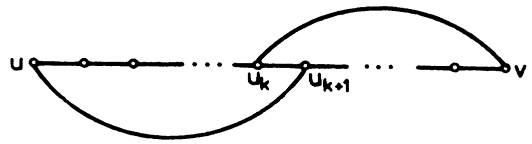


Figure 7.10

If we examine the proof, then we see that the following lemma is valid: Let G be a simple graph with p vertices and having u and v as nonadjacent vertices with $d(u) + d(v) \geq p$. Then G is Hamiltonian if and only if $G \bigcup uv$ is Hamiltonian. But this means: If G has a pair of such nonadjacent vertices u and v with $d(u)+d(v) \geq p$, then we can add the edge uv without changing the Hamiltonicity of G. If there again exists in the new graph $G_1 = G \bigcup uv$ such a pair, then we can again add the corresponding edge, etc. The process stops if we have finally constructed a graph H with the property: $d_H(u) + d_H(v) < p$ for each pair of vertices u and v that are nonadjacent in H. The reader can easily convince himself that it is immaterial in which order the edges are added—we always obtain the same final graph H. Thus the following definition is not ambiguous.

DEFINITION. *Let G be a simple graph on p vertices. The (Hamiltonian)* **closure**, *$H = h(G)$, of G is the smallest graph containing G for which $d_H(u) + d_H(v) < p$ holds for all vertices u and v that are not adjacent in H, where $d_H(v)$ denotes the degree of v in H.*

Figure 7.11 shows the construction of a closure. The added edges are in bold type.

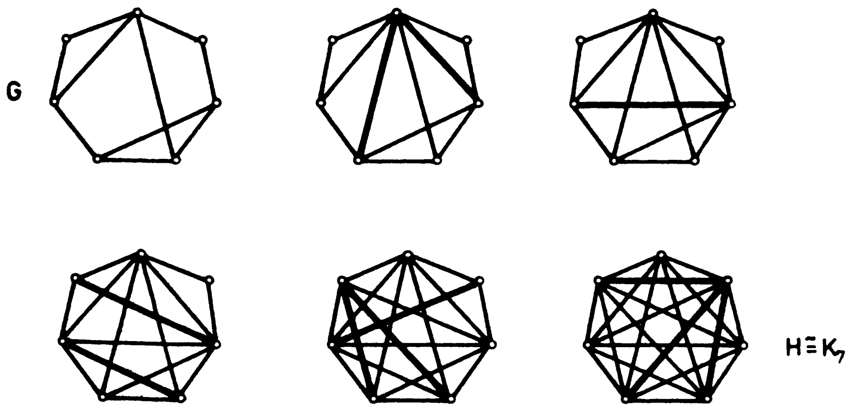


Figure 7.11

We immediately obtain the following basic theorem from our observations.

7.11 THEOREM (Bondy-Chvátal). *A simple graph G is Hamiltonian if and only if its closure $h(G)$ is Hamiltonian.*

If we can infer from the graph G that $h(G)$ is Hamiltonian, then G must also be Hamiltonian. In particular the following corollary holds:

7.12 COROLLARY. *If G is a simple graph whose closure is a complete graph, then G is Hamiltonian.*

This last result gives us a method for deriving sufficient conditions from the degree sequence.

7.13 THEOREM (Chvátal). *Let G be a simple graph on p vertices. We number the vertices $v_1, \ldots, v_p$ so that $d_1 \le d_2 \le \ldots \le d_p$ holds for the degree sequence, where we abbreviate, $d_i = d(v_i)$. If $d_k \le k < p/2$ always implies that $d_{p-k} \ge p - k$, then G is Hamiltonian.*

Proof. We will prove that the hypothesis of the theorem forces the completeness of the closure $H = h(G)$. Assume this is not the case; then we could pick a pair of vertices u and v that are not joined in H for which $d_H(u) + d_H(v)$ is maximally large. Let $d_H(u) \le d_H(v)$. From the definition of H, we must have $d_H(u) + d_H(v) \le p - 1$, whence it follows that $d_H(u) < p/2$. We set $k = d_H(u)$ and show that with this index k, the condition of the theorem is violated. Let A and B be the sets of vertices in H not adjacent to u resp. v. We have $|A| = p - 1 - k$,

$|B| = p - 1 - d_H(v) \geq d_H(u) = k$. By the choice of u and v, $d_H(x) \leq k$ holds for all $x \in B$ and $d_H(x) \leq d_H(v)$ holds for all $x \in A$. From $d(x) \leq d_H(x)$ we see that there are at least k vertices in G (those from B) for which $d(x) \leq k$, i.e., $d_k \leq k < p/2$ and that there are at least $p - k$ vertices in G (those of A and u) for which $d(x) \leq d_H(v) \leq p - 1 - k$, i.e., $d_{p-k} \leq p - k - 1$ holds and this is our desired contradiction. ∎

How good is this theorem? It naturally encompasses Dirac's result (this is trivial, since $d_1 \leq p/2$) and also Ore's, as one can readily show. However it can not be applied to the graph in Figure 7.11. There we have the degree sequence $(2,2,3,3,3,3,4)$ and the condition in 7.13 is not satisfied, even for $k = 2$, since $d_5 = 3 < 5$. Therefore the statement 7.12 is stronger. But in the following sense, 7.13 is the best possible. Let $d = (d_1 \leq d_2 \leq \cdots \leq d_p)$ and $d' = (d'_1 \leq d'_2 \leq \cdots \leq d'_p)$ be degree sequences. We say that d' **dominates** d if $d_i \leq d'_i$ for all i, in symbols, $d < d'$. By '$<$', the sequences become ordered sets. It is immediately clear that if a sequence d satisfies Chvátal's condition in 7.13, then the same is true for every sequence that dominates d. In the language of order theory this means: The set of sequences that satisfy 7.13 form an upper ideal with respect to the order $<$, which by 7.13 consists solely of Hamiltonian sequences. Now this ideal can not be enlarged inside the class of Hamiltonian sequences. Namely, suppose $d = (d_1 \leq d_2 \leq \cdots \leq d_p)$ is a sequence that violates 7.13, then there is a $k < p/2$ with $d_k \leq k$ and $d_{k-p} \leq p - k - 1$. Thus the sequence d is dominated by

$$d(k,p) = (\underbrace{k,k,\ldots,k}_{k},\underbrace{p-k-1,\ldots,p-k-1}_{p-2k},\underbrace{p-1,\ldots,p-1}_{k}).$$

But there is a graph $G(k,p)$ that corresponds to this degree sequence and it is uniquely determined as the reader can easily convince himself (see Figure 7.12):

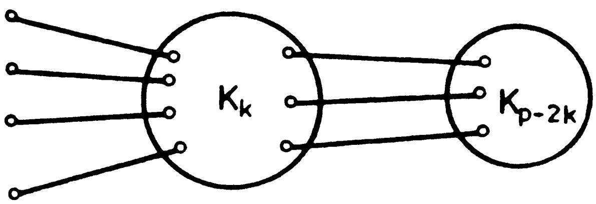


Figure 7.12

A $\qquad$ B $\qquad$ C

$|A| = k$ $\qquad$ $|B| = k$ $\qquad$ $|C| = p - 2k$

As specified, the vertex set of $G(k,p)$ splits into 3 parts, A, B, and C. The subgraphs on B and C are complete, there are no edges between the vertices of A, and any two vertices from A and B resp. B and C are joined. However, $G(k,p)$ is not Hamiltonian, as we learn from $c(G-B) = k+1 > k = |B|$ and 7.7.

One more thing follows from our observations. Let $G(V,E)$ be a non-Hamiltonian graph with degree sequence $d = (d_1 \leq \cdots \leq d_p)$, then d is dominated by a sequence $d(k,p)$ for a particular $k < p/2$. Thus, in particular, by 1.2,

$$|E| = \frac{1}{2}\sum d_i \leq \frac{1}{2}(k^2 + (p-2k)(p-k-1) + k(p-1)) = \binom{p-2}{2} + k^2.$$

Now one computes immediately that

$$\binom{p-k}{2} + k^2 \leq \binom{p-1}{2} + 1$$

always holds, with equality when $k = 1$ or $k = 2$, $p = 5$. Thus we obtain:

7.14 COROLLARY. *If a simple graph has p vertices and more than $\binom{p-1}{2} + 1$ edges, then it is Hamiltonian. The only non-Hamiltoninan graphs with $\binom{p-1}{2} + 1$ edges are the graphs in Figure* 7.13.

$G(1,p)$ $G(2,5)$

Figure 7.13

Regular graphs are naturally of particular interest. If G is k-regular with $k \geq p/2$, then it follows from 7.10 that G is Hamiltonian. The first case that is not covered by 7.10 or 7.13 is $k = (p-1)/2$, but indeed the corresponding result also holds there. The (not difficult) proof can be left to the reader.

7.15 THEOREM (Nash-Williams). *A simple k-regular graph on $p = 2k+1$ vertices is Hamiltonian.*

Can one sharpen the preceding result further? Clearly a k-regular graph on $p = 2k+2$ vertices need no longer be Hamiltonian. The graph G consisting of 2 components, each isomorphic to K_{k+1}, is a counterexample. But if we assume G to be connected, then a Hamiltonian graph must again result. Indeed, the following even holds: If the simple graph G is 2-connected and k-regular with $k \geq p/3$, then G is Hamiltonian. Again our ever present Petersen graph stands in the way of sharpening this result.

Other interesting sufficient conditions for the existence of Hamiltonian circuits arise from considering the exponentiation of graphs. The **m-th power**, G^m, of a graph G, is the graph whose vertex set corresponds to that of G and vertices are joined in G^m if and only if they are joined in G by a path of length $\leq m$. Figure 7.14 shows a tree T and its square T^2.

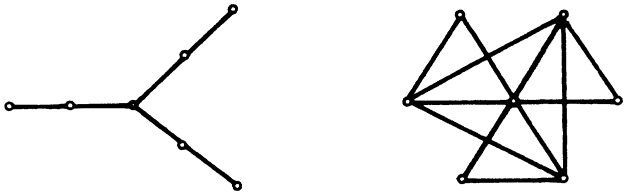


Figure 7.14

In 1960 Sekanina showed that the third power G^3 of every connected graph G is Hamiltonian. This does not hold for the square G^2, as is shown by the non-Hamiltonian graph T^2 of Figure 7.14. As a matter of fact, Harary and Schwenk have proven that the square of a tree G is Hamiltonian if and only if G does not contain the tree T in Figure 7.14 as a subgraph. But, in general, we know of no characterization of those graphs whose square is Hamiltonian. On the positive side, Fleischner proved the following theorem in 1974 which represents, together with Tutte's result 7.5, the most significant advance concerning Hamiltonian circuits. We refer the reader to the original paper for the difficult proof.

7.16 THEOREM (Fleischner). *The square G^2 of any 2-connected graph is Hamiltonian.*

Before we conclude this chapter on Hamiltonian circuits, we will go into another interesting aspect, that has its origins in 7.14. There we showed that a simple graph on p vertices and at least $\binom{p-1}{2}+2$ edges always contains a Hamiltonian circuit. Or reformulated: A simple graph on p vertices that contains *no* circuit C_p of length p, has at most $\binom{p-1}{2}+1$ edges. This suggests the following general **extremal problem**: Let H be a simple graph. We denote by $\mathrm{ex}(H,p)$ the maximal number of edges in a simple graph on p vertices that does *not* contain H as a subgraph.

We concentrate on two important classes of graphs: $H = P_k$ (path of length k) and K_k.

7.17 LEMMA. *Let G be a connected simple graph with $p \geq 3$ vertices in which $d(u)+d(v) \geq k$ holds for any two nonadjacent vertices u, v. If $k = p$, G contains a (Hamiltonian) path P_{p-1}. If $k < p$, G contains a path P_k of length k.*

Proof. The case $k = p$ was settled in 7.10. Now suppose $k < p$, and let $P = (u_0, u_1, \dots, u_l)$ be a longest path in G. All neighbors of u_0 and u_l must lie on P. Assume $l \le k - 1$; then G can not contain a circuit of length $l + 1$. For, by $l + 1 \le k \le p - 1$, there is a vertex outside of C that would have to be joined with a vertex of C (G is connected!) and we would have a path of length $l + 1$. Thus, in particular, $u_0 u_l \notin E$. As in the proof of 7.10, we set $A = \{u_i : u_0 u_{i+1} \in E\}$, $B = \{u_i : u_i u_l \in E\}$. By hypothesis, $|A| + |B| = d(u_0) + d(u_l) \ge k$, $u_l \notin A \bigcup B$ and $A \bigcap B = \emptyset$, since otherwise we would obtain the circuit $\{u_0, u_{i+1}, u_{i+2}, \dots, u_l, u_i, u_{i-1}, \dots, u_1, u_0\}$ of length $l+1$. Finally it follows from this that $k \le d(u_0) + d(u_l) \le l$. Therefore, P indeed has length at least k. ■

7.18 THEOREM. ex$(P_k, p) \le p(k-1)/2$ *holds, with equality if and only if all components are isomorphic to* K_k.

Proof. Let k be arbitrary but fixed. For $p \le k$ the assertion is trivial. We assume inductively that the result is correct for fewer than p vertices. Now let $G(V, E)$ be a graph on $p > k$ vertices, that does not have a path of length k. If G is not connected, then we can immediately apply the induction asumption to the individual components. Therefore, let G be connected. Then G contains no complete subgraph K_k and furthermore, by 7.17, a vertex u with $d(u) \le (k-1)/2$. Since by induction $G - u$ is not an extremal graph (i.e., for which equality holds in 7.18), we deduce that

$$|E| = d(u) + |E(G-u)| < \frac{k-1}{2} + \frac{k-1}{2}(p-1) = \frac{k-1}{2}p,$$

whereby both assertions are proven. ■

The exact value of ex(P_k, p) is unknown for arbitrary k and p. For the case $k = p - 1$, i.e., for the maximal possible number of edges in a simple graph without a Hamiltonian path, the formula ex$(P_{p-1}, p) = \binom{p-1}{2}$ can be verified (similarly as in 7.14) by the reader. The extremal graphs consist of one component K_{p-1} and one additional isolated vertex, and there is one other graph ($p = 4$, $k = 3$), namely $K_{1,3}$.

The most important and historically also the first extremal theorem concerns graphs that contain no complete subgraphs K_k. The corresponding formula for ex(K_k, p) was found by Turán (in a prisoner of war camp, as he himself reports) in 1941. We present a refinement of his result by Erdős who published the most important work in this area of graph theory.

7.19 THEOREM. *Let $G(V,E)$ be a simple graph that contains no complete subgraph K_k. Then there is a $(k-1)$-chromatic graph H on V such that $d_H(v) \geq d(v)$ holds for all $v \in V$.*

Proof. The claim is clear for $k = 1$ or 2. Therefore let $k \geq 3$ and suppose the theorem is true for all smaller values of k. We choose a vertex u of maximal degree in G and denote by G_0 the subgraph of G induced on the neighborhood U of u. G_0 contains no graphs K_{k-1} since otherwise we would obtain together with u a graph K_k in G. Therefore by the induction hypothesis, there is a $(k-2)$-chromatic graph H_0 on U with $d_H(v) \geq d_{G_0}(v)$ for all $v \in U$. We now construct H by adding to H_0 all edges between vertices from U and $V-U$. Clearly H is $(k-1)$-chromatic (we can color all vertices from $V-U$ alike). If $v \in U$, then $d_H(v) = d_{H_0}(v) + |V| - |U| \geq d_{G_0}(v) + |V| - |U| \geq d(v)$, and if $v \in V-U$, then by the choice of u we likewise have $d_H(v) = |U| = d(u) \geq d(v)$. ■

Now with the help of 7.19 we can quickly determine $\mathrm{ex}(K_k, p)$ and the extremal graphs. (The reader will certainly have noticed the similarity to the argument prior to 7.14.) Suppose the $(k-1)$-chromatic graph H has the color classes $A_1, \ldots, A_{k-1}$ with $|A_i| = p_i$, where we assume that $p_1 \leq p_2 \leq \cdots \leq p_{k-1}$. Then

$$|E(G)| = \frac{1}{2}\sum_{v\in V} d(v) \leq \frac{1}{2}\sum_{v\in V} d_H(v) = |E(H)| \leq \sum_{1\leq i<j\leq k-1} p_i p_j.$$

The expression $\sum_{1\leq i<j\leq k-1} p_i p_j$ is maximal (and indeed unique) if the p_i's are as equal as possible, i.e., if $p_1 = p_2 = \cdots = p_{k-1-r} = m$, and $p_{k-r} = \cdots = p_{k-1} = m+1$, where $p = m(k-1) + r$ with $0 \leq r \leq k-2$. We shall denote the corresponding $(k-1)$-chromatic graph by $T(k-1,p)$. Therefore, $T(k-1,p)$ consists of $k-1$ independent sets of size m or $m+1$ and all edges between vertices from different sets. Figure 7.15 shows the graph $T(3,7)$.

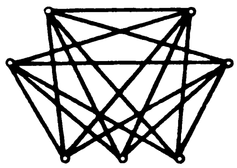


Figure 7.15

$T(k-1,p)$ contains $(1/2)[(k-1-r)m(p-m) + r(m+1)(p-m-1)]$ edges and this is equal to

$$\frac{(k-2)(p^2-r^2)}{2(k-1)} + \binom{r}{2},$$

as one verifies by an easy calculation. Accordingly, this last number is an upper bound for $\mathrm{ex}(K_k, p)$. However, since the graph $T(k-1,p)$ clearly contains no complete subgraph K_k, this number must be the exact value of $\mathrm{ex}(K_k, p)$. The extremal graphs are also determined by this. Namely, if G is an extremal graph, then we must have $d(v) = d_H(v)$ and $H = T(k-1,p)$. Hence it immediately follows by induction that G is $(k-1)$-chromatic and thus must be equal to $T(k-1,p)$. Summarizing, we obtain the following theorem:

7.20 THEOREM (Turán). *For $k \geq 2$,*

$$\mathrm{ex}(K_k, p) = \frac{(k-2)(p^2 - r^2)}{2(k-1)} + \binom{r}{2},$$

where $p = m(k-1) + r$, $0 \leq r \leq k-2$. $T(k-1,p)$ is the unique extremal graph.

For example, a simple graph without triangles on p vertices has at most $\lfloor p^2/4 \rfloor$ edges, and $K_{\lfloor p/2 \rfloor, \lceil p/2 \rceil}$ is the unique extremal graph.

Let us solve another related problem. Let $G(V, E)$ be a simple graph on p vertices and let $\overline{G}$ be its complement. We want to count the **total number**, ∇, of triangles in G and $\overline{G}$. A triple $(u, v, w) \subseteq V$ is said to be **bad**, if u, v, w, yield a triangle neither in G nor in $\overline{G}$. Then exactly 2 of the vertices u, v, w have the property that one incident edge lies in G and the other in $\overline{G}$. Since for an arbitrary vertex $u \in V$ there are clearly exactly $d(u)(p-1-d(u))$ such edge pairs incident with u, the number of bad triangles is $(1/2)\sum_{u \in V} d(u)(p-1-d(u))$ and thus the total number of triangles in G or $\overline{G}$ is

$$\nabla = \binom{p}{3} - (1/2)\sum_{u \in V} d(u)(p-1-d(u)).$$

Since $d(u)(p-1-d(u)) \leq \frac{p-1}{2}^2$, we obtain

$$\nabla \leq \binom{p}{3} - \frac{p}{2}\left(\frac{p-1}{2}\right)^2 = \frac{p(p-1)(p-5)}{24}.$$

In particular, it follows for every graph G with $p \geq 6$ vertices, that either G or $\overline{G}$ contains a triangle. For $p = 5$ this is no longer true: $G = C_5$ with $\overline{G} = C_5$ is a counterexample.

This last result is a familiar puzzle: Show that at every party with 6 people, there are 3 people who are mutual acquaintances, or three who are mutually nonacquainted. The generalization of this puzzle leads to one of the most beautiful and important theorems in all of combinatorics, the Ramsey Theorem, with which we will conclude this chapter.

7.21 THEOREM (Ramsey). *Let $k, l \geq 1$ be natural numbers. Then there is a smallest number $R(k,l)$, called* **Ramsey number**, *such that for every simple graph G on $p \geq R(k,l)$ vertices, either G contains a complete subgraph K_k or the complementary graph $\overline{G}$ contains a complete subgraph K_l.*

Proof. We use induction on $k+l$. In case $k=1$ or $l=1$, then the claim is trivial. For $k=2$ and $l \geq 2$ we have $R(2,l)=l$ since a graph G on l vertices either contains an edge, therefore a subgraph K_2, or $\overline{G}$ is isomorphic to K_l. Similarly, $R(k,2)=k$ holds for $k \geq 2$. Now we assume by induction that the theorem holds for all pairs (k',l') with $2 \leq k' \leq k-1$, $2 \leq l' \leq l$ and for all pairs (k',l') with $2 \leq k' \leq k$, $2 \leq l' \leq l-1$. We define p_1 to be $R(k-1,l)$, and p_2 to be $R(k,l-1)$ and verify that

$$(\star) \qquad R(k,l) \leq p_1 + p_2$$

holds. Therefore let $G(V,E)$ be an arbitrary graph on $p \geq p_1 + p_2$ vertices and let $v \in V$. Among the $p-1$ vertices $\neq v$ there are, since $p-1 \geq p_1+p_2-1$, at least p_1 that are joined to v or at least p_2 that are not joined to v. In the first case by the induction assumption there is a complete subgraph K_{k-1} in the subgraph of these $p_1 = R(k-1,l)$ vertices, which together with v yields a K_k in G, or a complete subgraph K_l in $\overline{G}$. In the second case there exists a complete subgraph K_k in G on the $p_2 = R(k,l-1)$ vertices, or a K_{l-1} in the complement which together with v yields a K_l in $\overline{G}$. Thus in both cases the claim has been established, and the proof is finished. ■

Therefore the content of our puzzle mentioned above, is that $R(3,3) = 6$. Only a few Ramsey numbers are known, and from complexity considerations there is no hope whatsoever of obtaining a formula for the Ramsey numbers in closed form. The significance of Ramsey's Theorem lies in the fact that it encompasses, as does also Hall's Theorem 6.3, a general *existence assertion*. The reader can obtain some insight about the many variations and generalizations of 7.21 from the book by Graham-Rothschild-Spencer.

EXERCISES FOR CHAPTER 7

1. Complete the proof of 7.2(v).
2. Verify Eberhard's Theorem for $p_3 = 2$, $p_4 = 4$, $p_5 = p_9 = 1$. What is the smallest p_6?

3°. Prove that the graph in Figure 7.8 is not Hamiltonian.

4. Which cubes Q_n are Hamiltonian? Which graphs $K_{m,n}$?

5. Prove: If G is Hamiltonian, then so is the line graph $L(G)$. Show that the converse is false.

6°. A graph G is called almost Hamiltonian if G is not Hamiltonian but every subgraph $G-u$ is. Show that there are no almost Hamiltonian graphs with $p \leq 8$ vertices. Give an example of such a graph with 10 vertices.

7°. Let the numbers $\alpha(G)$ and $\kappa(G)$ be defined as in 7.8. Show: If $d(u)+d(v) \geq p$ for every pair of nonadjacent vertices u and v, then $\kappa(G) \geq \alpha(G)$.

8. Construct a non-Hamiltonian graph on 10 vertices for which $d(u) + d(v) \geq 9$ holds for any two nonadjacent vertices.

9*. Prove: If $d(u)+d(v) \geq p+1$ holds for any two nonadjacent vertices u and v, then there is a Hamiltonian path between any two vertices of the graph. Show that $d(u)+d(v) \geq p$ does not suffice.

10°. Show that the Hamiltonian closure (see 7.11) does not depend on the order in which the edges are added.

11. Show that 7.13 directly implies Ore's result, 7.10.

12°. Prove the following analogue of 7.13: Let $d_1 \leq d_2 \leq \cdots \leq d_p$ be the degree sequence of the simple graph G. If $d_k \leq k-1 < (p-1)/2$ always implies $d_{p-k+1} \geq p-k$, then G has a Hamiltonian path.

13. Deduce from the preceding exercise that a simple graph with p vertices and more than $\binom{p-1}{2}$ edges always has a Hamiltonian path.

14°*. Prove 7.15. (Hint: Show first that G has a Hamiltonian path in any case.)

15*. Show that G^3 is always Hamiltonian when G is connected.

16°. From exercise 13 above, we know that $\mathrm{ex}(P_{p-1}, p) \leq \binom{p-1}{2}$. Show that equality may hold and determine the extremal graphs.

17*. Prove: If G is simple and 2-connected with $d(u) \geq k$, $k < p/2$ for all vertices u, then G contains a circuit of length at least $2k$.

18°. Deduce the following from the preceding exercise: If G is a simple graph on p vertices and more than $k(p-1)/2$ edges, then G contains a circuit of length $\geq k+1$. Determine all graphs with $k(p-1)/2$ edges that do not have a circuit of length $\geq k+1$.

19. Let $\alpha(G)$ be the independence number and $d(G)$ the support number (=minimal number of vertices that meet all edges, see 6.1). Show that $\alpha(G)+d(G)=p$ always holds, and deduce from this that for simple graphs without triangles, $q \leq \alpha d$ and $q \leq p^2/4$.

20*. Somewhat more difficult. We know that a simple graph G on p vertices, and without triangles, has at most $p^2/4$ edges. Let $p=2n$ and suppose G has $(p^2/4)+1$ edges. Show that then G contains not only one, but at least n triangles. Analogously: If G has $(p^2/4)+2$ edges, then G contains at least $p=2n$ triangles. Generalization for G containing $(p^2/4)+k$ edges?

21. Generalize Ramsey's Theorem. For $l_1, \ldots, l_r$ there is a smallest number $R(l_1, \ldots, l_r)$ such that the following holds: For $p \geq R(l_1, \ldots, l_r)$ vertices and for each coloring of the edges of K_p with r colors there is an i, such that K_p contains a complete subgraph on l_i vertices, all of whose edges are colored with color i.

22. Let $r_m = R(\underbrace{3,3,\ldots,3}_{m})$. We know that $r_2=6$. Show by induction that $r_{m+1} \leq (m+1)(r_m-1)+2$ holds for all $m \geq 2$.

8. MATROIDS

As was also the case with the previous topics, our last theory begins in the 30's—more precisely, with Whitney's Theorem 4.11. Let us recall the central idea: Let $G(V, E)$ be a graph. We call $G^\star(V^\star, E^\star)$ a Whitney dual if there is a bijection $\varphi : E \to E^\star$ such that $C \subseteq E$ is a polygon in G if and only if $\varphi C \subseteq E^\star$ is a bond in $G^\star$. The content of Whitney's Theorem was: It is precisely the planar graphs G which have a W-dual, namely the dual graph $G^\star$ of any plane realization. But what if G is not planar, can we still define a meaningful "dual" structure?

To answer this question, Whitney investigated the exact structure of polygons and bonds, and by that, came back to Veblen's concept of the incidence matrix. Let $A = (a_{ij})$ be the vertex-edge incidence matrix of the graph $G(V, E)$. If we think of A as a matrix over $GF(2)$, then the columns s_j (corresponding to the edges e_j) are 0,1-vectors of length $|V|$, and we have seen that the minimal linearly independent sets $\{s_{j_1}, \dots, s_{j_t}\}$ correspond exactly to the polygons $\{e_{j_1}, \dots, e_{j_t}\}$ of G. Formulated differently, $\{s_{i_1}, \dots, s_{i_l}\}$ is linearly independent if and only if $I = \{e_{i_1}, \dots, e_{i_l}\}$ contains no polygons, i.e., I is a **forest**.

What structural properties do forests or circuits possess? The abstraction of these properties (and corresponding to this, the linear independence, resp. dependence in vector spaces) led to a class of combinatorial structures that, since the work of Tutte and Edmonds in the 60's, have been prominent in obtaining entirely new insights and interconnections. We begin with the following fundamental definition.

DEFINITION. *A* **matroid** $\mathcal{M} = \mathcal{M}(S)$ *is a finite set* S *together with a family* $\mathcal{I} \subseteq 2^S$ *of subsets of* S *that satisfy the following axioms:*

1) $\emptyset \in \mathcal{I}$; $I \in \mathcal{I}$, $J \subseteq I \Rightarrow J \in \mathcal{I}$.

2) *For each* $A \subseteq S$, *all maximal subsets of* A *that lie in* $\mathcal{I}$ *have the same size. We call this size the* **rank**, $r(A)$, *of* A.

The sets of $\mathcal{I}$ are called the **independent sets** of the matroid.

In words: Every subset of an independent set is independent, and the maximal independent subsets of a set $A \subseteq S$ all have the same size. If we replace "independent" by **linearly independent**, these axioms are well-known theorems of Linear Algebra. Thus, if V is any vector space, then every finite set S of vectors, with the linearly independent sets as the family $\mathcal{I}$, forms a matroid. Every matroid obtained in this way is called a **linear matroid**. From this example which can always be represented by a matrix, Whitney, of course, derived the name **matroid**.

The following isomorphism concept is clear: Two matroids $\mathcal{M}(S)$ and $\mathcal{M}'(S')$ are said to be **isomorphic** if there is a bijection $\varphi : S \to S'$ with $I \in \mathcal{I}$ if and only if $\varphi I \in \mathcal{I}'$.

The second class—which interests us primarily—are graphs. Let $G(V, E)$ be an arbitrary graph. We declare an edge set $I \subseteq E$ to be **independent** if the subgraph $G(V, I)$ is a forest in G. For the sake of brevity, we always identify $A \subseteq E$ with the subgraph $G(V, A)$. Axiom 1) is clearly satisfied for the family of forests, but 2) is also. Indeed, if $A \subseteq E$, and if $c(A)$ denotes the number of components of A, then we know that $|I| = |V| - c(A)$ holds for each maximal forest I in A. Therefore the size of a maximal independent set in A is always the same, namely $r(A) = |V| - c(A)$. The matroid so determined is called the **polygon matroid** $\mathcal{P}(G(V, E))$ of G. Figure 8.1 shows a graph G and its maximal forests.

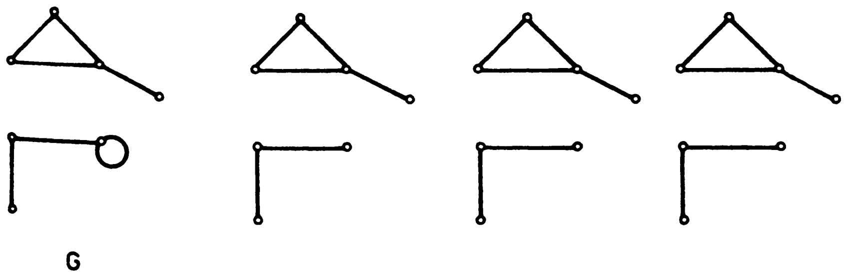


Figure 8.1

An abstract matroid that is isomorphic to a polygon matroid is called **graphic**. The smallest non-graphic matroid is (up to isomorphism) $S = \{a, b, c, d\}$, $\mathcal{I} = \{\emptyset, a, b, c, d, ab, ac, ad, bc, bd, cd\}$. The reader may prove for himself why there can not be a graph $G(V, S)$ with $S = \{a, b, c, d\}$ whose forests correspond exactly to the sets from $\mathcal{I}$. In $\mathcal{P}(G(V, E))$, the loops of G are exactly the one-element dependent sets and each pair of multiple edges are the two-element dependent sets. In analogy with this, in an arbitrary matroid $\mathcal{M}$, one also calls the dependent one-element sets the **loops** of $\mathcal{M}$. $\mathcal{M}$ is called **simple** if $\mathcal{M}$ contains no one-element or two-element dependent sets.

Let S and a family $\mathcal{I} \subseteq 2^S$ be given that satisfies axiom 1). A useful observation is that axiom 2) is equivalent to the following **exchange axiom**:

2′) If $I, J \in \mathcal{I}$, with $|I| < |J|$, then there is a $p \in J - I$ with $I \bigcup p \in \mathcal{I}$.

Indeed, if namely, 2) is satisfied and we set $A = I \bigcup J$, then all maximal $\mathcal{I}$-sets of A have the same size $r(A)$; therefore, since $|I| < |J| \leq r(A)$, there must be such a $p \in J - I$. Conversely, if 2′) is satisfied, then clearly two maximal $\mathcal{I}$-sets in A can not have distinct sizes.

In the paper of 1935 in which he introduced matroids, Whitney gave other axiomatic descriptions of which we shall discuss two. We call each maximal $\mathcal{I}$-set in A a **basis** of A; and a maximal $\mathcal{I}$-set of S, a **basis of the matroid**. Clearly $\mathcal{I}$ determines the family $\mathcal{B}$ of bases of $\mathcal{M}$ (these are precisely the maximal members), but the converse is also correct, for by axiom 1), $\mathcal{I}$ consists of exactly the subsets of the bases.

8.1 THEOREM. *Let $\mathcal{M}(S)$ be a matroid and $\mathcal{B}$ the family of bases. Then the following hold:*

1) $B \neq B' \in \mathcal{B} \Rightarrow B \not\subseteq B', B' \not\subseteq B$.

2) *Let $B \neq B' \in \mathcal{B}$. Then for each $b \in B$ there is a $b' \in B'$ with $(B - b) \bigcup b' \in \mathcal{B}$.*

Conversely, if $\mathcal{B}$ is a nonempty family of sets that satisfies 1) and 2), then there is a unique matroid that has $\mathcal{B}$ as a basis family.

Proof. 1) is clear and 2) follows from axiom 2′) with $B - b$, $B' \in \mathcal{I}$, since $|B - b| < |B'|$. On the other hand, if $\mathcal{B}$ satisfies the conditions of the theorem, then we define $\mathcal{I}$ to be the family of all subsets of $\mathcal{B}$-sets. Then axioms 1) and 2) are immediately verified. ∎

If $A \subseteq S$ and $A \notin \mathcal{I}$, then we call A **dependent**. Along with $\mathcal{I}$, the family $\mathcal{A} = 2^S - \mathcal{I}$ of dependent sets also determines the matroid. Since by axiom 1), every set containing a dependent set is again dependent, then the minimal dependent sets alone suffice to uniquely determine the matroid. In the polygon matroid of a graph, these minimal dependent sets are, of course, exactly the **polygons**, i.e., the edge-sets of circuits in the graph. Leaning on that one calls the minimal dependent sets in an arbitrary matroid $\mathcal{M}$ the **circuits** of $\mathcal{M}$.

8.2 THEOREM. *Let $\mathcal{M}(S)$ be a matroid and let $\mathcal{K}$ be the family of circuits. Then the following hold:*

1) $\emptyset \notin \mathcal{K}; C \neq C' \in \mathcal{K} \Rightarrow C \not\subseteq C', C' \not\subseteq C$.

2) $C \neq C' \in \mathcal{K}, p \in C \bigcap C' \Rightarrow$ *there is a* $D \in \mathcal{K}$ *with* $D \subseteq (C \bigcup C') - p$.

Conversely, if $\mathcal{K}$ is a family of sets that satisfies 1) and 2), then there is a unique matroid that has $\mathcal{K}$ as its family of circuits.

We leave the proof to the reader. It is interesting that 2) is equivalent to the following seemingly sharper condition:

2′) $C \neq C' \in \mathcal{K}$, $p \in C \bigcap C'$, $q \in C - C' \Rightarrow$ there is a $D \in \mathcal{K}$ with $q \in D \subseteq (C \bigcup C') - p$.

Condition 2′), carried over to graphs, yields a theorem on polygons that is by no means obvious.

We return to the starting point. Is there a dual structure for a nonplanar graph G? If we identify G with the polygon matroid $\mathcal{P}(G)$, then Whitney gives an elegant answer: *Every* matroid $\mathcal{M}$ has a natural dual $\mathcal{M}^*$ which, in the case of $\mathcal{P}(G)$ with G plane, is exactly the polygon matroid $\mathcal{P}(G^*)$ of the dual graph G^*.

8.3 THEOREM. *Let $\mathcal{M}(S)$ be a matroid with basis family $\mathcal{B}$. The family $\mathcal{B}^* = \{A \subseteq S : S - A \in \mathcal{B}\}$ of complementary sets satisfies the conditions of* 8.1 *and thus defines a unique matroid that we call the* **dual matroid** *$\mathcal{M}^*(S)$ of $\mathcal{M}$. Clearly, $\mathcal{M}^{**} = \mathcal{M}$.*

Proof. Condition 1) in 8.1 is clear. Now if $A \neq A' \in \mathcal{B}^*$ and $a \in A$, then we must find $a' \in A'$ with $(A - a) \bigcup a' \in \mathcal{B}^*$. If $a \in A \bigcap A'$, then we can pick $a' = a$. Hence we may assume that $a \in A - A'$. Since the basis $S - A$ is maximal independent in $\mathcal{M}$, there must be a circuit C in $(S - A) \bigcup a$ (which, by 8.2, is uniquely determined) that contains a, i.e., $a \in C \subseteq (S - A) \bigcup a$. C can not lie in the $\mathcal{B}$-set $S - A'$. Therefore, there is an $a' \in C - (S - A') \subseteq (S - A) \bigcup a - (S - A') = A' - A$. The set $((S - A) - a') \bigcup a$ is independent, since otherwise it would contain the uniquely determined circuit C from $(S - A) \bigcup a$, which contradicts $a' \in C$. Thus, because $|((S - A) - a') \bigcup a| = |S - A|$, $((S - A) - a') \bigcup a \in \mathcal{B}$ and hence $(A - a) \bigcup a' \in \mathcal{B}^*$. ∎

The following assertion clarifies once more the connection between a matroid and its dual via the families of circuits.

8.4 THEOREM. *Let $\mathcal{M}(S)$ be a matroid with the family of circuits $\mathcal{K}$. A set A is a circuit in $\mathcal{M}^*(S)$ if and only if $|A \bigcap C| \neq 1$ for all $C \in \mathcal{K}$, and if $A \neq \emptyset$ is minimal with respect to this property.*

Proof. We split the proof into two parts:

a. A a circuit in $\mathcal{M}^* \Rightarrow |A \bigcap C| \neq 1$ for all $C \in \mathcal{K}$.

b. $|A \bigcap C| \neq 1$ for all $C \in \mathcal{K} \Rightarrow A$ is dependent in $\mathcal{M}^*$.

First, on the basis of 8.3, we note that A is independent in $\mathcal{M}^*$ if and only if $S - A$ contains a basis of $\mathcal{M}$, or, restated: A is dependent in $\mathcal{M}^*$ if and only if $S - A$ contains no basis of $\mathcal{M}$.

Thus, A is a circuit of $\mathcal{M}^*$ if $S - A$ contains no basis of $\mathcal{M}$ and is maximal with respect to this property. Let A be a circuit in $\mathcal{M}^*$. Assume $C \in \mathcal{K}$ exists with $C \bigcap A = \{p\}$. By what we just said, $S - A$ contains no basis of $\mathcal{M}$, but $(S - A) \bigcup p$ does; therefore, the set $C - p \subseteq S - A$ which is independent in $\mathcal{M}$ (C is a circuit) can be extended to a basis $B \subseteq (S - A) \bigcup p$ with $p \in B$. But since B contains the dependent set C, this can not be. Conversely, suppose A is a set with the property $|A \bigcap C| \neq 1$ for all $C \in \mathcal{K}$. If A were independent in $\mathcal{M}^*$, then $S - A$ would contain a basis B of $\mathcal{M}$. For $p \in A$, $B \bigcup p$ is dependent in $\mathcal{M}$, therefore it contains a circuit C, for which clearly $A \bigcap C = \{p\}$. ■

Let us apply 8.4 to the polygon matroid $\mathcal{P}(G(V, E))$ of a graph in order to describe graph-theoretically the matroid dual to $\mathcal{P}(G)$.

CLAIM: *The circuits in $\mathcal{P}(G)^*$ are precisely the bonds of G.*

Since a bond A is a bipartition, then as was remarked in Chapter 4, $|A \bigcap C|$ must be even for each polygon $C \in \mathcal{K}$, hence in particular $|A \bigcap C| \neq 1$ must hold. Conversely, if $A \subseteq E$ satisfies the condition $|A \bigcap C| \neq 1$ for all polygons C, then A is a separating set of edges. Indeed, if $e = uv \in A$, then the vertices u and v can no longer be joined in $G - A$, since otherwise there would be a polygon C with $C \bigcap A = \{e\}$. Since the bonds are precisely the minimal separating edge sets, our claim is proven. On the basis of this, we speak of $\mathcal{P}(G)^*$ as the **bond matroid** $\mathcal{B}(G(V, E))$ of the graph G, and we call the matroid $\mathcal{M}$ **cographic** if it is isomorphic to a bond matroid. We already know that $r(\mathcal{P}(G(V, E))) = |V| - c(G)$. Now it follows from this that $r(\mathcal{B}(G(V, E))) = |E| - |V| + c(G)$.

If we examine the definition of a W-dual from Chapter 4, then Whitney's Theorem 4.11, in its matroid formulation, reads: A graph G is planar if and only if its bond matroid $\mathcal{B}(G)$ is graphic (or, equivalently: its polygon matroid is cographic).

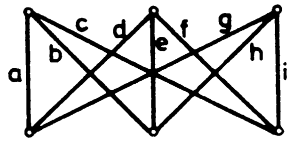


Figure 8.2

Let $K_{3,3}$ be as depicted in Figure 8.2. Since $K_{3,3}$ is not planar, $\mathcal{B}(K_{3,3})$ is not graphic. $\mathcal{P}(K_{3,3})$ has rank $|V| - c(K_{3,3}) = 5$ and thus $\mathcal{B}(K_{3,3})$ has rank $|E| - 5 = 4$. An example of a basis of $\mathcal{B}(K_{3,3})$ is $\{a, b, d, h\}$, and $\{g, h, i\}$ is an example of a circuit.

Until now the reader has perhaps gotten the impression that matroid theory yields only a formal framework for well-known graph theorems. The following examples, that yield an entirely new and elegant approach to transversal and decomposition theorems, will correct this objection.

Let a function $f : 2^S \to \mathbb{N}_0$ with $f(\emptyset) = 0$. f is called **monotone** if $A \subseteq B$ implies $f(A) \le f(B)$, and it is called **submodular** if $f(A \cap B) + f(A \cup B) \le f(A) + f(B)$ holds for all A and B. If f is monotone and submodular, then, as one convinces oneself immediately, so is the function $\hat{f}$, where $\hat{f}(A)$ is defined to be $\min_{B \subseteq A}(f(B) + |A - B|)$.

From the definition of $\hat{f}$, we have:

$$(\star) \qquad \hat{f}(A) = |A| \quad \Rightarrow \quad |B| \le f(B) \text{ for all } B \subseteq A.$$

The following remarkable theorem shows that *every* monotone submodular function generates a matroid.

8.5 THEOREM (Edmonds). *Let $f : 2^S \to \mathbb{N}_0$ be a monotone and submodular function with $f(\emptyset) = 0$. The family $\mathcal{I} \subseteq 2^S$ defined by*

$$I \in \mathcal{I} \Leftrightarrow |B| \le f(B) \text{ for all } B \subseteq I$$

is the family of independent sets of a matroid $\mathcal{M}_f$. The rank function in $\mathcal{M}_f$ is given by

$$r(A) = \min_{B \subseteq A}(f(B) + |A - B|).$$

Proof. We consider the family $\mathcal{A} = \{A \subseteq S : f(A) < |A|\}$ and verify that the minimal sets in $\mathcal{A}$ satisfy the circuit axioms in 8.2. Condition 1) is clear. Let $C \ne C'$ be minimal sets in $\mathcal{A}$ with $p \in C \cap C'$. Then $|C| \ge 2$ and thus

$$|C| - 1 = |C - p| \le f(C - p) \le f(C) < |C|.$$

Therefore $f(C) = |C| - 1$ and $f(D) \ge |D|$ for all $D \subseteq C$, and similarly $f(C') = |C'| - 1$. In particular, we have $|C \cap C'| \le f(C \cap C')$, and by the submodularity of f,

$$\begin{aligned} f(C \cup C' - p) &\le f(C \cup C') \le f(C) + f(C') - f(C \cap C') \\ &\le |C| - 1 + |C'| - 1 - |C \cap C'| \\ &= |C \cup C'| - 2 < |(C \cup C') - p|. \end{aligned}$$

Hence $(C \cup C') - p$ is in $\mathcal{A}$ and thus contains a minimal set from $\mathcal{A}$. As a result, the minimal sets from $\mathcal{A}$ generate a matroid $\mathcal{M}_f$. From the monotonicity and submodularity of $\hat{f}$, it can easily be derived that $\hat{f}$ is the rank function of a certain matroid $\mathcal{M}'$ on S. But since the independent sets A in $\mathcal{M}'$ are obviously characterized by the condition $\hat{f}(A) = |A|$, then because of $(\star)$, the matroids $\mathcal{M}_f$ and $\mathcal{M}'$ contain exactly the same independent sets. Therefore they are identical and $\hat{f}$ is in fact the rank function in $\mathcal{M}_f$, as claimed. ∎

The theorem proven above is perhaps the most fruitful construction method for matroids that has been discovered up to now. Every monotone submodular function, no matter how defined, gives a matroid. Two examples will illustrate this.

Let $G(S \bigcup T, E)$ be a bipartite graph. As in Chapter 6, we set $R(A) = \{v \in T : \exists u \in A \text{ with } uv \in E\}$ for $A \subseteq S$. Let $f : 2^S \to \mathbb{N}_0$ be defined by $f(A) = |R(A)|$, $A \subseteq S$. Clearly, along with $A \subseteq B$ we also have $R(A) \subseteq R(B)$; thus f is monotone. Furthermore, we have $R(A \bigcup B) = R(A) \bigcup R(B)$ and $R(A \bigcap B) \subseteq R(A) \bigcap R(B)$, and hence $f(A) + f(B) \geq f(A \bigcup B) + f(A \bigcap B)$, and, trivially, also $f(\emptyset) = 0$. By our theorem, f generates a matroid $\mathcal{M}_f$ on S whose independent sets I are characterized by the condition

$$I \in \mathcal{I} \quad \Leftrightarrow \quad |B| \leq |R(B)| \text{ for all } B \subseteq I.$$

But that is exactly Hall's condition 6.3 for the existence of a matching M in G that contains in S exactly the vertices from I. If we interpret $G(S \bigcup T, E)$, as in Chapter 6, to be a family of sets $(S; \mathcal{A} = (A_1, \ldots, A_m))$, then the independent sets in $\mathcal{M}_f$ are exactly the **transversals** of subfamilies of $\mathcal{A}$ or as we say the **partial transversals**. For this reason, $\mathcal{M}_f$ is called the **transversal matroid** generated by $(S; \mathcal{A})$, and we see that the rank formula 8.5 is nothing other than formula 6.2.

Now we can apply all the theorems of matroid theory. For example, the definition of a matroid says that every partial transversal can be extended to a transversal of maximal size (which are all of the same size) or that the exchange axiom 2') holds for partial transversals—neither result is at all obvious.

The second example refers directly to the rank function r of a matroid $\mathcal{M}(S)$. Clearly, r is monotone with $r(\emptyset) = 0$. To prove the submodularity we choose a basis C of $A \bigcap B$ and extend C to a basis D of $A \bigcup B$. Therefore both $C = (D \bigcap A) \bigcap (D \bigcap B)$ and $D = (D \bigcap A) \bigcup (D \bigcap B)$ hold, and thus

$$\begin{aligned} r(A \bigcap B) + r(A \bigcup B) &= |C| + |D| \\ &= |(D \bigcap A) \bigcap (D \bigcap B)| + |(D \bigcap A) \bigcup (D \bigcap B)| \\ &= |D \bigcap A| + |D \bigcap B| \leq r(A) + r(B), \end{aligned}$$

since $D \bigcap A$ and $D \bigcap B$ are independent sets in A, resp. B.

Now let t matroids $\mathcal{M}_1, \ldots, \mathcal{M}_t$ be given on S with the rank functions $r_1, \ldots, r_t$. Since each individual function r_i is monotone and submodular, this also holds for the sum $f = \sum_{i=1}^t r_i$. The matroid induced by f, via 8.5, is called the **sum** $\sum_{i=1}^t \mathcal{M}_i(S)$. A surprisingly simple description of the independent sets in $\sum \mathcal{M}_i$ was given by Nash-Williams: A is independent in $\sum \mathcal{M}_i$ if and only if $A = \bigcup_{i=1}^t A_i$ with A_i independent in $\mathcal{M}_i$ for all i. It is clear that this condition is equivalent to the requirement that $A = \bigcup_{i=1}^t A_i$ is the *disjoint* union of independent sets A_i since we can omit elements that occur several times except for one occurrence. That each such union is independent in $\sum \mathcal{M}_i$ follows immediately from 8.5. Namely, if $B \subseteq A$, $B = \bigcup_{i=1}^t B_i$, and $B_i \subseteq A_i$ for all i, then we have

$$|B| \le \sum_{i=1}^t |B_i| = \sum_{i=1}^t r_i(B_i) \le \sum_{i=1}^t r_i(B_i) = f(B).$$

The converse may be checked in the book by Aigner, p.291.

If in particular $\mathcal{M}_1 = \cdots = \mathcal{M}_t = \mathcal{M}$, then we obtain very interesting packing and covering results that can be derived directly only with considerable difficulty. If r is the rank function of $\mathcal{M}$, then by 8.5 the rank function r' of the t-fold sum $\sum_t \mathcal{M}$ is given by:

$$r'(A) = \min_{B \subseteq A} (t \cdot r(B) + |A - B|).$$

The **packing problem** consists in packing as many disjoint "large" sets (i.e., sets that contain a basis) as possible in S. The **covering problem** asks for a covering of S with as few "small" (=independent) sets as possible. We can now easily give the answer to both questions.

8.6 THEOREM.

i) *S contains t disjoint bases* $\Leftrightarrow |S - B| \ge t(r(S) - r(B))$ *for all* $B \subseteq S$.

ii) *S is the union of t independent sets* $\Leftrightarrow |B| \le t \cdot r(B)$ *for all* $B \subseteq S$.

Proof. By the result of Nash-Williams, S contains t disjoint bases if and only if the rank, $r'(S)$, in $\sum_t \mathcal{M}$ is at least $t \cdot r(S)$. But by the formula for r' this is equivalent with i). Condition ii) is just as easily disposed of. That S is the union of t independent sets means precisely that S is independent in $\sum_t \mathcal{M}$, i.e., $|S| = r'(S)$, and this is equivalent to $|S| \le t \cdot r(B) + |S - B|$ for all B. ∎

Let the **packing number**, $\pi(\mathcal{M})$, be defined as the largest number of disjoint bases in $\mathcal{M}$, and the **covering number**, $\beta(\mathcal{M})$, as the smallest number of independent sets (which, as mentioned at the beginning, can be assumed to be disjoint) that form a covering of $\mathcal{M}$. Application of 8.6 yields:

8.7 COROLLARY. *The following hold:*

i) $$\pi(\mathcal{M}) = \min_{\substack{B\subseteq S,\\ r(B)\neq r(S)}} \left\lfloor \frac{|S-B|}{r(S)-r(B)} \right\rfloor,$$

ii) $$\beta(\mathcal{M}) = \min_{\substack{B\subseteq S,\\ r(B)\neq 0}} \left\lceil \frac{|B|}{r(B)} \right\rceil,$$

where $\mathcal{M}$ *is assumed to be without loops in* ii), *since otherwise an independent covering does not exist.*

As we know, the forests are exactly the independent sets in the polygon matroid $\mathcal{P}(G(V,E))$. Accordingly, $\pi(\mathcal{P})$ resp. $\beta(\mathcal{P})$ give the maximal number of disjoint maximal forests, resp. the minimal number of forests into which K can be decomposed. Therefore, $\beta(\mathcal{P})$ is nothing other than the **arboricity** of the graph G defined at the end of Chapter 6, and the reader can easily convince himself that 8.7(ii) gives exactly formula 6.19.

After this excursion to general matroids, we will now consider Veblen's ideas for the 4-color problem from Chapter 3, and study Tutte's analysis of these ideas. In Chapter 4, as a generalization of Veblen's attack, we considered the vertex-edge-incidence matrix A for an arbitrary graph $G(V,E)$ and the circuit-edge incidence matrix B. We established that (with the usual identification, set=characteristic vector) the sets from the subspace $\hat{A}$ generated by A are exactly the **bipartitions** of G which we also called **cocycles**, while the sets from $\hat{B}$ are precisely the **Euler subgraphs**, resp. **cycles**. Furthermore, $\hat{A} = \hat{B}^{\perp}$ and $\hat{B} = \hat{A}^{\perp}$ hold.

Let us now discuss the connection with matroids. As we know, the minimal linearly independent sets of columns of A correspond exactly to the polygons of G. But this means that the linear matroid generated by the columns of A is isomorphic to the **polygon matroid** $\mathcal{P}(G)$. Analogously, we obtain for B that the minimal linearly dependent sets of columns correspond exactly to the bonds, and thus that the linear matroid generated by the columns of B is isomorphic to the **bond matroid** $\mathcal{B}(G)$. The algebraic translation of the fact that $\mathcal{P}(G)$ and $\mathcal{B}(G)$ are dual matroids, is given by the connection $\hat{A} = \hat{B}^{\perp}$. Therefore, it is natural to consider arbitrary 0,1-matrices R (not only, as it were, "graphic" matrices), and after that to define matroids. Then the dual matroid will be given by $\hat{R}^{\perp}$. Before we carry out this program, let us state the connection with the coloring problem.

In Chapter 3 we derived an equivalent formulation of the 4-color conjecture by means of two operators δf_1 and δf_2. Now we shall generalize this to arbitrary graphs. To this end, we interpret the cocycles and cycles in the following way: Let $G(V,E)$ be an arbitrary graph. A mapping $g : V \to GF(2)$ is called a **0-chain**, a mapping $f : E \to GF(2)$ a **1-chain**. The carrier of a 1-chain f, $\|f\|$, is defined to be $\{e \in E : f(e) = 1\}$ and analogously for 0-chains. Therefore f is merely the characteristic vector of $\|f\|$. For $v \in V$ and $e \in E$ we define the incidence symbol $\eta(v,e)$ by

$$\eta(v,e) = \begin{cases} 1, & \text{if } v \in e; \\ 0, & \text{if } v \notin e \text{ or } e \text{ is a loop.} \end{cases}$$

The 0-chains, resp. 1-chains form with the usual addition and scalar multiplication, two vector spaces over $GF(2)$ that we denote by $\mathcal{K}_0$, resp. $\mathcal{K}_1$. The **boundary operator**, ∂, is a mapping $\partial : \mathcal{K}_1 \to \mathcal{K}_0$ defined by

$$(\partial f)(v) := \sum_{e \in E} \eta(v,e) f(e) \qquad (f \in \mathcal{K}_1).$$

A 1-chain f is called a **cycle** if $\partial f \equiv 0$. Since $\partial(f+g) = \partial f + \partial g$, the cycles form a subspace of $\mathcal{K}_1$ that we call the **cycle space**, $Z(G)$ of G. Therefore f is a cycle if and only if $\partial f(v) = \sum_{\substack{e \in E \\ e \ni v}} f(e) \equiv 0 \pmod 2$ for all $v \in V$, i.e., if and only if the carrier $\|f\|$, as a subgraph of G, has even degree at each vertex, therefore it is a cycle in the sense used previously. We remark further that the **bridges** are exactly those edges e for which $f(e) = 0$ holds for every cycle f.

We next define the **coboundary operator**, $\delta : \mathcal{K}_0 \to \mathcal{K}_1$ by

$$(\delta g)(e) := \sum_{v \in V} \eta(v,e) g(v) \qquad (g \in \mathcal{K}_0).$$

Therefore $(\delta g)(e)$ is nothing other than the sum $(\delta g)(e) = g(u) + g(v)$ of the two g-values at the end points u and v of e. A 1-chain f is called a **cocycle** if $f = \delta g$ for some $g \in \mathcal{K}_0$. Since $\delta(g+h) = \delta(g) + \delta(h)$, the cocycles form a subspace of $\mathcal{K}_1$, the **cocycle space**, $C(G)$. In Figure 8.3, you find a 1-chain on the left and the corresponding boundary ∂f, whose values are circled. On the right a 0-chain g and the corresponding cocycle δg are shown, where the edges e with $(\delta g)(e) = 1$ are marked in boldface.

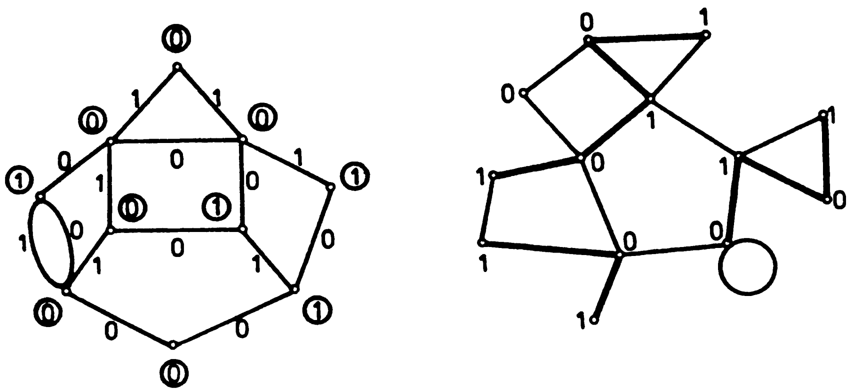


Figure 8.3

We want to show that the carriers $\|f\|$ of the cocycles correspond exactly to the cocycles (=bipartitions) of the graph. But this is almost obvious. For, when $f = \delta g \in C(G)$, with $f \neq 0$, then $\|f\|$ is exactly the bipartition generated by the vertex sets $V_0 = \{v \in V : g(v) = 0\}$ and $V_1 = \{v \in V : g(v) = 1\}$. Conversely, if $\|f\| \neq \emptyset$ is a bipartition with vertex sets V_0 and V_1, then we define $g : V \to GF(2)$ by $g(v) = 0$ or 1 depending on whether $v \in V_0$ or V_1. Then we obtain $f = \delta g$. Clearly we have $f(e) = 0$ for each cocycle f if and only if e is a *loop*.

We now come to the *coloring* of graphs. Assume G is 2^l-colorable. (Naturally the case $l = 2$ is of particular interest.) We take as the color set M, all l-tuples with entries 0,1. A mapping $g : V \to M$ can be described in terms of the coordinate functions $g_1, \ldots, g_l : V \to GF(2)$ by $g(v) = (g_1(v), g_2(v), \ldots, g_l(v))$. If g is an admissible coloring, then $g_i(u) \neq g_i(v)$ must hold for at least one i, whenever u and v are joined. In other words: $g : V \to M$ is a coloring if and only if $(\delta g_1(e), \ldots, \delta g_l(e)) \neq (0, \ldots, 0)$ holds for all $e \in E$. Let us state this as our first result.

8.8 THEOREM. *A graph $G(V,E)$ is 2^l-colorable if and only if there are l cocycles $\delta g_1, \ldots, \delta g_l$ such that $(\delta g_1(e), \ldots, \delta g_l(e)) \neq (0, \ldots, 0)$ for all $e \in E$. Or, if we identify 1-chains with their carriers: $G(V,E)$ is 2^l-colorable if and only if there are l bipartitions that cover all edges.*

The reader will have noted that this theorem is completely analogous to the (dual) statement in 3.3. From now on we concentrate exclusively on the 4-coloring of graphs.

8.9 COROLLARY. *A graph $G(V,E)$ is 4-colorable if and only if there are two cocycles $f_1, f_2 \in C(G)$ with $(f_1(e), f_2(e)) \neq (0,0)$ for all $e \in E$.*

We can make a corresponding statement also for cycle groups, which lets Tait's Theorem 1.7 appear in a new light.

8.10 THEOREM. *Let $G(V, E)$ be a 3-regular graph. G is 3-edge-colorable if and only if there are two cycles $f_1, f_2 \in Z(G)$ with $(f_1(e), f_2(e)) \neq (0,0)$ for all $e \in E$.*

Proof. Assume $f : E \to M$ is a 3-coloring of the edges, where we take $(1,0)$, $(0,1)$ and $(1,1)$ as the colors. We write $f(e) = (f_1(e), f_2(e))$ so that $(f_1(e), f_2(e)) \neq (0,0)$ always holds, and then verify that the 1-chains f_1, f_2 are cycles. Let the (arbitrary) vertex v be incident with the edges k, l, m. Since k, l, m are colored differently, exactly two lie in the carrier $\|f_1\|$ of f_1 and exactly two in $\|f_2\|$. But this means that $(\delta f_1)(v) = (\delta f_2)(v) = 0$, i.e., that f_1 and f_2 lie in $Z(G)$. Conversely, if f_1, $f_2 \in Z(G)$ are given with $(f_1(e), f_2(e)) \neq (0,0)$, then we define $f : E \to M$ by $f(e) = (f_1(e), f_2(e))$. Since each cycle assumes the value 0 on a bridge, G can not contain any bridges, and by 3-regularity, also no loops. Again, let k, l, m be the edges that are incident with a vertex v. Now if for example $f(k) = f(l)$ holds, then since $(\delta f_1)(v) = f_1(k) + f_1(l) + f_1(m) = 0$, we must have $f_1(m) = 0$ and similarly $f_2(m) = 0$, in contradiction to $(f_1(m), f_2(m)) \neq (0,0)$. ∎

Tait's Theorem is an immediate consequence of 8.10. Namely, if G is a bridgeless 3-regular plane graph, then, as we know, the cycles of G correspond exactly to the cocycles of the dual graph G^*. Therefore, by 8.9, G is 3-edge-colorable if and only if the dual graph G^* is 4-colorable, or what is the same thing, if and only if the countries of G can be colored with 4 colors.

The striking similarity of Theorems 8.9 and 8.10 led Tutte to study colorings of *arbitrary* subspaces of $\mathcal{K}_1$. In this connection, he succeeded in obtaining a significant deepening of the original ideas of Veblen.

DEFINITION. *Let S be a finite set, $|S| = n$. By a* **chain** *on S over $GF(2)$ we mean a mapping $f : S \to GF(2)$. Let $\mathcal{V}_n$ be the vector space of all chains. A* **chain group** *Γ on S is a subspace of $\mathcal{V}_n$. That is: $f, g \in \Gamma$ implies $f + g \in \Gamma$. As before, $\|f\| = \{s \in S : f(s) = 1\}$ is called the* **carrier** *of f, and $f \in \Gamma$ is called* **elementary**, *if $f \neq 0$ and $\|f\|$ is minimal (with respect to inclusion) among all carriers of Γ. The dimension of the subspace is the* **rank** *$r(\Gamma)$ of Γ.*

For example, in $Z(G)$ the elementary chains are precisely those whose carriers are polygons, and in $C(G)$ those whose carriers are bonds.

Two chains f, $g \in \mathcal{V}_n$ are called **orthogonal** if $\sum_{s \in S} f(s)g(s) = 0$. Those chains that are orthogonal to all chains of a group Γ again form a group, the **dual group** Γ^*. Obviously $\Gamma^{**} = \Gamma$ holds and furthermore $r(\Gamma^*) = n - r(\Gamma)$. In our example of a graph, that we will always keep in sight, $C(G) = Z(G)^*$.

We already know that $C(G)$ and $Z(G)$ each generate matroids, and indeed mutually dual matroids. Now we will prove the corresponding result for arbitrary chain groups.

8.11 THEOREM. *Let Γ be a chain group on S and Γ^* the dual group. Then the following hold:*

i) *If we define the circuits C as the carriers of the elementary chains from Γ^*, then we obtain by means of this family of circuits, $\mathcal{K}$, a matroid $\mathcal{M}(\Gamma)$ on S. Analogously, we obtain a matroid $\mathcal{M}(\Gamma^*)$.*

ii) *Every carrier in Γ^* is the disjoint union of circuits of the matroid $\mathcal{M}(\Gamma)$ and analogously every carrier in Γ is the disjoint union of circuits of the matroid $\mathcal{M}(\Gamma^*)$.*

iii) $\mathcal{M}(\Gamma)^* = \mathcal{M}(\Gamma^*)$. *In words: The dual matroid of a group is exactly the matroid of the dual group.*

iv) $r(\mathcal{M}(\Gamma)) = r(\Gamma)$, $r(\mathcal{M}(\Gamma^*)) = r(\Gamma^*)$.

Proof. To prove i) we must verify the axioms in 8.2. Condition 1) is clear. In order to simplify the calculation, we recall the connection established in Chapter 3: If $A = \|f\|$ and $B = \|g\|$, then the carrier of $f+g$ is the symmetric difference $A+B = \|f+g\|$. Therefore, because of the group property, for each two carriers their symmetric difference is also a carrier. Now if $C \neq C'$ are minimal carriers $\neq \emptyset$ in Γ^* and $p \in C \bigcap C'$, then $C + C' \subseteq (C \bigcup C') - p$ holds. Hence the carrier $C+C'$ contains a minimal carrier D, i.e., a circuit D with $D \subseteq (C \bigcup C') - p$. To prove ii), let A be a carrier in Γ^* and A_0 a maximal subset of A, which is the disjoint union of circuits, i.e., of minimal carriers from Γ^*. Thus A_0 is also a carrier and hence so is $A - A_0 = A + A_0$. If $A_0 \subsetneq A$, then $A - A_0$ contains a circuit C, so that we obtain $A_0 \subsetneq A_0 \dot{\bigcup} C \subseteq A$, which contradicts the maximality of A_0.

By definition, the circuits of $\mathcal{M}(\Gamma^*)$ are precisely the minimal carriers $A = \|f\|$, $f \in \Gamma$, while by 8.4, the circuits of the matroid $\mathcal{M}(\Gamma)^*$ dual to $\mathcal{M}(\Gamma)$ are exactly the minimal sets A with $|A \bigcap C| \neq 1$ for all $C \in \mathcal{K}$. Hence, for iii), we must show that these two families of sets coincide. Because of the minimality, the verification of the following claims suffice:

a. *If $A = \|f\|$, $f \in \Gamma$, is a minimal carrier in Γ, then $|A \bigcap C| \neq 1$ holds for all circuits $C = \|g\|$, $g \in \Gamma^*$, g elementary.*

b. *If A is a minimal set with $|A \bigcap C| \neq 1$ for all $C \in \mathcal{K}$, then $A = \|f\|$ with $f \in \Gamma$.*

To prove a. we only need to note that it follows from the orthogonality of $f \in \Gamma$ and $g \in \Gamma^*$ that $|A \bigcap C|$ is even for all $C \in \mathcal{K}$, therefore in particular $|A \bigcap C| \neq 1$.

Now let A be a set that satisfies b. We must show that $|A \bigcap C|$ is even for all $C \in \mathcal{K}$, $C = \|g\|$, $g \in \Gamma^\star$, since this means, because of ii), that $|A \bigcap B|$ is even for *all* carriers B of $\Gamma^\star$ or, in other words, that f with $\|f\| = A$ is orthogonal to all $g \in \Gamma^\star$, and therefore is in Γ. Assume this is false; then let $C \in \mathcal{K}$ be so chosen that $|A \bigcap C|$ is minimal odd. Since $|A \bigcap C| \neq 1$, we must have $|A \bigcap C| \geq 3$. Let $\{p, q\} \subseteq A \bigcap C$. Since A is a circuit in $\mathcal{M}(\Gamma)^\star$, then $A - p$ is independent and thus contained in a basis $B^\star$ of $\mathcal{M}(\Gamma)^\star$, $A - p \subseteq B^\star$, $p \notin B^\star$. By 8.3, $B = S - B^\star$ is a basis of $\mathcal{M}(\Gamma)$ with $B \subseteq (S - A) \bigcup p$, $p \in B$, $q \notin B$. The set $B \bigcup q$ is dependent in $\mathcal{M}(\Gamma)$ and thus contains a circuit $D \in \mathcal{K}$ with $q \in D \subseteq B \bigcup q \subseteq (S - A) \bigcup \{p, q\}$. A and D have q and at most also p in common, and because $|A \bigcap D| \neq 1$, then $A \bigcap D = \{p, q\}$ holds. From this, we obtain for the symmetric difference $C + D$ that $|A \bigcap (C + D)| = |A \bigcap C| - 2$. Since by ii), the carrier of $C + D$ in $\Gamma^\star$ decomposes into a disjoint union of circuits from $\mathcal{K}$, $C + D = C_1 \bigcup \cdots \bigcup C_t$. Hence there must be a circuit C_i with $|A \bigcap C_i| < |A \bigcap C|$ and $|A \bigcap C_i|$ is odd, in contradiction to the choice of C.

The proof of iv) is now easy. Let Γ be the given chain group. We index $S = \{s_1, \ldots, s_n\}$ and keep this numbering fixed. Now we write the chains of Γ as rows of a matrix R, and the chains of $\Gamma^\star$ as rows of a matrix $R^\star$. Therefore, R and $R^\star$ are both 0,1-matrices with n columns. By definition, the rank $r(\Gamma)$ is the row rank of R, and analogously $r(\Gamma^\star)$ is the row rank of $R^\star$. Because $\hat{R}^\star = \hat{R}^\perp$, it follows from i) and ii) that a set $A \subseteq S$ is independent in $\mathcal{M}(\Gamma)$ if and only if the set A of the corresponding columns of R is **linearly independent**; analogously for $\Gamma^\star$. Hence: the rank $r(\mathcal{M}(\Gamma))$ of the matroid $\mathcal{M}(\Gamma)$ is merely the column rank of R; therefore $r(\mathcal{M}(\Gamma)) = r(\Gamma)$ indeed holds. Analogously $r(\mathcal{M}(\Gamma^\star)) = r(\Gamma^\star)$. ∎

Let us stress the connection discussed at the conclusion of the preceding proof. Instead of *all* chains from Γ, we naturally need to write down only a generating set (e.g., a basis) of Γ as the rows of a matrix. Each matrix R obtained in this way has the following properties: The **row space** is the chain group Γ, the **columns** yield with linear independence exactly the matroid $\mathcal{M}(\Gamma)$. Therefore each matroid constructed in this way is linear—more precisely, linear over $GF(2)$, or as we say, **binary**. We call every such matrix R a **coordinatization matrix** of $\mathcal{M}(\Gamma)$. But the converse is also correct: If $\mathcal{M}$ is a binary matroid and we write the elements of $\mathcal{M}$, i.e. the vectors, as the columns of a matrix, then the rows of this matrix generate a chain group Γ whose corresponding matroid $\mathcal{M}(\Gamma)$ is precisely the initial matroid $\mathcal{M}$. And furthermore, our theorem says: If R is a coordinatization matrix of $\mathcal{M}$, then we obtain a coordinatization matrix $R^\star$ of the dual matroid $\mathcal{M}^\star$ by writing down a basis of the space dual to the row space of R, i.e., of the solution space of the system of equations $Rx = 0$.

EXAMPLE. *Our matroids $\mathcal{P}(G(V,E))$ and $\mathcal{B}(G(V,E))$ should again clarify this. We know that the chain group corresponding to $\mathcal{P}(G(V,E))$ consists exactly of the cocycles $C(G)$ and the group corresponding to $\mathcal{B}(G(V,E))$ is precisely $Z(G)$. For each vertex $v \in V$ we define the 0-chain $g_v : V \to GF(2)$ by*

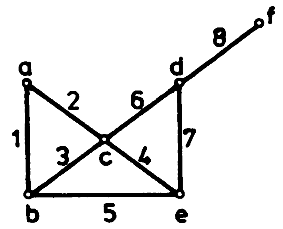
$$g_v(x) = \begin{cases} 1, & \textit{if } x = v; \\ 0, & \textit{otherwise.} \end{cases}$$

The corresponding cocycle δg_v satisfies

$$\delta g_v(e) = \begin{cases} 1, & \textit{if } v \in e, \textit{ where } e \textit{ is not a loop;} \\ 0, & \textit{otherwise.} \end{cases}$$

Therefore $\delta g_v(e)$ is nothing other than the incidence vector of v. Since the set $\{g_v : v \in V\}$ clearly spans the 0-chain space $\mathcal{K}_0$, then the cocycles $\{\delta g_v : v \in V\}$ generate the cocycle group C. If we write the vectors δg_v as rows of a 0,1-matrix, then what results is just the vertex-edge incidence matrix of G. In that way we obtain Veblen's original situation: The vertex-edge incidence matrix R is a coordinatization matrix of $\mathcal{P}(G(V,E))$, i.e., the linear matroid generated (with linear independence) by the columns of R is precisely the matroid $\mathcal{P}(G(V,E))$.

Figure 8.4 shows a graph G, the coordinatization matrix R of $\mathcal{P}(G)$, and a coordinatization matrix $R^\star$ of the bond matroid $\mathcal{B}(G)$ which one obtains by solving the system of equations $R\mathbf{x} = 0$.



$$R = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R^\star = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Figure 8.4

The rows of $R^\star$ yield the cycle basis $(123, 1245, 467)$. For example, $\{1,3,4\}$ is a basis of $\mathcal{B}(G)$, whose complementary basis $\{2,5,6,7,8\}$ is a tree in G.

Because of the correspondence $C(G) \leftrightarrow \mathcal{P}(G)$, $Z(G) \leftrightarrow \mathcal{B}(G)$, we also call the groups $C(G)$ **graphic** groups and the groups $Z(G)$ **cographic** groups.

We can now generalize our coloring concept to arbitrary chain groups. We give the following definition, motivated by 8.9 and 8.10.

DEFINITION. *Let Γ be a chain group on S. A* **coloring** *of Γ is a pair of chains $f, g \in \Gamma$ such that $(f(s), g(s)) \neq (0,0)$ holds for all $s \in S$. Or, in terms of sets: A pair of carriers $\|f\|$, $\|g\|$ from Γ, that together cover all elements of S. We say that Γ is* **colorable** *or* **chromatic** *if there is a coloring, otherwise* **achromatic**.

In this terminology, the statements 8.9 and 8.10 read as follows: $C(G)$ is chromatic if and only if G is 4-colorable. For a 3-regular graph G, $Z(G)$ is chromatic if and only if G is 3-edge-colorable. For example, $C(K_5)$ is achromatic and likewise $Z(P)$, where P is the Petersen graph.

In order to further analyze the existence of colorings we recall Hadwiger's conjecture from Chapter 5, for $n = 5$: A graph G that is not 4-colorable contains a subgraph that can be contracted to a graph K_5. The question of 4-colorablity—that is, in our terminology, the colorability of the cocycle group $C(G)$—therefore depends solely on the existence of such a K_5. In order to carry this over to arbitrary chain groups, we must thus first translate the concepts "subgraph" and "contraction".

DEFINITION. *Let Γ be a chain group on S and let $A \subseteq S$. For $f \in \Gamma$ we define f_A to be the restriction of f to A, i.e., $f_A(s) = f(s)$ for all $s \in A$. The set of distinct chains f_A, $f \in \Gamma$ form again a chain group, the* **restriction**, *$\Gamma.A$, of Γ.*

If $\mathcal{M}(\Gamma)$ is the matroid of Γ, then we clearly obtain the matroid corresponding to $\Gamma.A$ by taking all independent sets of $\mathcal{M}(\Gamma)$ that lie entirely in A. We call this matroid likewise the **restriction** of A, with the notation $\mathcal{M}(\Gamma).A = \mathcal{M}(\Gamma.A)$. Clearly, $r(\mathcal{M}(\Gamma).A) = r(A)$ holds, where $r(A)$ is the rank of A in $\mathcal{M}(\Gamma)$. Thus, for a graph $G(V,E)$, $\mathcal{P}(G(V,E)).A$ is exactly the polygon matroid of the subgraph $G(V,A)$. To retain the analogy, we also write $G.A$ for this subgraph and thus have:

$$C(G).A = C(G.A), \qquad \mathcal{P}(G).A = \mathcal{P}(G.A).$$

If (f,g) is a coloring of Γ, then clearly (f_A, g_A) yields a coloring of $\Gamma.A$, so that we can note:

8.12 THEOREM. *If the chain group Γ is colorable, then so also is every restriction $\Gamma.A$.*

Next we carry over the contractions.

DEFINITION. *Let Γ be a chain group on S and $A \subseteq S$. The set of chains $f \in \Gamma$ which vanish identically outside of A, form, with their restrictions, f_A, to A, another chain group, the* **contraction**, $\Gamma \times A$, *to A.*

What does this mean for graphs? We define the **contraction** $G \times A$ of G on $A \subseteq E$ in the following way: Let the vertices be the components of the subgraphs $G.(E - A)$, let the edges be those from A, where we join e with those components from $G.(E - A)$ that contain the end points of e in G. For example, if we choose in the graph G from Figure 8.5 the set $A = \{1, 2, 3, 4, \ldots, 9\}$, then $G.(E - A)$ has the components $\{a, b\}$, $\{c\}$, $\{d, e, f\}$, $\{g\}$, $\{h, i\}$ and we obtain as the contraction $G \times A$ the graph depicted on the right in Figure 8.5.

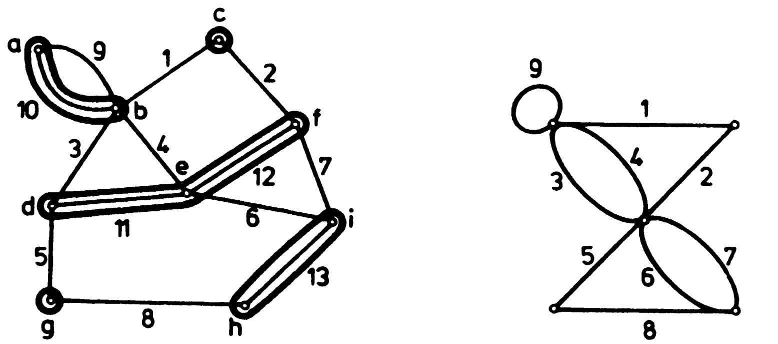


Figure 8.5

Therefore our construction $G \to G \times A$ proceeds in such a way that we contract the components to single vertices in $G.(E - A)$, or, in other words, that we contract the edges from $E - A$ one after the other—thus the name.

The reader can easily convince himself (by carrying out the contraction edge after edge as in the proof of 4.8) that the cocycles of $G \times A$ are just the chains from $C(G) \times A$. By the definition of a contraction, in the matroid $\mathcal{M}(\Gamma) \times A$ corresponding to $\Gamma \times A$ exactly those sets $B \subseteq A$ are independent for which $B \bigcup C$ is independent in $\mathcal{M}(\Gamma)$, for all independent sets $C \subseteq S - A$. We call $\mathcal{M}(\Gamma) \times A$ the matroid **contracted** through A, for whose rank, according to what we just stated, $r(\mathcal{M}(\Gamma) \times A) = r(\mathcal{M}(\Gamma)) - r(S - A)$ holds. Therefore, with this notation, we have

$$C(G) \times A = C(G \times A), \quad \mathcal{P}(G) \times A = \mathcal{P}(G \times A).$$

How can we interpret the restrictions and contractions of the cycle group, resp. of the bond matroid of a graph? To this end we collect some useful formulas. First of all, the following fact is clear:

8.13 LEMMA. *Let Γ be a chain group on S and $A \subseteq B \subseteq S$. Then:* $(\Gamma.B).A = \Gamma.A$, $(\Gamma \times B) \times A = \Gamma \times A$.

We now come to an important concept, that combines the two operators, restriction and contraction.

DEFINITION. *A* **minor** *of a chain group Γ on S is a chain group of the form $(\Gamma.B) \times A$ with $A \subseteq B \subseteq S$. Analogously, we speak of minors of the matroid $\mathcal{M}(\Gamma)$.*

Since $\Gamma.S = \Gamma$ and $\Gamma \times S = \Gamma$, every restriction and every contraction, and thus also Γ itself, is a minor. The following lemma, that can be obtained directly from the definition, shows that the order of the operations plays no role.

8.14 LEMMA. *Let Γ be a chain group on S and $A \subseteq B \subseteq S$. Then:*

$$(\Gamma.B) \times A = (\Gamma \times (S - (B - A))).A,$$

$$(\Gamma \times B).A = (\Gamma.(S - (B - A))) \times A.$$

Thus: Every minor of a minor is a minor.

8.15 LEMMA. *Let Γ be a chain group on S, Γ^* the dual group, and $A \subseteq S$. Then:*

$$(\Gamma.A)^* = \Gamma^* \times A, \quad \textit{and} \quad (\Gamma \times A)^* = \Gamma^*.A.$$

Proof. We prove the first assertion from which the second will follow by duality. Let f be an arbitrary chain on A and $\hat{f}$ be the chain on S that coincides with f on A, and is identically 0 outside A. f is orthogonal to all chains from $\Gamma.A$ (and thus is in $(\Gamma.A)^*$) if and only if $\hat{f}$ is orthogonal to all chains from Γ. But this means precisely that $f \in \Gamma^* \times A$. ■

As an application to the graph matroids we obtain:

$$Z(G.A) = Z(G) \times A, \qquad Z(G \times A) = Z(G).A,$$

$$\mathcal{B}(G).A = \mathcal{B}(G \times A), \qquad \mathcal{B}(G) \times A = \mathcal{B}(G.A).$$

Let us explicitly state these results for graphs.

8.16 THEOREM. *Every minor of a graphic group $C(G)$ is again graphic, every minor of a cographic group $Z(G)$ is again cographic. Or, in matroid form: Every minor of a graphic matroid $\mathcal{P}(G)$ is again graphic and every minor of a cographic matroid $\mathcal{B}(G)$ is again cographic.*

Let us assume the given chain group Γ is not colorable. This is trivially always the case if there is an element $a \in S$ on which all chains vanish. In the cocycle group $C(G)$ these are precisely the loops and in the cycle group $Z(G)$ they are the bridges of G. If Γ exhibits no such element, then we call Γ a **full chain group**. If the full group Γ is not colorable, then, by 8.12 and 8.13, there is a minimal (full) restriction $\Gamma.A$ that also is not colorable. Now we try to contract $\Gamma.A$ so long as the noncolorablity remains. With this we arrive at the following fundamental definition.

DEFINITION. *An* **irreducible chain group** *is a full achromatic chain group that contains no full achromatic minor except itself.*

Every full achromatic chain group must contain an irreducible group as a minor. Accordingly, if we succeed in making a list of all irreducible chain groups, then the problem of colorability is in principle solved: If the group Γ is achromatic then it must contain a minor from the list. Thus our (admittedly ambitious) program is: *Find all irreducible chain groups!*

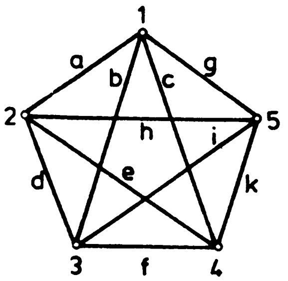
Because of 8.16, the irreducible groups contained in a cocycle group $C(G)$ must again be graphic, and analogously for $Z(G)$. Thus in this formulation Hadwiger's conjecture reads: *The cocycle group $C(K_5)$ is the only irreducible graphic group.*

That $C(K_5)$ is indeed irreducible follows immediately from the fact that every proper subgraph of K_5 and every graph on fewer than 5 vertices is 4-colorable.

Which cycle groups $Z(G)$ are irreducible? If G is 3-regular, then by 8.10, this means that in any case G is bridgeless and does not have a 3-edge-coloring. Naturally, it is the Petersen graph P that immediately comes to mind. The reader can convince himself (not entirely easily), that $Z(P)$ is in fact irreducible. Tutte took this as the occasion to state the following conjecture: *The cycle group $Z(P)$ is the only irreducible cographic group.*

We have already noted that Hadwiger's conjecture encompasses the 4-color conjecture, but this is also true for Tutte's conjecture. In graph terminology, it states: Every bridgeless 3-regular graph without 3-edge-coloring has a subgraph that can be contracted to P. But if G is a plane bridgeless 3-regular graph, then G can have no minor isomorphic to the nonplanar graph P. Thus it is 3-edge-colorable—whence with 1.7 the 4-color theorem would be proved.

Let us now consider what the matroids $\mathcal{M}(\Gamma)$ of irreducible chain groups look like. We take as an example $\mathcal{P}(K_5)$ with the incidence matrix as the coordinatizing matrix, whereby we can omit a row, e.g. the last one, since $r(\mathcal{P}) = 4$. We already know (Theorem 8.11 and the remarks following it) that $\mathcal{P}(K_5)$ is given precisely by the linear independence of the columns. Now we think of the columns as points in the projective space PG_3 over $GF(2)$ and try to describe geometrically the configuration obtained in this way—this was precisely Veblen's original idea.



$$R = \begin{array}{c} \begin{array}{cccccccccc} a & b & c & d & e & f & g & h & i & k \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Figure 8.6

What does one mean by the **projective space**, PG_{m-1}, over $GF(2)$? Nothing other than the subspace structure of the vector space $\mathcal{V}_m$ of dimension m over $GF(2)$. The **points** of PG_{m-1} are the 1-dimensional subspaces of $\mathcal{V}_m$, i.e. all vectors $\neq 0$ (geometric dimension 0). The **lines** are the 2-dimensional subspaces of $\mathcal{V}_m$ (geometric dimension 1), and in general a subspace of geometric dimension $k-1$ is a k-dimensional subspace of $\mathcal{V}_m$. Every line contains 3 points, namely, with P and Q, we get $P+Q$ as a third point. We denote by V_m the set of points of PG_{m-1}, that is all vectors $\neq 0$; and by $\overline{U}$ the subspace generated by $U \subseteq V_m$. For example, for $P \neq Q$, $\overline{PQ} = \{P, Q, P+Q\}$ is the line generated by P and Q.

In order to avoid confusion, we will always consider the **vector space dimension** and, as previously, speak of the **rank** $r(A)$ of a set $A \subseteq V_m$. The geometric dimension is always one smaller.

Consider $\mathcal{P}(K_5)$ in Figure 8.6. If we think of the columns of R as points in PG_3 then we see, e.g., that $a+b=d$, so that $\{a,b,d\}$ forms a line (corresponding to the polygon $\{a,b,d\}$ in K_5). As another example take $\{e,h,i,f\}$. Since this is a polygon in the graph, then of the corresponding 4 columns, each 3 of them must be independent, but all 4 together are dependent. In a geometric realization, and up to rank 4 (=geometric dimension 3) this can still be pictured, we shall try if possible to draw lines with 3 points as affine real lines, and all planes (therefore subspaces of rank 3) as affine real planes. If we carry this out for $\mathcal{P}(K_5)$, then we obtain the configuration depicted in Figure 8.7. The ten 3-point lines correspond exactly to the 10 triangles in K_5.

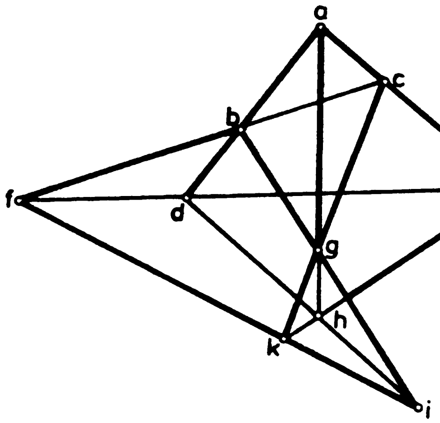


Figure 8.7

This configuration is well known in projective geometry. It reproduces the situation of Desargues' Theorem with center g and axis $\{f, d, e\}$. Thus we call the figure the **Desargues Block**.

In the general case we proceed in exactly the same way. For a given chain group Γ on S we take any coordinatizing matrix R with m rows $(m \geq r(\Gamma))$ and $n = |S|$ columns and think of the column vectors as points in PG_{m-1}. We call the point-configuration obtained in this way the corresponding **block** of Γ in PG_{m-1}. Now our problem consists in describing the blocks of **irreducible** chain groups **geometrically**. In the following we shall formulate the theorems purely algebraically, i.e. in the vector space $\mathcal{V}_m$, and then interpret them geometrically.

As the first step we characterize the full achromatic (but not necessarily minimal) configurations.

8.17 THEOREM. *Let Γ be a full chain group on S. Γ is achromatic if and only if the corresponding block B has rank ≥ 3 and, in $\mathcal{V}_m$, B has nontrivial intersection with every subspace of rank $m-2$.*

Proof. We can restrict ourselves to the case $m = r(\Gamma) = r(B)$. For, if B has nontrivial intersection with every subspace $U \subseteq \overline{B}$ with $r(U) = r(B) - 2$, then this also holds for every subspace $W \subseteq \mathcal{V}_m$ with $r(W) = m - 2$, since, by the rank formula for subspaces, $r(W \cap \overline{B}) = r(W) + r(\overline{B}) - r(W + \overline{B}) \geq m - 2 + r(B) - m = r(B) - 2$. Hence $W \cap \overline{B}$ contains a subspace of $\overline{B}$ of rank $\geq r(B) - 2$. We denote the vectors in $\mathcal{V}_m$ corresponding to $s_j \in S$ by b_j, $j = 1, \ldots, n$, and set $B = \{b_j : j = 1, \ldots, n\}$. Let $f \in \Gamma$, $f \neq 0$. We think of f as the mapping $f : B \to GF(2)$ defined by $f(b_j) = f(s_j)$. Since $r(B) = m$, B is a generating system of $\mathcal{V}_m$, so f can be uniquely extended to all $\mathcal{V}_m$, and we know from Linear

Algebra that $U_f = \{x \in \mathcal{V}_m : f(x) = 0\}$ forms a subspace of rank $m-1$, thus a hyperplane. On the other hand, since $r(\Gamma) = m$, all hyperplanes can be obtained in this way, and naturally $U_f \neq U_g$ holds for $f \neq g \in \Gamma$. Therefore: By $U = U_f$, each of the $2^m - 1$ hyperplanes U corresponds uniquely to a chain $f \neq 0$ in Γ.

Now let Γ be chromatic. If $m \leq 2$, then the condition $r(B) \geq 3$ is violated. Therefore we assume $m \geq 3$. By the definition of colorability, there are chains f and g in Γ such that $f(s) \neq 0$ or $g(s) \neq 0$ for each $s \in S$. But this means that the subspace $U_f \bigcap U_g$ (of rank $m-2$) does not contain a single point of B.

Conversely, assume that B does not satisfy both conditions. If $r(B) = m \geq 3$ is violated, then we have $m = 1$ or 2 ($m = 0$ is not possible since Γ is full), i.e., Γ contains a linearly independent chain f (identically 1) or two linearly independent chains f and g. In every case, $\{f, f\}$ resp. $\{f, g\}$ is a coloring; thus Γ is chromatic. The case remains where $m \geq 3$ and there is a subspace W of rank $m-2$ with $W \bigcap B = \emptyset$. But then $W = U_f \bigcap U_g$ with f and g in Γ; thus $W = \{x \in \mathcal{V}_m : f(x) = g(x) = 0\}$. Since $W \bigcap B = \emptyset$, it follows immediately that $\{f, g\}$ is a coloring of Γ. ∎

Let us call a block that satisfies the hypotheses of 8.17 a **2-block**. Therefore: Γ is achromatic if and only if B is a 2-block. In order to get from Γ to an irreducible minor, by 8.13 and 8.14,we can proceed in two steps. First, we determine a minimal achromatic restriction and from that a minimal full achromatic contraction. Clearly, the minimal achromatic restrictions correspond to the **minimal 2-blocks**. Therefore we only need to clarify how the blocks of Γ and a contraction $\Gamma \times A$ are related.

8.18 LEMMA. *Let Γ be a full achromatic chain group, $r(\Gamma) = m$, whose 2-block $B \subseteq V_m$ is minimal. Γ has a proper full achromatic contraction $\Gamma \times A$ if and only if there exists a C, $\emptyset \neq C \subseteq B$ with $r(C) \leq m-2$, such that every $(m-2)$-dimensional subspace W that contains C also contains a point $b \in B - \overline{C}$.*

Proof. We again denote by $b_j \in B$ the point corresponding to $s_j \in S$. Let $\Gamma \times A$ be a proper full achromatic contraction and $C \subseteq B$ the points of the block corresponding to $S - A$. By 8.17, we have $r(\Gamma \times A) \geq 3$. Therefore $r(C) = r(S - A) = r(\Gamma) - r(\Gamma \times A) \leq m - 3$. By definition, all chains $f \in \Gamma \times A$ vanish on $S - A$. If $b_j \in B - C$ were dependent on C, then we would have $f(s_j) = 0$ for all $f \in \Gamma \times A$, i.e., $\Gamma \times A$ would not be full. We deduce from this that $B - C = B - \overline{C}$. Let $U \subseteq \mathcal{V}_m$ be a subspace of dimension $m - r(C) = m - r(\overline{C})$

with $U \bigcap \overline{C} = \{0\}$. Each point $b \in B - C$ determines a subspace $\overline{C \bigcup \{b\}}$ of rank $r(C) + 1$, and thus a unique point $b' = \overline{C \bigcup \{b\}} \bigcap U$. But the configuration $B' = \{b' : b \in B - C\}$ is a block corresponding to the group $\Gamma \times A$. For, $b'_1, \ldots, b'_h$ in B' are linearly independent points if and only if $\{b_1, \ldots, b_h\} \subseteq B$ is independent of C. However, this was exactly the concept of independence in the matroid $\mathcal{M}(\Gamma) \times A$ (see the remark before 8.13). Hence, if W is a subspace of rank $m - 2$ that contains C, then, by 8.17, we must have $W \bigcap B' \neq \emptyset$. If $b' \in W \bigcap B'$, then we have $W \supseteq \overline{C \bigcup \{b'\}} = \overline{C \bigcup \{b\}}$, $b \in B - C$, therefore W contains a point $b \in B - C = B - \overline{C}$. The inference just carried out can easily be reversed. ■

With 8.17 and 8.18 we have completed the geometric characterization of irreducible chain groups. Let B be a 2-block and $\emptyset \neq C \subseteq B$. A **tangent** of C in B is a subspace of rank $m - 2$ that contains all of C (and thus $\overline{C}$), but no other point $b \in B - \overline{C}$. The name tangent is clear: If $C = \{b\}$ is a single point then a tangent of b is a subspace U with $r(U) = m - 2$, that intersects B *exactly in* b. If each nonempty subset $C \subseteq B$ with $r(C) \leq m - 2$ has a tangent, then we call B a **tangential 2-block**. Every tangential 2-block is minimal. For suppose there were a point $b \in B$ whose complementary set $B - \{b\}$ were a 2-block, then by definition every subspace of rank $m - 2$, and thus also every tangent of b, would have to intersect the set $B - \{b\}$.

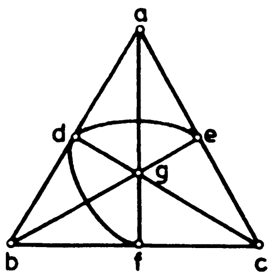
As a summary of our analysis, we formulate the following main result:

8.19 THEOREM (Tutte). *A chain group Γ is irreducible if and only if every corresponding block (i.e., every geometric realization) is a tangential 2-block.*

Hence, we now face the geometric problem: *Find all tangential 2-blocks!*

As a start, we will determine all tangential 2-blocks B with $r(B) \leq 4$. From 8.17 we know that a 2-block has rank at least 3. For rank 3, resp. geometric dimension 2, the situation is trivial. The full projective plane PG_2 (i.e., $\mathcal{V}_3 - \{0\}$) is a tangential 2-block (each point is its own tangent), and this is also the only one since every proper subset $B \subseteq PG_2$ misses at least one point, i.e., a subspace of rank $r(B) - 2 = 1$ and thus can not be a 2-block.

In geometry, one calls PG_2 the Fano plane, so that in our terminology we will call PG_2 the **Fano block**, F. A planar realization is given in Figure 8.8. F has 7 points and 7 three-point lines, of which one is drawn curved.



$$\begin{array}{c} \begin{array}{ccccccc} a & b & c & d & e & f & g \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

Figure 8.8

We obtain a coordinatization matrix of the corresponding matroid $\mathcal{M}(S)$ by writing all vectors $\neq 0$ of length 3 side by side. Is $\mathcal{M}(S)$ graphic? Assume there is a graph $G(V,S)$, $S = \{a, b, \dots, g\}$ with $\mathcal{P}(G) \cong \mathcal{M}(S)$. Since any two elements from S are contained in a circuit, G must be connected, whence by $r(M) = 3$ and $r(\mathcal{P}(G)) = |V| - c(G)$ it follows that $|V| = 4$. But of course every graph with 4 vertices is 4-colorable, in contradiction to the fact that the Fano block is tangential. We can, as easily, convince ourselves that $\mathcal{M}(S)$ can not be cographic.

Now we proceed to rank 4, resp. geometric dimension 3. Let B be a tangential 2-block of rank 4. The full space $PG_3 = \mathcal{V}_4 - \{0\}$ contains $2^4 - 1 = 15$ points. We already know *one* such tangential 2-block, Desargues' Block D in Figure 8.7. We will show that D is the only one (up to isomorphism). It is easiest to describe D in terms of the complementary set $D' = PG_3 - D$. Let us inspect Figure 8.6: The columns $\neq 0$ missing there are

$$D' = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The corresponding 5 points P_1, P_2, P_3, P_4, P_5 have the obvious property that each 4 are linearly independent. We say that $P_1, \dots, P_5$ are in **general position**. Furthermore, we see that D consists of exactly the points $P_i + P_j$, $1 \leq i < j \leq 5$. Conversely, if 5 points $P_1, \dots, P_5$ are given in general position, then by a suitable coordinate tranformation they can be brought to the above form D' so that the complementary set is precisely the Desargues Block D in Figure 8.7. We can also see this directly. Let $D = PG_3 - \{P_1, \dots, P_5\} = \{P_i + P_j : 1 \leq i < j \leq 5\}$. We define $\{P_1, \dots, P_5\}$ to be the vertices of a graph and declare $P_i + P_j$ to be the edge from P_i to P_j so that K_5 results. A point set $\{P_i + P_j, P_k + P_l, \dots\}$ of D is minimally dependent if and only if each P_i that occurs at all occurs exactly twice, therefore if and only if the corresponding edge set is a polygon in K_5.

Conversely, let B be an arbitrary tangential 2-block with $r(B) = 4$. Since B is a minimal 2-block, B can not contain a Fano block, i.e., no plane in PG_3. Hence the complementary set $B' = PG_3 - B$ must intersect each plane nontrivially. It remains to show that B' contains five points in general position, for then, by minimality, B must be a Desargues Block. Since B as a 2-block intersects every line, B' can not contain an entire line. This means in particular: If $P, Q \in B'$, then we must have $P + Q \in B$. Conversely, since B' intersects every plane, B' must contain at least 3 independent points P_1, P_2, P_3, since for P_1 and P_2 there exists a plane E with $E \bigcap \overline{P_1P_2} = P_1 + P_2 \in B$. All 3 points $P_1 + P_2$, $P_1 + P_3$, $P_2 + P_3$ are in B and form a line. This line is contained, together with $\overline{P_1P_2P_3}$, in two other planes, therefore there must be a $P_4 \in B'$ that is independent of P_1, P_2, P_3, i.e., $\{P_1, P_2, P_3, P_4\} \subseteq B'$ is independent. The 6 points $P_i + P_j$, $1 \leq i < j \leq 4$, and $P_5 = P_1 + P_2 + P_3 + P_4$ determine a plane. Since all points $P_i + P_j$ are in B, we must have $P_5 \in B'$, and we are finished since the 5 points $P_1, \ldots, P_5 \in B'$ are in general position.

Already the case rank $= 4$ indicates that the finding or exclusion of tangential 2-blocks is not an easy task. Tutte was able to show further that there is no rank 5 tangential 2-block and that the only such block of rank 6 is the **Petersen block** P, corresponding to the bond matroid of the Petersen graph. It is known by now that there also are no examples for rank 7, and perhaps one can verify some day Tutte's boldest conjecture: *The Fano, Desargues and Petersen blocks are the only tangential 2-blocks.*

This last conjecture encompasses Hadwiger's conjecture and thus the 4-color conjecture. The extremely tedious handling of even the cases with rank only 5, 6 and 7 indicates that no advance will be made here without fundamentally new techniques.

We have arrived at the end of the last big theoretical idea for solving the 4-color problem and chronologically we have ploughed forward to the middle of the 60's. In the roughly 30 years since the theorems of Menger and Kuratowski and the book by König, graph theory took an enormous upswing, its results were applied in many different areas—in short, it became an established discipline within discrete mathematics.

And the 4-color problem? It was as unsolved as ever. In every attack one was stopped one step before the goal. Whether this was Heawood's 5-color theorem, or Vizing's result on the edge colorability of 3-regular graphs, or Tutte's result on the existence of Hamiltonian circuits in 4-connected plane graphs—the last decisive breakthrough was not achieved. The success of graph theory as a whole on the one hand, and the apparent impossibility of deciding the

4-color conjecture on the other hand, were perhaps the basic reason why interest in the 4-color problem diminished significantly in the 50's and 60's. Perhaps it was really only a topological curiosity after all, as it was thought in earlier times, and possibly it was not decidable. In retrospect, it is perhaps only natural that a renewed vitalization with new impulses, and advances leading towards the final solution occurred where one would have least guessed it—namely in the original program of Kempe: Find an unavoidable set of reducible configurations. And so we will open the last two chapters of the 4-color problem.

EXERCISES FOR CHAPTER 8

1. Complete the proof of 8.1.

2. Let S be a finite set and t a natural number. A family $\mathcal{H} \subseteq 2^S$, $|\mathcal{H}| \geq 2$, is called a t-partition of S if i) $A \in \mathcal{H}$ implies $|A| \geq t$, ii) every t-subset of S is contained in exactly one set from $\mathcal{H}$. Show that $\mathcal{I} = \{I \subseteq S : |I| \leq t\} \bigcup \{J \subseteq S : |J| = t+1, J \not\subseteq A \text{ for } A \in \mathcal{H}\}$ defines a matroid on S. Determine the rank $r(A)$ for $A \subseteq S$ and give examples of 2- and 3-partitions.

3°. Construct the smallest nongraphic matroid (on 4 elements).

4. Prove 8.2.

5. Computer problem. Design programs to list the bases and circuits of a matroid given by the family $\mathcal{I}$ of independent sets.

6. Let $\mathcal{M}(S)$ be the matroid on S with the family $\mathcal{I} = \{A \subseteq S : |A| \leq k\}$ as independent sets. What does $\mathcal{M}^*(S)$ look like?

7*. Let S be a finite set and $r : 2^S \to \mathbb{N}_0$ a function with the following properties: i) $A \subseteq B \Rightarrow r(A) \leq r(B)$, ii) $r(A \bigcap B) + r(A \bigcup B) \leq r(A) + r(B)$, iii) $0 \leq r(A) \leq |A|$ for all $A, B \subseteq S$. Show that there is a unique matroid on S that has r as rank function. With this complete the proof of 8.5. (Hint: Define I to be independent if and only if $r(I) = |I|$.)

8°. Let r be the rank function of $\mathcal{M}(S)$ and r^* the rank function of the dual matroid $\mathcal{M}^*(S)$. Show that $r^*(A) = |A| - r(S) + r(S - A)$ holds for all $A \subseteq S$.

9*. Prove: A matroid $\mathcal{M}$ is binary if and only if $|C \bigcap D|$ is even for all circuits C in $\mathcal{M}$ and circuits D in $\mathcal{M}^*$.

10. Show that every matroid on at most 5 elements is isomorphic to a transversal matroid. Does this still hold for 6 elements?

11°*. Let $(S, \mathcal{I})$ be a matroid with rank function r and let $A = \{A_1, \ldots, A_m\}$ be a set system on S. We call a partial transversal T independent if $T \in \mathcal{I}$. Show the following generalization of Hall's Theorem: A has an independent transversal if and only if $r(\bigcup_{i \in I} A_i) \geq |I|$ holds for all $I \subseteq \{1, \ldots, m\}$. In what sense is this a generalization of 6.5? (Hint: Consider a submodular function as in 8.5.)

12°. Let $\mathcal{M}$ be the binary matroid on the set $\mathcal{V}_n$ of all 0,1-vectors. Determine $\pi(\mathcal{M})$ and $\beta(\mathcal{M})$.

13. Show that in the case of graphs, 8.7(ii) reduces to 6.19.

14. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two matroids on S with rank functions r_1 and r_2. Show: There is a t-set $B \subseteq S$ which is independent in $\mathcal{M}_1$ as well as in $\mathcal{M}_2$ if and only if $r_1(A) + r_2(S - A) \geq t$ holds for all $A \subseteq S$.

15. Deduce from the preceding exercise a theorem on common transversals of two set systems $(S; \mathcal{A})$ and $(S; \mathcal{B})$.

16. Repeat, in detail, the construction of the contraction $G \times A$ of a graph.

17. Show: Every restriction of a transversal matroid is again a transversal matroid, but this in not true for contractions.

18. Describe the circuits and bases in a contraction $\mathcal{M} \times A$.

19°. Show that the Fano block is not cographic.

PART III: FINALE

9. BACK TO THE BEGINNING

We again take up Kempe's ideas on the solution of the 4-color problem, and primarily their exposition (in the dual form, to be precise) by Birkhoff (Chapter 3). We know, from Heawood's 5-color theorem 2.1, that every plane graph G has chromatic number $\chi(G) \leq 5$. If the 4-color conjecture is false, then there must be 5-chromatic plane graphs. It was Kempe's idea to use an induction proof to study the minimal graphs among these, with the intent of showing their non-existence by contradiction, and thus proving the 4-CC. For that, we begin with the following definition, where it is clear that we can restrict ourselves to simple graphs.

DEFINITION. *An* **irreducible graph**, *G, is a simple 5-chromatic plane graph with a minimal number of vertices.*

Now, how will the contradiction be attained? In two steps: First, by setting up a list of the configurations that can not occur in an irreducible graph—these configurations are called **reducible**. And, secondly, by verification that certain sets of reducible configurations are **unavoidable** in the sense that at least one of the configurations must be contained in each irreducible graph.

The simplest unavoidable set is the set dual to that in Figure 1.14. Every plane graph must contain a vertex of degree ≤ 5, and we obtain the unavoidable set depicted in Figure 9.1.

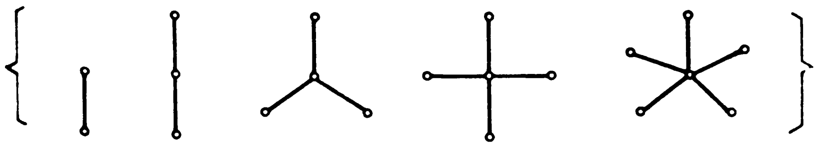


Figure 9.1

As we have already seen in Chapter 1, Kempe verified the reducibility of the first 4 configurations in Figure 9.1. However, the corresponding proof for the vertex of degree 5 contained an irreparable error. Despite Kempe's error, the program, or better said the hope, was to find more and more reducible configurations, until finally they in toto would yield an unavoidable set. In many papers, starting with Birkhoff, 1913, the list of reducible configurations was increased little by little. Yet, 50 years later this was still far from forming an unavoidable set. Indeed, most mathematicians participating in this research were primarily interested in finding a lower bound for the vertex number of an irreducible graph, the so-called **Birkhoff number**, b. If one could show that every plane graph with n vertices must contain at least one of the known reducible configurations, then it would follow that $b \geq n+1$. The larger the list of reducible configurations grew, the higher the Birkhoff number rose.

From 1913 until 1950, the Birkhoff number was raised from the trivial bound $b \geq 12$ to $b \geq 36$ (we shall become acquainted with several of these methods, some exceptionally clever, to verify reducibility). However in the final analysis, this slow advance was cause for pessimism. It appears that Heinrich Heesch from Hanover University, was the first to pursue the original Kempe program with renewed vigor. He expressed not only his conviction that an unavoidable set of reducible configurations exists, but also estimated concretely that one should be able to find an unavoidable set of several thousand configurations, which all have a size that is bounded by a definite quantity (which we shall explain more fully below). This was far too much to verify with pencil and paper, but in the 60's the first generation of high-speed computers was available, so that Heesch was able to test his ideas and hypotheses using computers. Heesch formulated path-breaking ideas on the reducibility question as well as on the unavoidability problem. His suggestions were taken up by Appel and Haken and finally led to success. Before going more fully into these ideas, we briefly sketch the historical development.

We summarize: Two problems must be settled.

1. *How does one prove the reducibility of a given configuration?*
2. *How does one generate unavoidable sets?*

In the early years, it was almost exclusively Question 1) that was studied, so we will also begin with it. Let us first collect a few properties that an irreducible graph must possess in any case. It follows from the Jordan curve theorem that every circuit C of a plane graph G decomposes the plane into an interior region and an exterior region. If each of these two regions contains at least one vertex, then we call C a **separating circuit**. Therefore the non-separating circuits are just the boundary circuits of the countries of G.

9.1 THEOREM. *Let $G(V,E)$ be an irreducible graph with p vertices and q edges. Then:*

i) *G is a maximal plane graph, i.e. a simple triangulation.*

ii) $\sum_{v\in V}(6-d(v)) = 12, \quad q = 3p-6.$

iii) *Every vertex v has degree $d(v) \geq 5$.*

iv) *Every separating circuit C has length $l(C) \geq 5$. Hence in particular, G is 5-connected.*

Proof. Assume that G is not a triangulation, and F is a country with boundary vertices $v_1, v_2, \ldots, v_t$, $t \geq 4$. Then there are (see the remark before 7.2) two nonadjacent vertices v_i and v_j that are not joined outside of F. We contract v_i and v_j inside of F to a single vertex (without violating the plane realization). The new graph G' has one vertex less than G and is thus 4-colorable. However, we can transfer this 4-coloring immediately to G by painting v_i and v_j with the same color, which contradicts the irreducibility of G. Condition ii) now follows from 7.2, and statement iii) is clear—the reducibility of vertices v with $d(v) \leq 4$ was Kempe's actual starting point. It remains to verify iv) (which in any case also includes iii)). Let C be a separating circuit with the vertices $v_1, v_2, \ldots, v_t$ in that order. We denote by A the subgraph generated by the interior of C together with C, and by B the subgraph generated by the exterior of C together with C. Since G is simple, we must have $t \geq 3$. The two proper subgraphs A and B are 4-colorable. If $t = 3$, i.e., $C = K_3$, then by a permutation of the colors, we can attain the situation where the colorings of A and B coincide on C, whence the entire graph G could be colored by 4 colors. It remains to consider the case $C = (v_1, v_2, v_3, v_4)$. The graph $A \bigcup v_1v_3$ is 4-colorable, as is $B \bigcup v_1v_3$. Let f be a 4-coloring of $A + v_1v_3$ and let g be a 4-coloring of $B + v_1v_3$. Then by a permutation we can again assume that $f(v_i) = g(v_i)$ for $i = 1, 2, 3$. Suppose $f(v_i) = g(v_i) = a_i$, $i = 1, 2, 3$, where a_1, a_2, a_3 are distinct, since v_1, v_2, v_3 form a triangle in $A \bigcup v_1v_3$ and $B \bigcup v_1v_3$. Now we remove again the edge v_1v_3 and hence have colored A and B with 4 colors that coincide on v_1, v_2, v_3. If $f(v_4) = g(v_4)$ also, then we are finished. Otherwise, without loss of generality, we can assume that $f(v_4) = a_4$ and $g(v_4) = a_2$. Let us consider A. If there is no chain that is alternately colored by a_2, a_4, (in brief, an a_2, a_4-chain) from v_2 to v_4 in A, then, as in Kempe's proof, we can interchange the colors of the a_2, a_4-component of v_4 in A so that v_4 obtains the color a_2 in A as well as in B. Hence we assume there does exist an a_2, a_4-chain from v_2 to v_4 in A (see Figure 9.2).

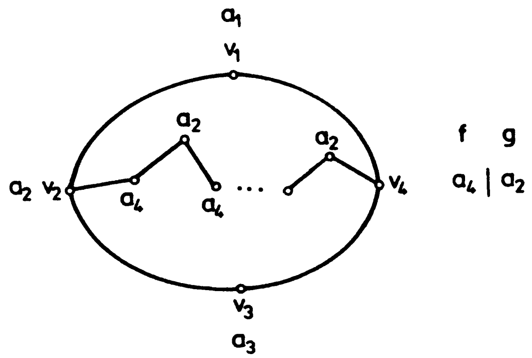


Figure 9.2

Then A contains no a_1, a_3-chain from v_1 to v_3, so we can change the coloring f of A in such a way that $f'(v_1) = f'(v_3) = a_1$ (where we still have $f'(v_2) = a_2$, and $f'(v_4) = a_4$). Now since $B + v_2v_4$ is also 4-colorable, there is a 4-coloring g' of B which assumes different values on v_1, v_2 and v_4, where by a permutation of the colors, we may assume that $g'(v_1) = a_1$, $g'(v_2) = a_2$ and $g'(v_4) = a_4$. Now if $g'(v_3) = a_3$, then g' coincides with f on C. If, on the other hand, $g'(v_3) = a_1$, then g' coincides with f' on C. Thus a 4-coloring of G is given for all cases, which contradicts the irreducibility of G. The final assertion about the 5-connectedness follows from 7.2(v). ■

Now, how does one recognize the reducibility of a given configuration? First: What is a *configuration*? In principle, this is any possible subgraph. In practice it has proven advantageous to understand by a **configuration** K a separating circuit C together with the interior H of C—and we shall also take this as our basic definition. We write $K = (H, C)$ and call H the **interior** of the configuration K and C the **ring**. The length of C is the **size of the ring**. Figure 9.3 shows three configurations with ring sizes 5, 6 and 8, where we draw the interior edges (that is, the edges of H) and the ring edges with solid lines and the edges leading from the interior to the ring with dashed lines.

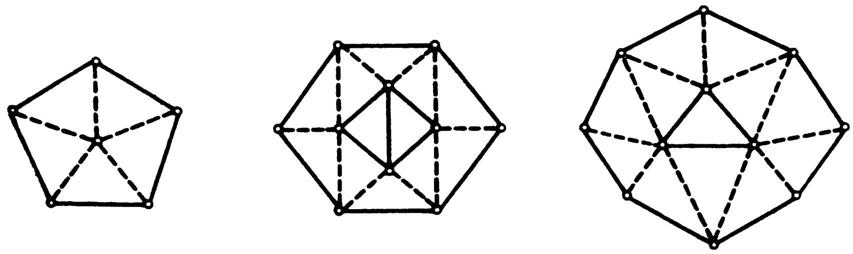


Figure 9.3

Now, verification of the reducibility of a given configuration follows essentially the pattern of the proof of 9.1. We separate from the configuration the interior H along with the edges incident to it, thus obtaining a hole in the graph G. The hole is now filled in any way with a new plane graph H', where we require as the only condition that H' contains fewer vertices than H. Because of the irreducibility of G, G' is 4-colorable. If we are now able to manipulate the 4-coloring of G' in such a way that a 4-coloring of G is obtained, then K must have been reducible!

Using the definition of a configuration, 9.1(iv) can be expressed as follows: Every configuration with ring size ≤ 4 is reducible. What is the situation for size 5? Here irreducible configurations certainly arise, e.g., the first one in Figure 9.3, in which the interior consists of a single vertex of degree 5. As the first significant advance after Kempe, Birkhoff showed in 1913 that in essence this is the only possible case.

9.2 THEOREM (Birkhoff). *Let C be a separating circuit of length 5 in the irreducible graph G. Then either the interior or the exterior of C consists of a single vertex.*

Proof. We assume to the contrary, that the interior A and the exterior B of C contains at least two vertices. If we replace B by a single vertex b which we join to all vertices of C, then the resulting graph is 4-colorable, where the vertices of C , since they are all adjacent to b, are colored with 3 colors. That means then: There is a 4-coloring f_A of $A \bigcup C$ whixh assigns to the vertices of C three colors, and an analogous coloring f_B of $B \bigcup C$. Let the vertices of C be numbered cyclically, $C = (v_1, v_2, \dots, v_5)$, and let the color set be $(\alpha, \beta, \gamma, \delta)$. Clearly, for each 3-coloring f of a 5-circuit there are two pairs of vertices that are colored the same, and one vertex v colored differently. We shall call v the **marked vertex** of f, $v = m(f)$. By a suitable numbering of the v_i's, and a permutation of the colors, we can assume that v_1 is the marked vertex of f_A with the color α, and furthermore that $f_A(v_2) = f_A(v_4) = \beta$ and $f_A(v_3) = f_A(v_5) = \gamma$. Similarly, by a permutation of the colors in $B \bigcup C$ we can attain the situation that the marked vertex of f_B is also colored with α and the two pairs with β and γ respectively. If $m(f_B) = v_1$, then it is clear that, after a possible exchange of β and γ in f_B, the colorings coincide on C, so that we are finished.

If $m(f_B) \neq v_1$, then two possibilities arise, depending on whether $m(f_B)$ is adjacent to v_1 or not. By a possible change of the cyclic order on C and an interchange of β and γ in f_B we obtain, without loss of generality, the possibilities shown in Figure 9.4.

For brevity, we write $f_A = (\alpha, \beta, \gamma, \beta, \gamma)$ and correspondingly for f_B. Before discussing the two cases, let us note that by using contractions as in the proof of 9.1 there is a 4-coloring of $A \bigcup C$, resp. $B \bigcup C$, that assigns the same color to a given pair of nonadjacent vertices of C.

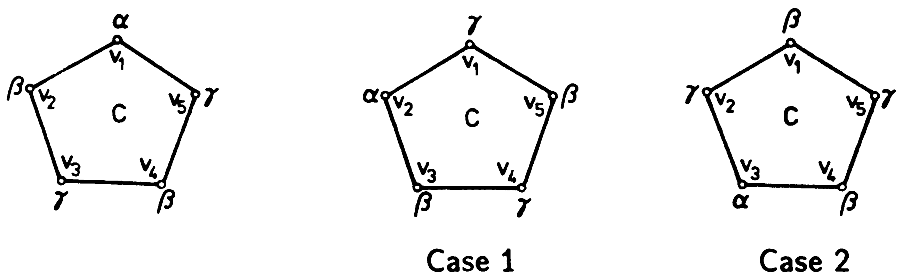


Figure 9.4

Case 1. If there is no α,γ-chain in $B \bigcup C$ from v_2 to v_4, then we can recolor v_4 with α, so that now v_1 is the marked vertex of the new coloring f'_B, and we are finished. In the other case there is no β,δ-chain in $B\bigcup C$ from v_3 to v_5. We change the colors in the β,δ-component of v_3 and obtain a new coloring $f'_B = (\gamma,\alpha,\delta,\gamma,\beta)$. We now color $A\bigcup C$ arbitrarily with 4 colors so that v_1 and v_4 obtain the same color. By a permutation of the colors we can assume that $f'_A = (\gamma,\alpha,\delta,\gamma,z)$ where $z=\alpha$, β, or δ. If $z=\beta$, then f'_A and f'_B coincide on C and we are finished, and, similarly for $z=\delta$, since then C is 3-colored and $m(f'_A)=m(f_B)$. The case $z=\alpha$ remains. Thus $f'_A=(\gamma,\alpha,\delta,\gamma,\alpha)$ with $m(f'_A)=v_3$. In this case we set $f'_B=f_B$. Let us summarize our analysis thus far: From a pair of 3-colorings f_A, f_B of C with $m(f_A)=v_1$ and $m(f_B)=v_2$ we have constructed another pair f'_A and f'_B with $m(f'_A)=v_3$, $m(f'_B)=v_2$. Now, if we interchange the roles of A and B, then we obtain another pair f''_A and f''_B with $m(f''_A)=v_3$ and $m(f''_B)=v_4$ and with a further iteration of this two-fold step, a pair f'''_A and f'''_B with $m(f'''_A)=v_5$ and $m(f'''_B)=v_1$. But now $m(f_A)=m(f'''_B)=v_1$ and we are again finished.

Case 2. Here, $f_B=(\beta,\gamma,\alpha,\beta,\gamma)$. If no α,γ-chain exists from v_3 to v_5 in $B\bigcup C$, then we can recolor v_5 by α, and the new coloring f'_B has the marked vertex $m(f'_B)=v_2$, i.e., we are again in case 1. Otherwise, there is no β,δ-chain from v_1 to v_4 so that, by a recoloring of v_4, we obtain a coloring f'_B with $f'_B=(\beta,\gamma,\alpha,\delta,\gamma)$. As in the preceding case we choose a 4-coloring f'_A of $A\bigcup C$ in which v_2 and v_5 get the same color, where without loss of generality we can assume that $f'_A=(z,\gamma,\alpha,\delta,\gamma)$. If $z=\beta$, then f'_A and f'_B coincide on C. In the cases $z=\alpha$, resp. $z=\delta$, f'_A and f_B have adjacent marked vertices (case 1), resp. f'_A and f_B have the same marked vertex. So in all cases we have verified the 4-colorability of G, in contradiction to the hypothesis. ∎

Looking at this tiresome case analysis, the reader can already imagine that the verification of reducibility for ring sizes ≥ 6 will encounter considerable difficulties. Accordingly, we will derive only one additional result, the reducibility of the so-called **Birkhoff diamond** (the second configuration in Figure 9.3) and otherwise content ourselves with citing the most important later results.

9.3 THEOREM (Birkhoff). *The Birkhoff diamond is reducible.*

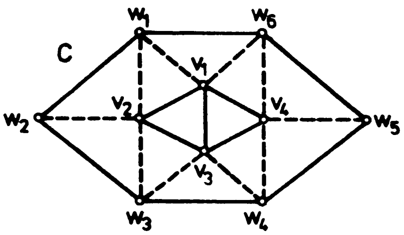


Figure 9.5

Proof. We rub out the interior together with the incident edges. The vertices w_4 and w_6 can not be adjacent (outside C), since otherwise we would obtain a separating circuit of length 3. We now modify the interior by contracting w_4 and w_6 and joining this vertex with w_2. The new graph G' is 4-colorable. Again calling the colors $\alpha, \beta, \gamma, \delta$, we obtain the situation of Figure 9.6. Let us show that in every case we can carry this 4-coloring over to the interior.

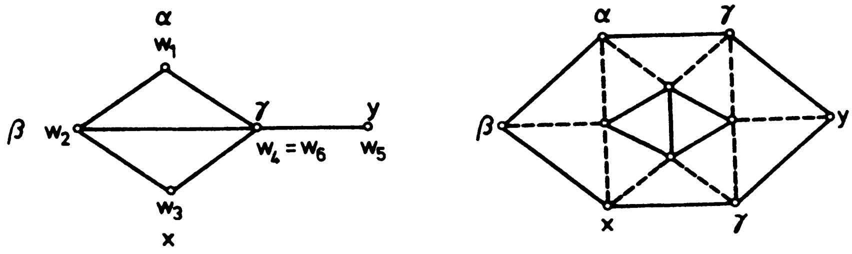


Figure 9.6

G' G

$x = \delta$ or α
$y = \alpha, \beta$ or δ

Case 1. $x = \delta$. Then we must color v_2 with γ. The following list completes this 4-coloring for every possible y-value.

y	v_1	v_3	v_4
α	β	α	δ
β	β	α	δ
δ	δ	β	α

Case 2. $x = \alpha$. Again, using the following table, we can extend the coloring to the interior when $y = \beta$ or δ.

y	v_1	v_2	v_3	v_4
β	δ	γ	β	α
δ	δ	γ	β	α

For $y = \alpha$ this no longer works. The vertices v_1 and v_3 must be colored with β and δ, since they are both adjacent to α, γ-vertices. Hence v_4 would be adjacent to vertices of all 4 colors.

But, the following Kempe argument overcomes this difficulty. Assume there is no α,δ-chain in $G - \{v_1, v_2, v_3, v_4\}$ from w_5 to w_3. If w_1 is in the α,δ-component of w_5, then we recolor w_3 by δ, otherwise we recolor w_5 with δ and we are back to the cases $x = \delta$ and $y = \alpha$, resp.

$x = \alpha$ and $y = \delta$, that were previously considered. However, if there is an α,δ-chain from w_5 to w_3, then there can not be a β,γ-chain from w_4 to w_2 nor a β,γ-chain from w_4 to w_6. We recolor w_4 with β and extend this coloring to the interior as in Figure 9.7. ■

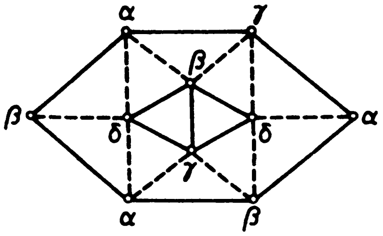


Figure 9.7

The terminology proposed by Heesch is useful for the following results: An **n-vertex** is a vertex of degree n. **Small vertices** are vertices of degree 5 or 6. **Large vertices** are the remaining ones. The symbols for vertices are listed in Figure 9.8:

Figure 9.8

For brevity we also omit the rings, so that for example the three configurations in Figure 9.3 now appear as follows (Figure 9.9):

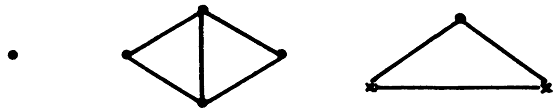
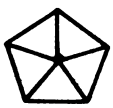
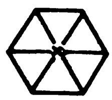


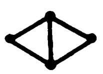
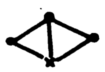
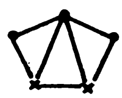
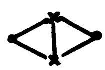
Figure 9.9

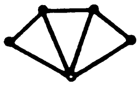
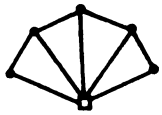
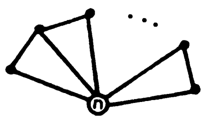
The ring size is then always uniquely determined.

The following theorem collects several of the earliest reducible configurations discovered.

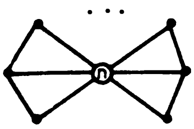
9.4 THEOREM. *The following configurations are reducible.*

i)  *and*  *(all exterior vertices are small)*

ii)  ,  ,  , 

iii)   *in general* 

with $n-3$ *successive* 5-*vertices,* $n \leq 13$.

iv) 

with $n-2$ *successive* 5-*vertices,* $n \geq 5$ *arbitrary.*

As previously mentioned, Heesch (who also found a whole series of reducible configurations), contributed two entirely new and decisive ideas in the 60's. For one thing, he analyzed more carefully than previously, the technique of reduction and, furthermore, he postulated heuristic principles on when a given configuration was probably reducible and when not.

About the first idea: Let $K = (H, C)$ be a configuration with interior H and ring C. Let us denote by J the set of all 4-colorings of C (where we naturally consider only the color schemes induced by permutations of the colors used), and by $J(H)$, the set of those 4-colorings J that can be extended from C *directly* to the interior H. It would of course be ideal if $J(H) = J$, for then (by induction) the 4-coloring of $G' = G - H$ could be extended directly to H, no matter what it may look like. However, that means that G is 4-colorable and hence that K is reducible. Unfortunately, however, $J(H)$ is a proper subset of J in most cases. If this occurs, then we apply Kempe's procedure. Let $\overline{J}(H)$ be the set of all colorings of C that arise from the colorings in $J(H)$ by interchanging the colors in one or several 2-color components. Now if $\overline{J}(H) = J$ then Heesch calls the configuration **D-reducible**.

If $\overline{J}(H) \subsetneq J$, then we can try a so-called **reductor** H' as in Theorems 9.1 through 9.3. That is, we replace H by H', where possibly we add edges to the interior, or contract vertices of C. Anything goes, so long as the exterior of C remains unchanged and the new subgraph $K' = (H', C')$ (that need no longer be a configuration in the previous sense) contains fewer vertices than K. Again let $J(H')$ be the set of 4-colorings of C that are compatible with H' (and C'). If $J(H') \subseteq J(H)$, then the 4-coloring, that exists by induction, of the new graph G' (with K' instead of K) is again directly extendible to H. If, at any rate, $J(H') \subseteq \overline{J}(H)$ holds, then in any case a Kempe exchange leads to success, and in this situation we say that

$K(H,C)$ is **C-reducible**. In our example of the Birkhoff diamond, $K' = (H', C')$ is the left graph in Figure 9.6. As we have seen, all colorings from $J(H')$ are also in $J(H)$ with the exception of the last coloring $(\alpha,\beta,\alpha,\gamma,\alpha,\gamma)$ which is in $\overline{J}(H)$. Therefore, D-reducibility is stronger than C-reducibility, just take the special case where the empty set leads to success as a reductor. The reader can easily convince himself that the Birkhoff diamond is even D-reducible. However, not every C-reducible configuration is also D-reducible. The smallest example for this is the last configuration from 9.4(ii).

Now Heesch suggested calculating the set $\overline{J}(H)$ *before* one breaks one's head over finding a reductor suited for the configuration. Indeed, testing small configurations it very often occurred that $\overline{J}(H) = J$, therefore $K = (H,C)$ is D-reducible, which naturally makes the search for a reductor superfluous. In general, C-reducibility appears to enlarge the class of reducible configurations only slightly as compared to D-reducibility. Hence all computer programs aim chiefly at D-reducibility. We shall return to this in detail in the next chapter.

At this point, at the very latest, the question of computability arises. A ring of size 13 already has 66430 different 4-colorings, and Heesch conjectured that, possibly, one would have to go to ring size 18 in order to construct an unavoidable set. For high-speed computers, numbers of this size do not present an insurmountable obstacle, especially since the generation of $\overline{J}(H)$ is easy to program. Therefore this idea is applicable so long as $\overline{J}(H) = J$, i.e., when the configuration is D-reducible. However, if this is not the case, then the amount of computation needed (search for H') will largely depend on the size and the inner complexity of H.

Heesch's second remarkable suggestion was made about the *structure* of a configuration. Until the 60's the list of reducible configurations increased into the hundreds, but so did the list of those types that resisted all attempts at reduction. Heesch noted that in most of these "non-reducible" configurations, one of three subconfigurations appeared, which he appropriately called **obstructions**. For this we need a couple of concepts: Let $K = (H,C)$ be a configuration with interior H and ring C. We call the edges that lead from an interior vertex v to the ring the **legs** of v. An interior vertex v is called an **articulation**, if $H - v$ decomposes, or equivalently, if not all legs of v follow each other successively on the ring.

Now the obstructions are:

(A) An interior vertex with at least 4 legs.

(B) An articulation with at least 3 legs.

(C) Two adjacent interior 5-vertices that are adjacent exactly to one additional interior vertex, and indeed to the same. (We call this a hanging 5-pair.)

Figure 9.10 gives examples for each of the 3 types, where the obstructions are circled.

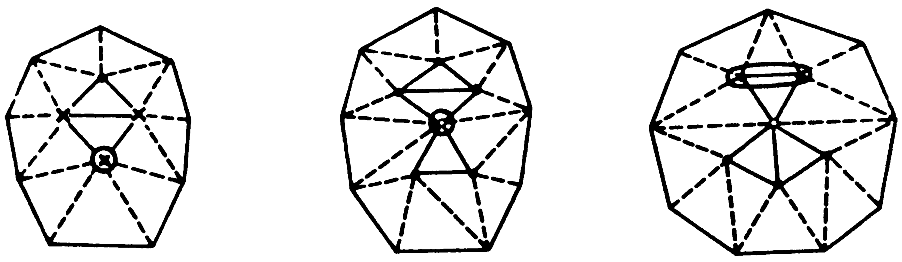


Figure 9.10

(A) (B) (C)

Until now no configuration that contains one of the obstructions was shown to be reducible, and it appears plausible that this will remain the case as long as one uses only the usual Kempe chain methods. In the other direction things appear less clear, since there is a whole series of configurations that contain no obstructions but have until now resisted every attempt at reduction. However, these types already have higher ring sizes. Up to ring size $n \leq 9$ there are no exceptions, for $n = 10$ there is one and for $n = 11$ there are three exceptions.

Heesch's idea of obstructions turned out to be indispensible as the leading principle: *In the search for an unavoidable set avoid those configurations that contain obstructions!*

This brings us to the second theme: Unavoidability, with the central question: How does one prove the unavoidability of a given set? We already know a simple example, the set $\{\bullet\}$ consisting of a 5-vertex. Indeed, we know from 9.1(ii) that an irreducible graph must contain at least 12 such 5-vertices. In general, formula 9.1(ii) forms the starting point for our considerations, which in its basic outline also goes back to Heesch.

Let G be an irreducible graph. We assign to each vertex v the **initial load** $a(v) = 6 - d(v)$ so that by 9.1(ii) the sum of the loads is $\sum_{v \in V} a(v) = 12$. Thus 5-vertices are assigned the initial load 1, 6-vertices the load 0, 7-vertices the load -1, and so forth. We recall, that 5- resp. 6-vertices were called *small* vertices, k-vertices with $k \geq 7$ *large* vertices. Therefore it is exactly the large vertices that have negative initial load.

Now we define a rule, by which we transfer a load in G without changing the total sum 12—we call this an **unloading algorithm**. In this way the **terminal load**, e, again satisfies $\sum_{v \in V} e(v) = 12$. According to the set-up of the terminal load, we obtain various necessary conditions on G that we can conveniently divide into three groups.

1. Since the total sum is > 0, there must be (finitely many) local places where a **positive terminal load** occurs—these local configurations form an **unavoidable set**. Or, turned around: If none of these configurations were present in G then we would have $\sum e(v) \leq 0$, which can not happen.
2. If we can show that $e(v) \leq s \leq 1$ holds at each vertex $v \in V$, then it follows that $12 = \sum_{v \in V} e(v) \leq |V|s$, and we obtain a bound for the Birkhoff number $b \geq 12s^{-1}$.
3. Let p_k be the number of k-vertices, $k \geq 5$. Then for the p_k's, 9.1(ii) implies $\sum_{k \geq 5}(6 - k)p_k = 12$. We try to arrange the terminal load so that $e(v) = 0$ for all 5-vertices v and $e(v) \leq f_k$ for $d(v) = k$, $k \geq 6$. From the relation $\sum_{k \geq 6} f_k p_k \geq 12$ we can derive inequalities between the p_k's.

The following examples should clarify these three viewpoints. The simplest unloading algorithm is the one where we do nothing at all, therefore we set $e = a$. Thus we obtain the unavoidable set $\{\bullet\}$ and the bound $b \geq 12$ in 2). Since until now $\{\bullet\}$ has not been reduced (and in the present state of reduction techniques we cannot expect that this will change) it is reasonable to replace $\{\bullet\}$ by a larger unavoidable set in the hope of coming closer to reducibility.

As a first attempt we stipulate that every 5-vertex is unloaded uniformly with $1/5$ placed on each of the adjacent large vertices (if there are any). What does the terminal load e look like? For the sake of simplicity we set v_k for a k-vertex, $k \geq 5$. Then a 5-vertex can only satisfy $e(v_5) > 0$ if not all of its neighbors are large vertices, i.e., if $\bullet\!\!-\!\!\!-\!\!\bullet$ or $\bullet\!\!-\!\!\!-\!\!\times$ occurs. Since 6-vertices are never involved, we always have $e(v_6) = 0$. If $e(v_7) > 0$ for a 7-vertex, then, since $e(v_7) = -1 + m/5$ (m=number of 5-neighbors), v_7 must have at least 6 such neighbors of which obviously two in turn must be adjacent. Thus $\bullet\!\!-\!\!\!-\!\!\bullet$ again results. Finally, for v_k, $k \geq 8$, we have $e(v_k) = (6 - k) + m/5 \leq (6 - k) + k/5 = (30 - 4k)/5 < 0$. We thus deduce:

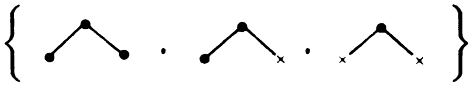
$$\{\bullet\!\!-\!\!\!-\!\!\bullet,\ \bullet\!\!-\!\!\!-\!\!\times\} \text{ is an unavoidable set.}$$

This result was one of the first unavoidability results found (by another, although similar counting method) by Wernicke in 1904.

Let us go a step further. We unload each 5-vertex by transferring $1/4$ to at most 4 adjacent large vertices. If for the terminal load $e(v_5) > 0$, then v_5 has at least two small neighbors, i.e.,

 or  or 

occurs. If $e(v_7) = -1+m/4 > 0$ then it follows that $m \geq 5$. Therefore again  occurs among the 5-neighbors of v_7. For $k \geq 8$ we obtain $e(v_k) = (6-k) + m/4 \leq (6-k) + k/4 = (24-3k)/4 \leq 0$. Conclusion:

 is unavoidable.

This result was essentially found by Franklin in 1922. We note that this unavoidable set is stronger than that of Wernicke.

It should be clear what we will try next. We unload every 5-vertex by transferring 1/3 to at most 3 adjacent large vertices. As in the preceding cases, the corresponding analysis leads to the following result: The set depicted in Figure 9.11 is unavoidable.

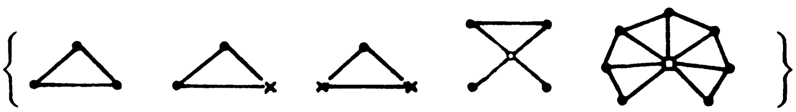


Figure 9.11

By 9.4(iii), the last configuration is reducible so that we can remove it from the list. The reader may determine the corresponding set for a transfer of 1/2. It is clear that the unavoidable sets become more and more complex, but at the same time chances increase that among them we will find ever more reducible configurations. This is precisely the gist of this method: That sooner or later the balance will tip towards the reducibility of *all* configurations of an unavoidable set.

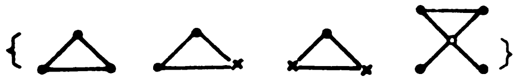
There is, of course, no limit to finding ever more sophisticated unloading algorithms. However, we shall be content with these examples as illustration of point 1). Let us summarize what we have proven so far.

9.5 THEOREM. *The following sets are all unavoidable.*

i) 

ii) 

iii) 

iv) 

Now to the second viewpoint: Bounds for the Birkhoff number. The following unloading algorithm, due to F. Bernhart, offers a beautiful example. We transfer the load from each 5-vertex v_5 as follows: By 9.4(i) we know that v_5 has at least one large neighbor.

a) If v_5 has exactly one large neighbor, then v_5 must have at least two 6-neighbors, since otherwise one of the two first reducible configurations in 9.4(ii) would result. We transfer 28/100 to the large vertex and 12/100 to each of exactly two of the 6-neighbors.

b) If v_5 has at least two large neighbors, then we give 26/100 to exactly two of them.

What is the situation for the terminal load $e(v)$? For v_5 we obtain in case a)

$$e(v_5) = 1 - \frac{28}{100} - 2\frac{12}{100} = \frac{48}{100} = \frac{12}{25}$$

and in case b)

$$e(v_5) = 1 - 2\frac{26}{100} = \frac{12}{25}.$$

Since by 9.4(ii), a 6-vertex v has at most four 5-neighbors we infer that $e(v_6) \leq 0 + 4(12/100) = 12/25$. A 7-vertex v_7 has at most five 5-neighbors, whence $e(v_7) \leq -1 + 5(28/100) = 10/25 < 12/25$. Finally for k-vertices v_k with $k \geq 8$ we have

$$e(v_k) \leq (6-k) + k\frac{28}{100} = 6 - \frac{72}{100}k \leq 6 - \frac{576}{100} = \frac{24}{100} < \frac{12}{25}.$$

Thus we have proven in all cases that $e(v) \leq 12/25$, where strict inequality holds at least once (since, by 9.4(i), large vertices must occur), so that for the Birkhoff number, b, we obtain $b \geq 26$.

This bound was given by Franklin in 1922. It is a hint of the superiority of the unloading technique if one compares the derivation just carried out with the tiresome (if, in principle, equivalent) counting method of Franklin.

The following illustration of the third viewpoint goes back to Chojnacki (1942) and Mayer (again our professor of literature, 1975). If we write the initial equation $\sum_{k\geq 5}(6-k)p_k = 12$ in the form

$$p_5 = 12 + \sum_{k\geq 7}(k-6)p_k,$$

then we obtain a first equation among the numbers p_k.

Now we unload the v_5's in the following way. A 5-vertex has either 3,2,1 or no 5-neighbors. The corresponding unloading procedures are given in Figure 9.12.

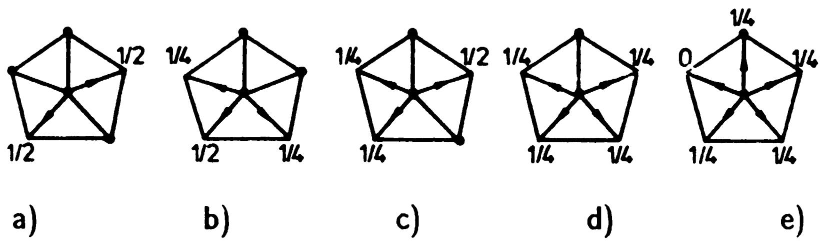


Figure 9.12

The last figure means that the four of the five neighbors that were assigned transfers of 1/4 were chosen arbitrarily.

How much does a k-vertex receive when $k \geq 6$? Let us consider the neighborhood C_k. The 5-vertices on C_k yield paths W_i, $i = 1, \ldots, t$, having $l_i \geq 1$ vertices, which are separated from one another by n-vertices, $n \geq 6$. If $l_i = 1$, then one of the cases (b), (d) , or (e) occurs and the isolated 5-vertex $v = W_i$ contributes at most 1/2. Let $W_i = (x_1, \ldots, x_{l_i})$, $l_i \geq 2$. Each of the interior 5-vertices in W_i transfers 1/2 to v_k (case (a) or (c)), each terminal vertex in W_i transfers 1/4 (case (b), (c), or (d)). Because of 9.4(iv), there must be in C_k at least two vertices that are not 5-vertices. Therefore, in particular, $l_i \leq k-2$, i.e., W_i contains terminal vertices, and we conclude that W_i transfers at most $\frac{1}{2}(l_i - 2) + \frac{1}{4}2 = \frac{1}{2}(l_i - 1)$ to v_k. Altogether this yields $e(v_k) \leq (6-k) + \frac{1}{2}\sum_{i=1}^{t} l_i - \frac{s}{2}$ (s = number of the $l_i \geq 2$). Since, as just mentioned, $\sum_{i=1}^{t} l_i \leq k-2$ holds, it follows that

$$\text{(a)} \qquad e(v_k) \leq (6-k) + \frac{1}{2}(k-3) = \frac{(9-k)}{2} \qquad \text{if } s \geq 1,$$

$$\text{(b)} \qquad e(v_k) \leq (6-k) + \frac{1}{2}\left\lfloor \frac{k}{2} \right\rfloor \leq 6 - \frac{3k}{4} \leq \frac{(9-k)}{2} \qquad \text{if } s = 0.$$

We set $f_k = (9-k)/2$ for $k \geq 14$ and thus always have $e(v_k) \leq f_k$. For $7 \leq k \leq 13$ we can improve this further. From 9.4(iii), we know that at most $k-4$ successive 5-vertices can appear in C_k. It follows immediately from this that for $7 \leq k \leq 13$:

(a′) $$e(v_k) \le (6-k) + \frac{1}{2}(k-4) = \frac{(8-k)}{2} \qquad \text{if } s \ge 1$$

(b′) $$e(v_k) \le (6-k) + \frac{1}{2}\left\lfloor \frac{k}{2} \right\rfloor \le \frac{(8-k)}{2} \qquad \text{if } s = 0.$$

Let us set $f_k = (8-k)/2$ for $7 \le k \le 13$, so, $f_7 = 1/2$, $f_8 = 0$, $f_9 = -1/2, \dots, f_{13} = -5/2$. By (a) and (b), we infer for $k = 6$ that $e(v_6) \le 3/2$ so that we take $f_6 = 3/2$.

In summary we obtain,

$$12 = \sum_{v \in V} e(v) \le \frac{3}{2}p_6 + \frac{1}{2}p_7 - \frac{1}{2}p_9 - \cdots - \frac{5}{2}p_{13} - \sum_{k \ge 14} \frac{(k-9)}{2} p_k,$$

and from this the following (slightly sharpened) inequality due to Chojnacki:

9.6 THEOREM. *Let p_k be the number of k-vertices of an irreducible graph G. Then:*

$$3p_6 + p_7 \ge 24 + p_9 + 2p_{10} + \cdots + 5p_{13} + \sum_{k \ge 14} (k-9)p_k.$$

It follows in particular that $\{\times, \circ\}$ is an unavoidable set. Incidentally, it was shown by Allaire that the 6-vertex alone already is unavoidable.)

By bringing in a series of complicated reducible configurations, Mayer was able to prove the following inequalities:

9.7 THEOREM. *For the numbers p_k the following inequalities hold:*

i) $2p_6 + p_7 \ge 24 + p_9 + \sum_{k \ge 10}(k-9)p_k$

ii) $p_6 + p_7 + p_8 \ge 24$

iii) $p_7 + 2\sum_{k \ge 8} p_k \ge 12$.

If we combine 9.7(iii) with the initial equation $p_5 = 12 + \sum_{k \ge 7}(k-6)p_k$, then we get $p_5 \ge 12 + p_7 + 2\sum_{k \ge 8} p_k \ge 24$ and from this, using 9.7(ii), $p_5 + p_6 + p_7 + p_8 \ge 48$. Thus, $b \ge 48$ holds for the Birkhoff number b. A short time later, Mayer succeeded in raising the Birkhoff number to $b \ge 96$ with the same methods.

Now in our history we have arrived in the 70's. The methods of unloading and reduction were well understood and had reached a high degree of sophistication. And yet, all known reducible configurations were, in toto, far from forming an unavoidable set. Even with very fast computers one stood before two enormous difficulties. First, it was certain that in every unavoidable set of reducible configurations there must occur types with ring size ≥ 14. Heesch and his student Dürre had already used 26 computer hours to test only *one* rather unpleasant configuration, with ring size 14, for D-reducibility. Since the estimated time increase to the next ring size is about a factor of 4, even the currently fastest and largest computers would have to capitulate for ring sizes from 16 upwards. The second difficulty was that nobody knew exactly how many configurations would be necessary. The usual estimate was several thousand—and this appeared to be well beyond the bounds of the available computer capacity.

And finally: A computer based proof is surely not what a mathematician thinks of as an elegant solution. On the other hand, most experts were pessimistic whether a proof of a reasonable length was within reach. Confronted with all these obstacles of mathematical and non-mathematical nature, Wolfgang Haken (originally from Kiel) and his colleague Kenneth Appel (from the University of Illinois) began in 1972 their work on the 4-color problem. The last chapter is devoted, of necessity, to a sketchy discussion of their work.

EXERCISES FOR CHAPTER 9

1°. Show that the Birkhoff diamond is D-reducible.

2. Show the reducibility of the Birkhoff diamond by choosing a different reductor than in 9.3.

3. Verify the reducibility of the second and fourth configurations in 9.4(ii).

4°. Show that the fourth configuration in 9.4(ii) is not D-reducible.

5°. Transfer the initial load of 5-vertices to at most two adjacent large vertices (each $\frac{1}{2}$) and determine the resulting unavoidable set.

6. Try to raise the Birkhoff number by a suitable unloading (as in 9.5) to beyond 26.

10. SOLUTION AND PROBLEM

We have seen in several examples, how the classical methods functioned. First, a list of reducible configurations was set up (reducibility). Then, secondly, it was attempted to show that at least one of the known reducible configurations appeared in a triangulation (unavoidability).

It was, perhaps, already the decisive step when Appel and Haken reversed this classical order. They first developed a comprehensive theory of unavoidability, where they followed Heesch's principle of obstructions, and only then tested for reducibility. Accordingly, we will also consider Appel's and Haken's ideas about these two themes in this order.

Hence, the first goal is determined: Find an unavoidable set of configurations that contain none of the obstructions. If such a set of moderate size existed (as regards the number of configurations as well as their size), then there would be reasonable hope for also finding sets all of whose members are reducible.

In order to gain perspective, Appel and Haken initially restricted themselves to the first two obstructions. They called a configuration, **geographically good**, if it contained neither obstruction (A) nor obstruction (B). We see from the definition, that a configuration is geographically good if every interior vertex has at most 3 legs, and has exactly 3 legs only if the corresponding ring vertices follow each other successively. A single interior 5-vertex has 5 legs and is accordingly geographically bad. Do there exist small unavoidable sets in which interior vertices have at most 4 legs? Yes, for example, the unavoidable set from 9.5(iv). As the next improvement Haken produced, by an ad hoc method, an unavoidable set with 68 configurations that avoided obstructions (A) and (B), with perhaps the exception of a single 4-legged interior vertex. In 1972 he and Appel wrote a computer program for an unloading algorithm that transferred the load from the 5-vertices, as in the preceding chapter, uniformly to the adjacent large vertices. The resulting unavoidable set still did not consist wholly of geographically good configurations, but two important pieces of information were obtained: First, geographically good configurations of moderate size (ring size ≤ 16) were found near most vertices with positive terminal loading. Second, the same configurations appeared relatively frequently, so that there was reason to hope that the final list would be of tolerable length.

With the results of the first computer run, Appel and Haken proceeded to modify their unloading algorithm accordingly. There began a dialogue with the computer that lasted about 6 months until they were sure that the modified program would indeed yield the desired result— an unavoidable set of geographically good configurations. The final proof, with an enormous number of cases to be considered, was a veritable tour de force that occupied them for over a year.

After the proof of the existence of an unavoidable set of geographically good configurations, they came to the next hurdle: How large would the list be and how complicated would the individual configurations be? For this, Appel and Haken tested their program on a special situation: They considered only triangulations in which adjacent 5-vertices did not occur. Clearly this was a strong restiction, but the results were encouraging: They obtained an unavoidable set of 47 configurations and ring size ≤ 16. Further experiments with the computer suggested that by also excluding the third obstruction (i.e., a hanging 5-vertex), the initial list of geographically good configurations would only be doubled and the ring size could perhaps be reduced to ≤ 14.

Only now was the reducibility question tackled. Let us call a configuration **probably reducible**, if it is without obstructions. Our analysis has led us to the following program: First, give an unavoidable set $\mathcal{M}$ of probably reducible configurations. Second, if a configuration K in $\mathcal{M}$ is not reducible, then modify the unloading algorithm in such a way that K is replaced by other configurations that are again without obstructions. Now test these for reducibility, and so forth. But can we be certain that this modification process comes to an end? Do we perhaps produce only larger and larger unavoidable sets with larger and larger ring sizes? The most important aspect of the work by Appel and Haken, beside the sophisticated unloading techniques, is a heuristic plausibility consideration that this case will not occur. We now go into this.

Let us anticipate the final result: In their computer based solution of the 4-color problem Appel and Haken constructed an unavoidable set of 1936 reducible configurations all with ring size $n \leq 14$.

Since to prove reducibility only Kempe's original method of color interchanges was used, this result appears bizarre at first glance. Why must one consider such a large set in order to rescue Kempe's proof, while smaller sets will not do? The following considerations will clarify this predicament. Namely, they will show that it is "extremely" probable that an unavoidable set with reducible configurations of ring size $n \leq 17$ exists and "very" probable that there is such a set with $n \leq 14$ while it is unlikely that such a set exists with $n \leq 12$. For $n = 13$ no prediction can be made with the available methods. The set constructed by Appel and Haken has $n \leq 14$ and thus confirms the positive result. On the other hand, in 1977, Moore found a triangulation that contains no reducible configuration with $n \leq 11$, so that also the negative result is almost reached. The only open cases are thus $n = 12$ and $n = 13$. Therefore, at second glance, the Kempe solution by Appel and Haken is in no way remarkable but on the contrary rather confirms the fact—in the words of Appel and Haken—that there are no surprises in mathematics.

We group our plausibility considerations into two parts: First, we study the probability of the *reducibility* of a configuration and then the probability of the *unavoidability* of a set of configurations.

DEFINITION. *A configuration $K = (H, C)$ with m interior vertices and n ring vertices is called an* **(m,n)-configuration**. *For the sake of clarity, we frequently write $K = (H_m, C_n)$.*

How can we estimate the probability of the reducibility of $K = (H_m, C_n)$? The larger m is with respect to n, the more we can assume that every 4-coloring of C can, at least by a Kempe interchange, be extended to H, and thus that K is D-reducible. This plausibility argument together with an analysis of small configurations lead to the following (unproven) conjecture, whose justification will be amply given in the following.

REDUCIBILITY CONJECTURE. *Every (m, n)-configuration with $m > \dfrac{3n}{2} - 6$ is D-reducible.*

Let us examine a couple of examples. Assume that the interior consists solely of one 5-vertex v. Then $m = 1$, $n = 5$ and the reducibility conjecture, briefly, the R-conjecture, will fail by one unit of m, since $1 < (3/2)5 - 6 < 2$. Next we take as the interior, the **first neighborhood** of v, i.e., v with its 5 neighbors. What ring size n is to be expected? It follows from 9.1(ii) that with $|V| = p$ the average degree of a vertex is $\overline{d} = 6 - (12/p)$. Therefore we only raise the average n, if we assume that all the adjacent vertices have degree 6. The sum of the degrees of the adjacent vertices $w_1, \ldots, w_5$ is then 30. If from each of these we subtract the 3 degrees from the interior, and take into account also the common ring vertices of adjacent w_i's, then we obtain $n \approx 30 - 3 \cdot 5 - 5 = 10$ (Figure 10.1).

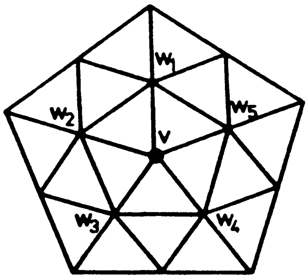


Figure 10.1

With $m = 6$ and $n = 10$, the R-conjecture doesn't apply, since $6 < (3/2)10 - 6 = 9 < 10$; we miss the R-conjecture by 4 units of m. This appears to be hopeless. By enlarging the configuration we are farther removed from reducibility than before. But if we go a step further and take as the interior the **second neighborhood** (i.e., v together with $w_1, \dots, w_5$ and their neighbors), then the situation is turned around again. As we have seen, now $m \approx 16$ and as one confirms by an easy calculation, $n \approx 15$. With $16 < (3/2)15 - 6 < 17$, the R-conjecture is again only a unit away from m. This is an (admittedly very weak) explanation of the phenomenon mentioned at the beginning: While small modifications of the original Kempe situation do not suffice, success is quite probable if we go up to around $n = 15$ to 17. This fact is perhaps theoretically of little interest, although practically it is of immense significance since the capacity of modern computers is quite adequate up to $n \leq 14$, but is hopelessly overtaxed from $n = 17$ onward. Moreover, the above consideration suggests that the second neighborhoods of small configurations might already suffice.

Now how does one arrive at the reducibility conjecture? To this end we must investigate the Kempe chain method more closely.

Let $K = (H, C)$ be an (m, n)-configuration that is embedded in an arbitrary triangulation, and let J_n be the set of distinct 4-color schemes on C_n and $j_n = |J_n|$. To make this clear once more, a color scheme corresponds exactly to a *partition* of the vertices of C_n into at most 4 classes, such that adjacent vertices lie in different classes. For example, if $n = 4$ and A, B, C, D are the vertices of C_4 in that order, then there are 4 color schemes: $AC|BD$, $AC|B|D$, $A|BD|C$, $A|B|C|D$. With the help of a recursion similar to that for chromatic polynomials in 3.4, one easily obtains:

$$j_n = 3j_{n-1} + 1 \quad (n \text{ even}), \quad j_n = 3j_{n-1} - 2 \quad (n \text{ odd}).$$

The following table gives some relevant values for j_n:

n	10	11	12	13	14	15
j_n	2461	7381	22144	66430	199291	597871

Suppose the subset $J(H) \subseteq J_n$ of the 4-colorings that are directly extendible to H form the x-th part of all schemes, i.e., $x = |J(H)|/j_n$. We shall call these colorings **good**. Now we take an arbitrary coloring $c \in J_n$ that is not directly extendible, and calculate the probability y that $c \in \overline{J}(H)$, i.e. that by a Kempe interchange c can be changed into a good coloring. We call such colorings **K-good**. Then the probability that K is D-reducible is $x + (1-x)y$, where we make the (somewhat problematical) assumption, that the individual probabilities are mutually independent. Therefore it all comes down to an estimation (depending on m and n) when $x + (1-x)y$ is close to 1.

First, let us estimate the dependence of y on x, and then the dependence of x on m and n. For this, we introduce the following terminology: Let G be a triangulation and $K = (H, C_n)$ be a configuration in G. We denote the residual graph by $U = G - H$. Therefore U is a triangulation up to one country, the interior of C_n, which has n edges. Let the colors be α, β, γ, δ. Every 4-coloring c' of U induces a 4-coloring c on C_n which we call the **restriction** of c'. Conversely, we call $c \in J_n$ **extendible** if c can be extended to all of U.

Let c' be a 4-coloring of U, and c the restriction on C_n. We denote by $\pi = \alpha\beta|\gamma\delta$ the partition of the color set into the two parts $\alpha\beta$, resp. $\gamma\delta$. In an obvious way π induces a decomposition of U into connected **α,β-components**, resp. **γ,δ-components**. The restriction of these components to C_n are called the **π-residues** of c', which set we denote by $R(\pi)$. $R_{\alpha\beta}(\pi)$, resp. $R_{\gamma\delta}(\pi)$ are the α,β-residues, resp. γ,δ-residues. If we consider only C_n, then by using the restriction c, π decomposes the vertices of C_n into connected α,β-, resp. γ,δ-arcs, which we call the **α,β-pieces**, resp. **γ,δ-pieces**, taken together the **π-pieces**. It is clear that each α,β-residue is composed of a number of α,β-pieces, that are joined by α,β-chains outside of C_n. Let us denote by $q(\pi)$ the number of π-pieces. It may happen (but only for even n) that c uses only the colors α and β, or only the colors γ and δ. In this case $q(\pi) = 1$. In all other cases the α,β-pieces, resp. γ,δ-pieces lie alternately around C_n, so $q(\pi)$ must be even.

Figure 10.2 illustrates these concepts. For reasons of space, the residual graph $U - C_n$ was placed in the interior of C_n (it is usually viewed as being outside C_n). Clearly, this makes no difference whatsoever for our argument.

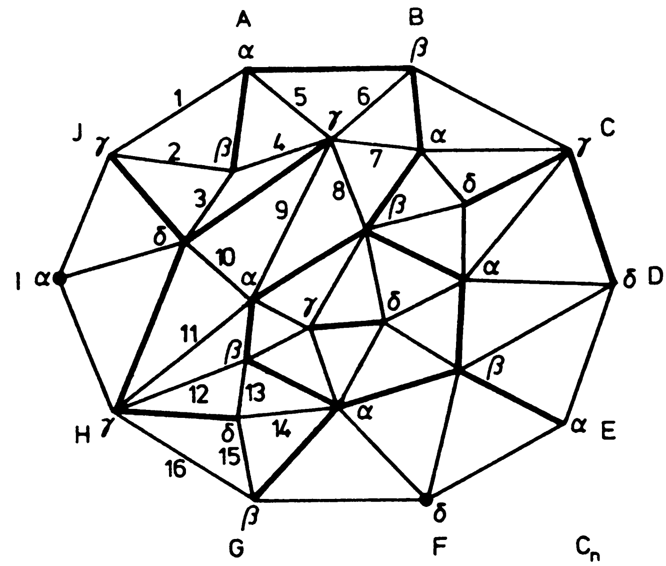


Figure 10.2

In the graph U shown above, there are two α,β-components and four γ,δ-components, where the bold edges display these components. $R(\pi) = \{ABEG, CD, F, HJ, I\}$. $R_{\alpha\beta}(\pi) = \{ABEG, I\}$. $R_{\gamma\delta}(\pi) = \{CD, F, HJ\}$. The number of π-pieces is $q(\pi) = 8$.

Clearly each admissible Kempe interchange (with respect to c') corresponds to a subset of $R(\pi)$, namely those residues in which the colors are interchanged. However, not all subsets give different color schemes. The initial coloring c has the scheme $\underset{\alpha}{AEI}|\underset{\beta}{BG}|\underset{\gamma}{CHJ}|\underset{\delta}{DE}$. The subset $\{I, CD, F\}$ corresponds to the interchange $I : \alpha \to \beta$, $C : \gamma \to \delta$, $D : \delta \to \gamma$, $F : \delta \to \gamma$, and we obtain the scheme $AE|BGI|DFHJ|C$. If we take $\{ABEG, HJ\}$, then there results the scheme $BGI|AE|C|DFHJ$, which is identical with the preceding. But we can immediately determine the exact number of the resulting schemes. Let $r = |R_{\alpha\beta}(\pi)|$ and $s = |R_{\gamma\delta}(\pi)|$. One convinces oneself easily that subsets $X, Y \subseteq R_{\alpha\beta}(\pi)$ yield the same scheme, after interchange, if and only if they are complementary sets in $R_{\alpha\beta}(\pi)$, and similarly for $R_{\gamma\delta}(\pi)$. In our example, $R_{\alpha\beta}(\pi) = \{ABEG, I\}$, $R_{\gamma\delta}(\pi) = \{CD, F, HJ\}$, and $\{I\}$, $\{ABEG\}$, resp. $\{CD, F\}$, $\{HJ\}$ are complementary sets. Thus, for the total number of schemes resulting from a Kempe interchange, we get $2^{r-1} \cdot 2^{s-1} - 1 = 2^{|R(\pi)|-2} - 1$ if $r \geq 1$ and $s \geq 1$. (We must subtract 1 since we exclude the empty interchange from which c again results.) If $r = 0$ or $s = 0$, then no new scheme is obtained.

Thus we face the problem of determining the number, $|R(\pi)|$, of residues. At first one would think that this number depends on the coloring c' of the entire residual graph U. However, we shall see that in fact it is already determined by the restriction, c, on C_n.

Let $S, T \in R(\pi)$, $S \neq T$. Suppose S consists of the pieces $S_1, \ldots, S_a$ and T of the pieces $T_1, \ldots, T_b$.

CLAIM: *It can not happen that S_i, T_k, S_j, T_l (possibly with intermediate spaces) lie cyclically around C_n for arbitrary $i \neq j$, $k \neq l$.* Or, in other words: T always lies entirely in one component of $C_n - S$ and S always entirely in one component of $C_n - T$. Assume $S, T \in R_{\alpha\beta}(\pi)$ (or analogously $\in R_{\gamma\delta}(\pi)$). Since an α, β-chain leads from S_i to S_j and similarly an α, β-chain from T_k to T_l, then, by the planarity of U, all 4 pieces must lie in the same α, β-component (Figure 10.3). If, on the other hand, $S \in R_{\alpha\beta}(\pi)$ and $T \in R_{\gamma\delta}(\pi)$, then clearly there can not simultaneously be chains from S_i to S_j and from T_k to T_l.

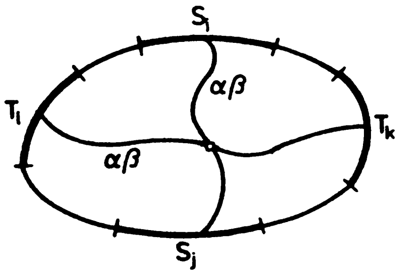


Figure 10.3

Let us call this **property (X)** of $R(\pi)$. Hence, it follows from (X) that two π-residues S and T are separated by exactly two arcs b and b' on C_n. For example, in the graph of Figure 10.2, CD and HJ are separated by the arcs $DEFGH$ and $JABC$. In the extreme case these arcs are single edges. Now we will prove the following **property (Y)** of $R(\pi)$.

CLAIM: *Let $S, T \in R(\pi)$ and b, b' be the separating arcs. If b is a single edge, then so is b'.* Proof: Since b is an edge, S and T are obviously in different color-components, say $S \in R_{\alpha\beta}(\pi)$ and $T \in R_{\gamma\delta}(\pi)$. Suppose the edge $b = AB$ has $A \in S$ and $B \in T$. Let $C \in S$ be the other endpoint of S and $D \notin S$ the adjacent vertex (see Figure 10.4). Thus we must show that $D \in T$.

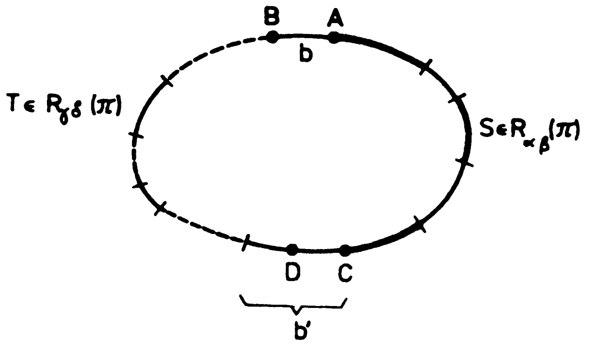


Figure 10.4

Let V be the α,β-component whose residue is S, and W the γ,δ-component whose residue is T. We call an edge of U a **V, W-bridge**, if one endpoint lies in V and the other in W. Since G is a triangulation, each edge of $U - C_n$ is contained in two triangles of the triangulation, and each edge of C_n in exactly one triangle. Clearly, a triangle always contains either, exactly two V,W-bridges, or, none. (See Figure 10.2: Each interior triangle contains two light edges = bridges.) We begin with the V,W-bridge b. b is contained in a triangle, let its other V,W-bridge be b_1. If b_1 is on C_n we stop, otherwise b_1 is contained in another triangle, let its other V,W-bridge be b_2. In this way we obtain a sequence $b, b_1, b_2, \ldots$ of successive incident V,W-bridges. It is easy to see that this sequence does not repeat itself. Consequently, a V,W-bridge $b_t \neq b$ must finally exist with $b_t \in C_n$. The endpoints A' and B' of b_t are by the construction of the sequence in S, resp. T, say $A' \in S$ and $B' \in T$.

It follows immediately from this and property (X) that $A' = C$ and $B' = D$, i.e., D is indeed in T.

To illustrate this argument consider again Figure 10.2 with $S = ABEG$, $T = HJ$. S and T are separated by the bridge $b = AJ$. We start with $b = 1$ and obtain the sequence, as shown in Figure 10.2, $b = 1, 2, 3, \ldots, 16 = b_t = GH$ with $G \in S$ and $H \in T$.

Properties (X) and (Y) suggest the following graph-theoretical description of $R(\pi)$. We construct the **chromodendron**, $X(\pi)$, corresponding to π in the following way: The vertices are the π-residues $R(\pi)$ and we join two residues by an edge if and only if both the separating arcs are single edges.

For our example in Figure 10.2 we obtain the chromodendron pictured in Figure 10.5.

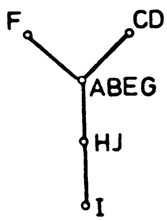


Figure 10.5

The name chromodendron hints that these graphs are always trees, which we are now going to prove.

10.1 THEOREM (Tutte-Whitney). *Let c' be a 4-coloring of U, and c the restriction to C_n. Suppose the partition $\pi = \alpha\beta|\gamma\delta$ induces $q(\pi)$ pieces on C_n. Then the following hold:*

i) *The chromodendron, $X(\pi)$, is a tree.*

ii) *If $q(\pi) > 1$ (and thus is even), then $X(\pi)$ has exactly $q(\pi)/2$ edges, and thus $q(\pi)/2 + 1 = |R(\pi)|$ vertices. If $q(\pi) = 1$ (i.e., c uses only the colors α, β or only the colors γ, δ), then $X(\pi)$ consists of a single vertex.*

Proof. We assume that $q(\pi) > 1$. By property (Y) the edges of $X(\pi)$ correspond to pairs of bridges between the residues $S \in R_{\alpha\beta}$ and $T \in R_{\gamma\delta}$ from which it immediately follows that $X(\pi)$ has exactly $q(\pi)/2$ edges. Furthermore, it is clear that $X(\pi)$ is connected since each residue T is cyclically accessible from every other residue S. Finally, it follows from property (X) that the removal of each pair of bridges separates the graph $X(\pi)$, therefore $X(\pi)$ is a tree. ∎

We have seen that each coloring c' of U induces a coloring c on C_n, all of whose colorings which are obtained from Kempe interchanges can be represented in one to one fashion by all possible chromodendra $X(\pi)$. Tutte and Whitney showed conversely (we leave the inductive proof to the reader) that for each arrangement of π-pieces that satisfies (X) and (Y), i.e., to each chromodenron $X(\pi)$, there is an extension from C_n to a certain residual graph U and there is a 4-coloring c' of U, such that c is exactly the restriction of c' and $X(\pi)$ is the chromodendron induced by c.

We summarize:

10.2 COROLLARY. *Let c be a 4-coloring of C_n.*

i) *The number of distinct Kempe arrangements is equal to the number of different chromodendra $X(\pi)$ corresponding to c.*

ii) *For a fixed partition π, the number of colorings obtained from c by Kempe interchanges with repect to π is $2^{(q(\pi)/2)-1} - 1$ when $q(\pi) > 1$ and 0 otherwise.*

Therefore, in particular, both numbers depend only on c regardless of the extension c'.

EXAMPLE. *Let us consider the specific case $n = 13$. Suppose c is a 4-coloring of C_{13}. Then $2 \leq q(\pi) \leq 12$ holds for each partition π of $\{\alpha, \beta, \gamma, \delta\}$ into 2 color pairs. Claim: There exists a partition π_1, with $q(\pi_1) = 10$ or 12, and another partition $\pi_2 \neq \pi_1$ with $q(\pi_2) \geq 8$. To prove this we assume that $\pi = \alpha\beta|\gamma\delta$ is a partition for which $q(\pi)$ is minimal; let π_1, π_2 be the two other partitions. By an α, β-**edge** we mean an edge of C_n one of whose endpoints is colored with α and the other with β, and analogously for γ, δ. All α, β-edges and γ, δ-edges must be bridges of π_1 and π_2 (i.e., they are not contained in any π_1-piece or in any π_2-piece).*

Case a. $q(\pi) = 2$. *Then there are at least* 11 α,β-*, resp.* γ,δ-*edges, whence* $q(\pi_1) \geq 11$, $q(\pi_2) \geq 11$ *follows, and thus* $q(\pi_1) = q(\pi_2) = 12$, *since* $q(\pi_i)$ *is even.*

Case b. $q(\pi) = 4$. *With the same reasoning it follows that* $q(\pi_1) \geq 13 - 4 = 9$ *and* $q(\pi_2) \geq 9$, *hence* $q(\pi_1) \geq 10$ *and* $q(\pi_2) \geq 10$.

Case c. $q(\pi) = 6$. *We obtain* $q(\pi_1) \geq 8$ *and* $q(\pi_2) \geq 8$. *Let* $e_1, \ldots, e_6$ *be the bridges of* π. *If* e_i *is not a bridge of* π_1, *then* e_i *is a bridge of* π_2, *and conversely. Therefore* $q(\pi_i) \geq 13-3 = 10$ *must hold for one of the partitions* π_1, π_2.

Case d. $q(\pi) = 8$. *With the same argument as in case c),* $q(\pi_i) \geq 14 - 4 = 9$ *holds, therefore* $q(\pi_i) \geq 10$ *for* $i = 1$ *or* 2.

In continuation of our example, let us determine the number of chromodendra for $n = 13$. Again let c be a 4-coloring of C_{13}. Assume $\pi = \alpha\beta|\gamma\delta$ is a partition with $q(\pi) = 8$. Claim: There are exactly 14 chromodendra $X(\pi)$ corresponding to π. We know that every chromodendron is a tree with $(q(\pi)/2)+1 = 5$ vertices. The possible trees with 5 vertices are depicted in Figure 10.6.

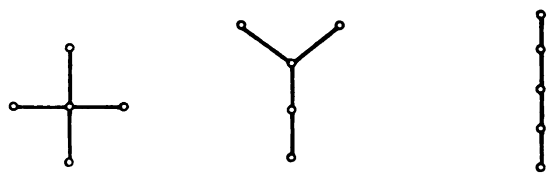
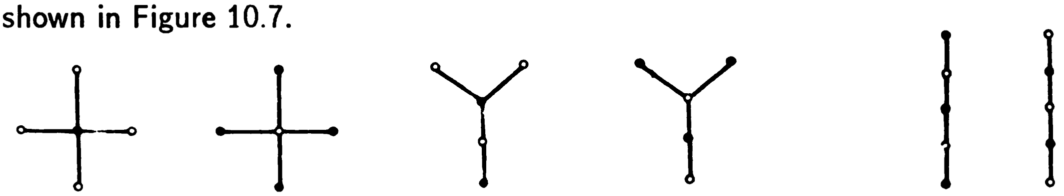


Figure 10.6

Each vertex corresponds to an α,β-residue or to a γ,δ-residue. We specify our trees more precisely by letting the black vertices correspond to $R_{\alpha\beta}$ and the white vertices to $R_{\gamma\delta}$. Since there must always be a black vertex and a white vertex among adjacent vertices, we obtain the list shown in Figure 10.7.



a) b) c) d) e) f)

Figure 10.7

Let $S_1, S_2, \ldots, S_8$ be the 8 π-pieces considered in a clockwise sense where we may assume that S_1, S_3, S_5, S_7 are the α,β-pieces and S_2, S_4, S_6, S_8 are the γ,δ-pieces. Clearly the structure of the α,β-residues (because of properties (X),(Y)) uniquely determines that of the γ,δ-residues. In this way we can easily determine the chromodendra. Figure 10.8 describes the case when S_1, S_3 and S_5 are joined by Kempe chains, and thus together yield an α,β-residue, while S_7 is a single α,β-residue. From this it follows that S_2 and S_4 are each γ,δ-residues and finally also $S_6 \bigcup S_8$, so that we obtain the chromodendron of type c) pictured on the right.

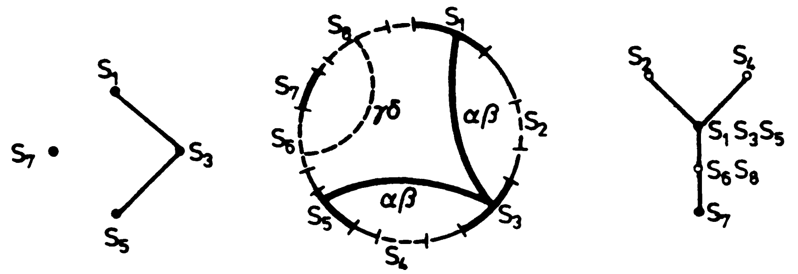


Figure 10.8

Figure 10.9 contains schemetically the 14 chromodendra where the letters a) to f) point out the correspondence to each of the the types from Figure 10.7.

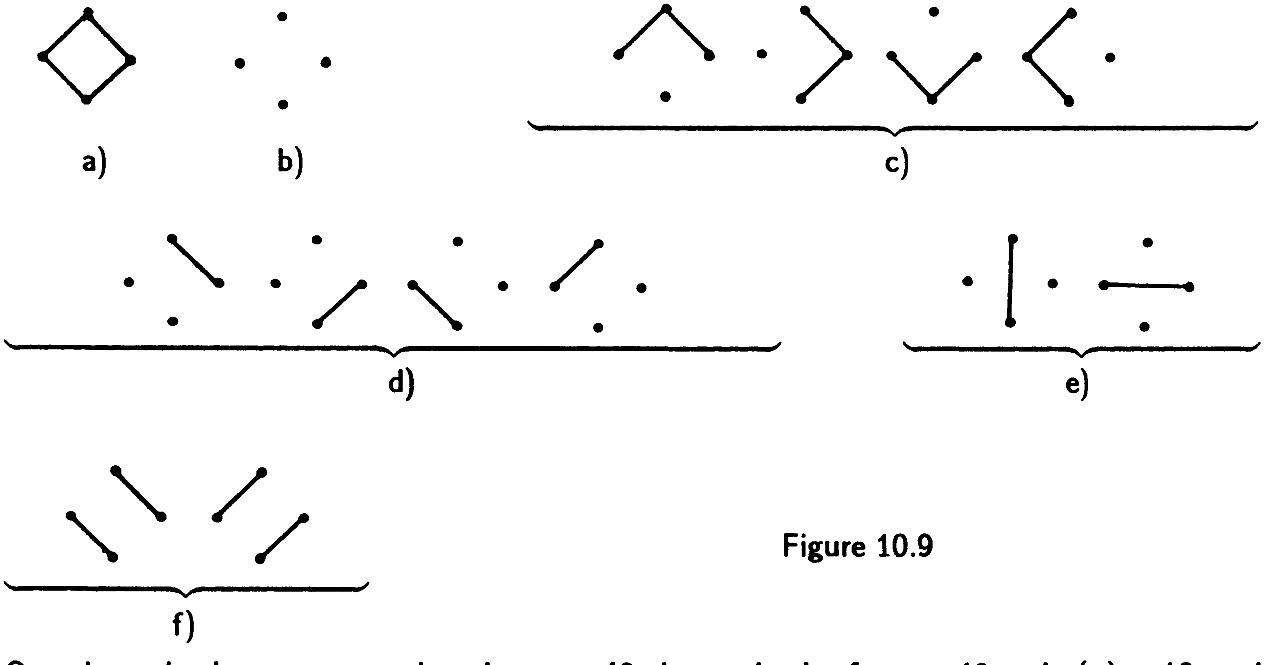


Figure 10.9

One shows in the same way, that there are 42 chromodendra for $n = 13$ and $q(\pi) = 10$, and 132 chromodendra for $n = 13$ and $q(\pi) = 12$.

Now we return to our original problem: Let the configuration $K = (H_m, C_n)$ be given. Let x be the probability that an arbitrary 4-coloring c of C_n is good. Assume c is not good. How large is the probability y that c is K-good?

Let us consider the case $n = 13$ just discussed in our example. We know that there is a partition π with $q(\pi) = 8$, 10, or 12. Let us denote the probabilities y by y_8, y_{10}, resp. y_{12}.

Let $q(\pi) = 8$. By 10.2(ii), for a fixed one of the 14 chromodendra there are exactly $2^3 - 1 = 7$ Kempe interchanges, and the probability that all 7 are *not* directly extendible is thus $(1-x)^7$. (In this connection we again make the somewhat unrealistic assumption that these events are mutually independent.) The probability that at least one of the 7 interchanges is good is accordingly $1-(1-x)^7$. Now if we bring in all 14 chromodendra, then we infer (independence assumed again) that

$$y_8 = \left[1 - (1 - (1-x)^7\right]^{14},$$

and analogously

$$y_{10} = \left[1 - (1-x)^{15}\right]^{42},$$

$$y_{12} = \left[1 - (1-x)^{31}\right]^{132}.$$

The functions y_i in dependence on x, are depicted in Figure 10.10.

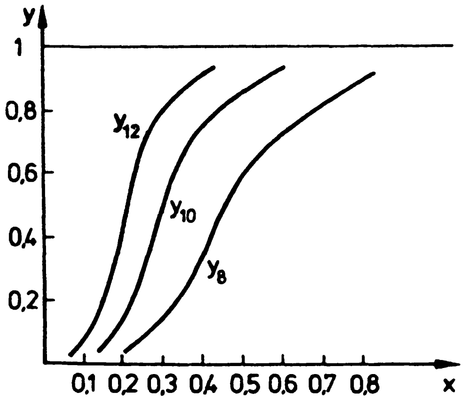


Figure 10.10

As a summary of our analysis we can say: An arbitrary 4-coloring c of C_{13} has probability x that it is good, and at least y_{10} (possibly at least y_{12}) that it is K-good. Thus the probability that K is D-reducible is at least $w(x) = x + (1-x)y_{10}$. Some values: $w(0.1) = 0.100062$, $w(0.2) = 0.422158$, $w(0.27) = 0.956694$. Similar considerations for the other values $n \leq 16$ lead immediately to the following assertion:

The Kempe method has only a slim chance of being successful when $x \leq 0.1$, a fair chance when x is around $x = 0.2$, and K is almost certainly D-reducible when $x \geq 0.3$.

Hence, we come to the second question: Can one reasonably estimate x in terms of m and n? First, we must expect (we have already established this at the beginning of this chapter) that the fraction x of the directly extendible colorings, with fixed n and increasing m, must rise rapidly. An analysis similar to that given before, made it appear plausible to Appel and Haken, that x and hence also $x + (1-x)y$, depends essentially on the difference $n - m$, and that for fixed n there is a **critical threshold** $\overline{m}$, so that $m > \overline{m}$ makes reducibility very probable, while $m < \overline{m}$ implies the opposite.

An investigation of the known obstruction free configurations by several authors, chiefly Bernhart, Allaire, Swart and Koch, indicated that up to $n = 11$ the critical value could be $\overline{m} = n-5$. Up to $n = 10$, all configurations with $m > \overline{m}$ are C-reducible, most of them even D-reducible. For $n = 11$ there is a single obstruction-free configuration with $m > \overline{m}$ that has not yet been reduced, and the value $\overline{m} = n - 5$ also appears quite plausible for higher n. Introducing the symbol $\varphi(K) = n - m - 3$ for an (m, n)-configuration $K = (H_m, C_n)$, we can summarize our considerations as follows:

If K is without obstructions and $\varphi(K) \leq 1$, then K is almost certainly reducible and very probably D-reducible.

But, now if we face an arbitrary configuration with $\varphi(K) \leq 1$, how do we know that K contains none of the three obstructions? As a first step toward answering this question, Appel and Haken proved the folowing result: *If K is an arbitrary configuration with $\varphi(K) \leq 0$, then K contains a geographically good subconfiguration $K^\star$ that also satisfies $\varphi(K^\star) \leq 0$. But $K^\star$ could still contain the third obstruction, a hanging 5-pair, in which case $K^\star$ would probably not be reducible.*

Now the following so-called m-lemma leads directly to the starting point of our considerations, the reducibility conjecture. One more concept: By an **admissible graph**, G, we mean a triangulation that exhibits all known properties of irreducible graphs (hence, in particular, all those proven in Chapter 9).

10.3 THEOREM (Appel-Haken). *Let K be an arbitrary (m, n)-configuration in an admissible graph G. If we assume $m > \frac{3n}{2} - 6$ (or equivalently $\varphi < 3 - \frac{n}{2}$), then K contains an obstruction-free subconfiguration $K^\star$ that also satisfies $m(K^\star) > \frac{3n(K^\star)}{2} - 6$. Furthermore: If u is an articulation of $K^\star$, then $m(W_i) \geq \frac{3}{2}n(W_i) - 6$ holds for at least one of the two subconfigurations W_1 and W_2 that arise upon removal of u.*

Proof. We use induction on m. If $m = 1$, then $n \leq 4$, and if $m = 2$ or 3, then $n \leq 5$. Thus by 9.1(iv) resp. 9.2, these cases can not occur. If $m = 4$, then $n \leq 6$. Therefore, by 9.2, $n = 6$. With the help of the Euler formula and the fact that every interior vertex must have degree ≥ 5, one sees immediately that in this case K must be the Birkhoff diamond. Since this is without obstructions, we can take $K = K^\star$. Now let $m \geq 5$ and suppose the assertion is correct up to $m - 1$.

OBSTRUCTION (A). Assume $u \in V(H_m)$ has $t \geq 4$ legs whose endpoints on the ring are $w_1, \ldots, w_t$. We can assume that the w_i's follow one another successively on C_n since otherwise we have an obstruction of type (B). The configuration K' as in Figure 10.11 satisfies $m' = m-1$, $n' \leq n-1$, and thus $m' = m-1 > (3/2)n-7 \geq (3/2)n'+(3/2)-7 > (3/2)n'-6$. By induction there is an obstruction-free subconfiguration K^* in K', and thus also in K.

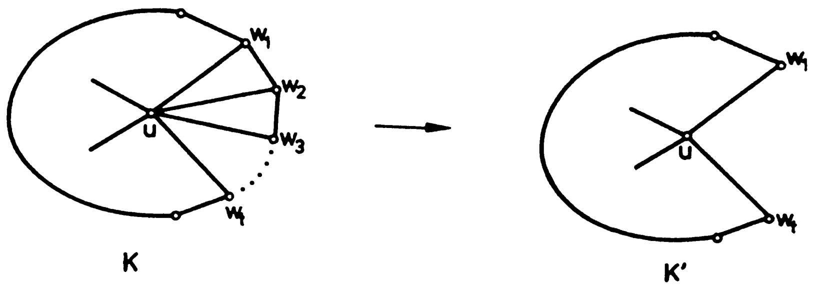


Figure 10.11

OBSTRUCTION (B). Let u be an articulation with at least 3 legs. Let the proper subconfigurations (i.e., those with interior vertices) that arise upon removal of u be $W_1, \ldots, W_t$ $(t \geq 2)$. If several legs follow one another in succession, then in the construction of the W_i's we take only the ones farthest out (see v_1, v_2, v_3 in Figure 10.12). Therefore when $t = 2$, then at least two legs of u are taken into account in only one W_i. We claim that at least one of the configurations W_i satisfies $m(W_i) > (3/2)n(W_i) - 6$. Let us assume the opposite. As is clear from Figure 10.12, we have

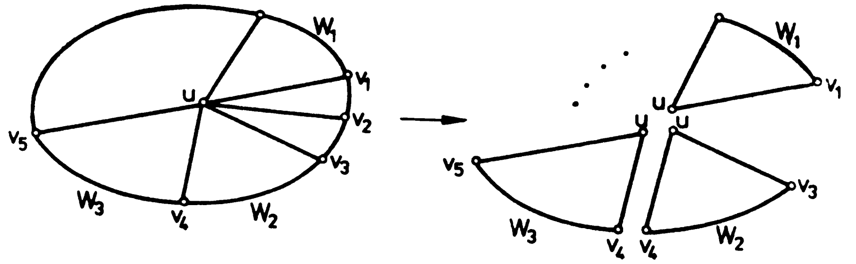


Figure 10.12

$$\sum_{i=1}^{t} m(W_i) = m - 1$$

$$\sum_{i=1}^{t} n(W_i) \leq n + 2t \quad \text{with} < \text{when } t = 2,$$

and thus

$$m = \sum_{i=1}^{t} m(W_i) + 1 \leq \frac{3}{2}\sum_{i=1}^{t} n(W_i) - 6t + 1 \leq \frac{3}{2}n - 3t + 1,$$

where, if $t = 2$, we have $\leq (3/2)n - 13/2$ in the last inequality. Since $m \leq (3/2)n - 3t + 1 \leq (3/2)n - 6$ for $t \geq 3$, and $m \leq (3/2)n - 13/2 < (3/2)n - 6$ for $t = 2$, we have a contradiction to the assumption in all cases.

OBSTRUCTION (C). Let (u, v) be a hanging 5-pair with the common interior neighbor w. Since we have already treated obstructions of type (B), we can assume the situation of Figure 10.13. For the subconfiguration K' in Figure 10.13, $m(K') = m - 3$ and $n(K') = n - 2$. Therefore $m(K') = m - 3 > (3/2)n - 3 - 6 = (3/2)(n - 2) - 6 = (3/2)n(K') - 6$. The last assertion is treated exactly as case (B). ■

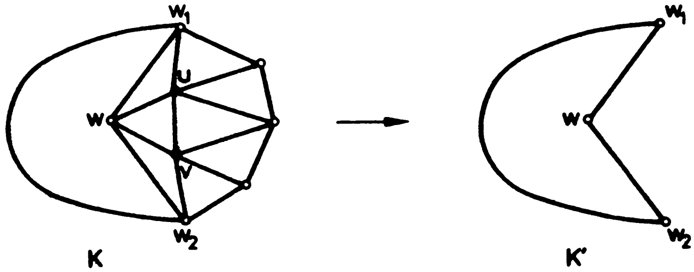


Figure 10.13

With this result we have concluded our probabilistic considerations regarding reducibility: *If an (m, n)-configuration K satisfies the inequality $m > (3/2)n - 6$, then it is almost certainly reducibile and very probably D-reducible.*

Let us now tackle the question of unavoidability. Let G be an admissible graph. We want to show that G almost certainly contains a D-reducible configuration with $n \leq 17$ (and very likely one with $n \leq 14$), while, on the other hand, it is to be expected that there are admissible graphs in which all (m, n)-configurations with $n \leq 12$ are not reducible.

For this, as in our considerations after the formulation of the reducibility conjecture, we must estimate the ring size n with respect to m for small neighborhoods. In this connection a concrete result was first obtained by Stromquist in 1975: *Every admissible graph contains a configuration K with $\varphi(K) \leq -1$, which is contained in the second neighborhood of a vertex and in which no vertex has degree* > 30. Together with the result of Appel and Haken, cited before 10.3, this proves once again the existence of an unavoidable set of geographically good configurations.

We introduce the following **neighborhood classes**:

Type	Interior
1	single vertex
2	edge
3	triangle
4	double triangle (=vertex with three successive neighbors)
5	triple triangle (=vertex with four successive neighbors)
6	first neighborhood of a vertex
7	first neighborhood of an edge
⋮	⋮
t	first neighborhood of a class of type $t-5$ $(t \geq 6)$
⋮	⋮

Every neighborhood class N_t has an average number m_t of interior vertices and an average number n_t of ring vertices with N_t as interior and thus an average value $\varphi_t = n_t - m_t - 3$. As already seen, we can initially assume the average degree of a vertex to be 6. For small neighborhoods this yields the following table:

t	1	2	3	4	5	6	7	8
m_t	1	2	3	4	5	7	10	12
n_t	6	8	9	10	11	12	14	15
φ_t	2	3	3	3	3	2	1	0

As an example, Figure 10.14 indicates the values for the first neighborhood of an edge $(t = 7)$.

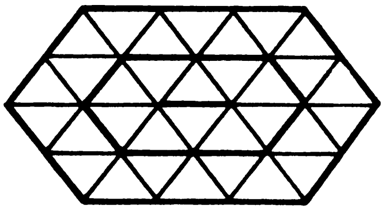


Figure 10.14

In general, one easily shows by induction (the reader may verify this) that from $t = 5$ on, φ_t decreases monotonically and that we obtain the values given in Figure 10.15. In the diagram, the φ-values are recorded as $\times$ with respect to n, where the number over the symbol $\times$ gives the neighborhood class. In particular, we see that starting with $n = 21$ (neighborhood class 13) the φ-values lie below the line $\varphi = 3 - n/2$. That means: Starting with $n = 21$, G almost certainly contains (the average degree is in reality < 6) an (m, n)-configuration that satisfies $m > (3/2)n - 6$, hence has very large probability of being reducible.

We can still sharpen our estimate considerably. By 9.1(ii), every admissible graph must contain a large number of 5-vertices that we have not yet taken into account, and consequently a large number of configurations with 5-vertices in the interior. Now for each t we consider neighborhood classes $\overline{N}_t$ in which a 5-vertex lies as closely as possible to the center. Thus $\overline{N}_1$ consists of a single 5-vertex, $\overline{N}_2$ of an edge one of whose endpoints has degree 5, etc. Now calculating the average value as previously, where we again assume degree 6 for all other vertices, then we obtain the φ-values provided with the symbol • in Figure 10.15. Again the class number is given over each •.

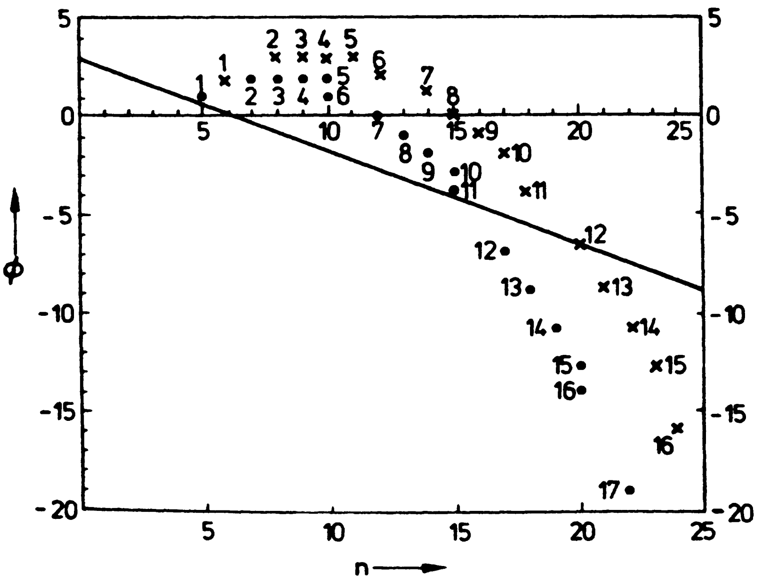


Figure 10.15

Now the φ-values starting with $n = 17$ (neighborhood class 12) are in the region of high probability for reducibility. On the other hand, it is to be expected that there are admissible graphs in which all configurations with $n \leq 12$ have φ-values ≥ 0. Also for $n = 13$, it is imaginable that admissible graphs exist whose configurations all have φ-values ≥ -1. An analogous analysis of the cases $n = 14$, 15 and 16 points out, on the contrary, that every admissible graph has a reducible configuration in this n-region.

With that we have concluded the probability analysis with the result mentioned at the beginning: *It is almost certain that an unavoidable set of reducible configurations exists with $n \leq 17$, and highly likely that $n \leq 14$ already suffices, while it is very likely that triangulations exist in which all configurations with $n \leq 12$ are irreducible (more precisely, not C-reducible).*

This was the situation in 1975. The probabilistic considerations convinced Appel and Haken that a computer based solution of the 4-color problem was possible, and they also indicated which configurations would arise. In particular, they suggested restricting all further work exclusively to configurations with $n \leq 14$. And finally their arguments stressed the perhaps bitter but insurmountable fact that, at least for the Kempe method alone, they could not do without a computer. Calculation by hand is simply too slow for the tens of thousands of individual steps (see the table for j_n).

In the last stage of their work improvements of the unloading algorithm and tests for reducibility were devised simultaneously (chiefly by John Koch). The final computer runs were so to speak "self-correcting". Appel, Haken and Koch began with a first approximation to their unloading algorithm. Whenever a large vertex with positive terminal loading resulted, the second neighborhood of this vertex was tested, to see whether an obstruction-free configuration appeared. If none occurred, the neighborhood was called **critical**. If an (m, n)-configuration without obstructions was found, then it was tested for reducibility, according to the guidelines worked out in their probability considerations: For $n > 14$, the configuration was not accepted. For $n \leq 14$ they would test first for D-reducibility, with the rule of thumb illustrated in Figure 10.10: If the percentage of good colorings was substantially below 20%, then the configuration was discarded. Otherwise D-reducibility was tried. If this led to a negative result, several types of reductors were tested for C-reducibility. Every configuration that could not be relatively quickly reduced in this way (90 minutes on an IBM 370-158, resp. 30 minutes on an IBM 370-168) was also discarded. In all these cases the neighborhood was also called critical. At the end of the run the critical neighborhoods were classified, whereupon the algorithm was modified accordingly, in order to exclude these neighborhoods.

This dialogue between man and machine lasted from January until June 1976. Reducibility was tested on three computers, using over 1000 hours, while the final modification of the unloading algorithm was carried out by hand. About 10000 neighborhoods of vertices with positive loading were analyzed by hand and over 2000 configurations were tested for reducibility by machine. In June 1976 this monumental undertaking was crowned with success: An unavoidable set with 1936 reducible configurations (all with $n \leq 14$) was found. By discarding duplications and making some small improvements, this number was soon reduced to 1482 and finally to 1405. The present state (March 1983) is 1258. Thus, finally after 124 years the 4-color problem became the 4-color theorem.

4-COLOR THEOREM. *Every plane map is 4-colorable.*

The announcement of the solution of the 4-color problem was greeted with mixed feelings by the mathematical world. Besides the enthusiasm that a 100 year old problem was finally solved, and the respect for the achievement of the mathematicians that took part in this monumental endeavor, there was also a feeling of disappointment, chiefly for two reasons. The first was of a mathematical-esthetic nature. The hope that one of the elegant refined theoretical concepts (that we discussed in Chapters 4–8) would lead to the goal had been thwarted, success was achieved by the oldest and most direct method, and that with the massive help of computers. The second and more important objection concerned the length of the proof. A computer based proof of such length is impossible to verify for those who do not have similar computer capacity at their disposal. It is, stated bluntly, more a matter of belief than a mathematical fact.

Both objections are understandable, but it seems that on second sight they are not really justified. All the other attacks discussed in this book require somewhat stronger theoretical means than the Euler formula which is the basis for the Kempe method. It can not be excluded that such means will be found, however, no one can guarantee that they in fact exist, or secondly, that they would significantly shorten the proof. As regards the use of computers, the proof of Appel-Haken-Koch follows in principle the usual mathematical procedures, but, on the other hand, it radically breaks new ground. All mathematics begins with an intuitive idea of what the final result could look like. This intuition is then carried through using logical inferences, and many experiments until a positive conclusion is reached. In the proof of the 4-color theorem sketched in this chapter, the guiding intuition is decisively influenced by probabilistic considerations (which for this reason were discussed at great length). The modification of the unloading algorithm follows exact logical steps (with the help of computer tests) and the experiments (reducibility tests) finally, because of their length, were left almost entirely to computers. The design of computer experiments of this size was unthinkable before the advent of high speed computers, and thus the resentment of many mathematicians with regard to these "computer slaves" is really more a psychological rather than a mathematically based problem. At any rate, it can be safely predicted that the two principle novel features of the Appel-Haken proof—probabilistic analysis of the success and the man-machine dialogue—will find in the near future a firm place in mathematical research.

Now to the question of the length of the proof. If a proof is short and on a high mathematical level, then it can and probably will be verified by many mathematicians that have the same background, so that finally logical or computational errors are excluded. But already here there are examples of theorems that were held to be correct over decades, until finally an error was found. It can be assumed as certain that there are numerous mathematical theorems that contain errors which were never discovered. In any case one arrives irrefutably at the conclusion that for very long proofs, and proofs provided with many calculations, the probability that the proof by hand is in error, is substantially greater than that of computer verification and the correctness of the machine program. The lasting significance of the 4-color theorem could lie in the fact that it is a particularly impressive example of a mathematical theorem, whose proof by nature must be very long, so long that it *can* not be acheived without computer support. Even if a proof of moderate length is finally found, the work of Appel and Haken still points far into the future. It could be the first example of many mathematical theorems, that can only be verified with the massive help of computers. Whether an enlarged or finally everyday contribution of large computers will be a blessing to mathematics, is yet to be seen, in any case it will alter it fundamentally, before all what concerns the traditional proof concept. The mathematical world will be forced to abandon total verification by hand (exactly as it is the case in physics in the microcosmos or the macrocosmos), but the breadth of mathematical knowledge will more than make up for this last. And so we come to the end. For more than 100 years the 4-color problem has inspired important mathematics, enabled an entire new discipline to emerge, and the solution opens the door to new ideas that perhaps for the next 100 years will influence to a high degree the future of mathematical development. Can one desire more?

EXERCISES FOR CHAPTER 10

1. Check the heuristic considerations at the conclusion of the R-conjecture.

2. Verify the recursion formulas for $j_n = |J_n|$: $j_n = 3j_{n-1} + 1$ (n even), $j_n = 3j_{n-1} - 2$ (n odd) and compute j_n explicitly from this recursion.

3°. Complete the proof of property (Y) of $R(\pi)$.

4. Sketch a proof that for each chromodendron, there exists an extension (see the remarks after 10.1).

5°. Show that there are 42 chromodendra for $n = 13$ and $q(\pi) = 10$.

6. Prove: If K is a configuration with $\varphi(K) \leq 0$, then K contains a geographically good subconfiguration $K^\star$ with $\varphi(K^\star) \leq 0$.

7°. Show that the Birkhoff diamond is the only (4,6)-configuration in an admissible graph.

8. Verify the values in the table in Figure 10.15.

BIBLIOGRAPHY

Every bibliographical selection is made with a certain degree of arbitrariness. Here we group by chapters—references to related reading and a brief selection of the most important original works (which throughout refer to the ideas and results of the corresponding chapter).

FOR CHAPTER 1

The book by Biggs-Lloyd-Wilson is an excellent presentation of the origins of graph theory from which we have taken most of the references to the early period of the 4-color problem. Among the many books in English, the one by Harary is certainly a good choice. The German literature is scarcer, for accompanying reading we recommend the book by Wagner. Whoever wishes to read an exhaustive presentation of the Jordan curve theorem may consult Aleksandrov.

Aleksandrov P. S., *Combinatorial Topology*. Vol. 1. Baltimore: Graylock Press 1956.

Biggs N. L., Lloyd E. K. and Wilson R. J., *Graph Theory 1736-1936*. Oxford: Clarendon Press 1976.

Harary F., *Graph Theory*. Reading: Addison-Wesley 1969.

Wagner K., *Graphentheorie*. Mannheim-Wien-Zürich: Bibliographisches Institut 1970.

Specialized References:

1. Cayley A., *On the Theory of the Analytical Forms Called Trees*. *Phil. Mag.* (4) **23** (1857), 172–176.
2. Cayley A., *On the Mathematical Theory of Isomers*. *Phil. Mag.* (4) **47** (1874), 444–446.
3. Cayley A., *On the Colouring of Maps*. *Proc. Royal Geog. Soc.* **1** (1879), 259–261.
4. Euler L., *Solutio problematis ad geometriam perinentis*. *Comm. Acad. Sci. Imp. Petropolitanae* **8** (1736), 128–140.
5. Euler L., *Letter to C. Goldbach* (1750).
6. Guthrie F., *Note on the Colouring of Maps*. *Proc. Royal Soc. Edinb.* **10** (1880), 727–728.

7. Kempe A. B., *On the Geographical Problem of the Four Colours.* *Amer. J. Math.* **2** (1879), 193–200.
8. Kempe A. B., *How to Colour a Map with Four Colours.* *Nature* **21** (1879–80), 399–400.
9. Kirchhoff G. R., *Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird.* *Ann. Phys. Chem.* **72** (1847), 497–508.
10. Tait P. G., *Remarks on the Colouring of Maps.* *Proc. Royal Soc. Edinb.* **10** (1878–80), 729.

FOR CHAPTER 2

The book by Fréchet and Fan is a beautiful introduction to combinatorial topology. Hilbert-Cohn-Vossen is unsurpassed in its clarity. Ringel's book gives a self-contained presentation of Heawood's color theorem.

Fréchet M. and Fan K., *Initiation to Combinatorial Topology.* Boston: Prindle, Weber and Schmidt 1967.

Hilbert D. and Cohn-Vossen S., *Anshauliche Geometrie.* Berlin: Springer 1932.

Ringel G., *Map Color Theorem.* Berlin-Heidelberg-NewYork: Springer 1974.

Specialized References:

1. Cauchy A.-L., *Recherches sur les polyèdres—premier mémoire.* *J. École Polytechn.* **9** (Cah. 16) (1813), 68-86.
2. Franklin P., *A Six-Color Problem.* *J. Math. Phys.* **13** (1934), 363–369.
3. Gustin W., *Orientable Embeddings of Cayley Graphs.* *Bull. Amer. Math. Soc.* **69** (1963), 272–275.
4. Heawood P. J., *Map-Colour Theorem.* *Quart. J. Pure Appl. Math.* **24** (1890), 332–338.
5. Heffter L., *Über das Problem der Nachbargebiete.* *Math. Ann.* **38** (1891), 477–508.
6. Lhuilier S.-A.-J., *Mémoire sur la polyèdrométrie.* *Ann. Mathématiques* **3** (1812-3), 169–189.
7. Möbius A. F., *Über die Bestimmung des Inhaltes eines Polyeders.* *Bericht K. Sächs. Ges. Wiss. Leipzig Math.-Phys. Cl.* **17** (1865), 31–68.

8. Ringel G., *Bestimmung der Maximalzahl der Nachbargebiete auf nichtorientierbaren Flächen. Math. Ann.* **127** (1954), 181–214.
9. Ringel G., *Färbungsprobleme auf Flächen und Graphen.* Berlin: VEB Deutscher Verlag d. Wiss. 1959.
10. Ringel G. and Youngs J. W. T., *Solution of the Heawood Map-Coloring Problem. Proc. Nat. Acad. Sci. USA* **60** (1968), 438–445.
11. Tietze H., *Einige Bemerkungen über das Problem des Kartenfärbens auf einseitigen Flächen. Jahresber. DMV* **19** (1910), 155–159.

FOR CHAPTER 3

Grünbaum's book is a beutiful introduction to convex geometry, taking one up to the frontiers of present-day research. König's classical graph theory book is still interesting.

Grünbaum B., *Convex Polytopes.* London: Wiley 1967.

König D., *Theorie der endlichen und unendlichen Graphen.* Leipzig: Akad. Verlagsgesellschaft 1936.

Specialized References:

1. Birkhoff G. D., *A Determinant Formula for the Number of Ways of Coloring a Map. Ann. Math.* (2) **14** (1912-3), 42–46.
2. Birkhoff G., *The Reducibility of Maps. Amer. J. Math.* **35** (1913), 115–128.
3. Franklin P., *The Four Color Problem. Amer. J. Math.* **44** (1922), 225–236.
4. Hamilton W. R., *Account of the Icosian Calculus. Proc. Royal Irish Acad.* **6** (1853-7), 415–416.
5. Heawood P. J., *On the Four-Colour Map Theorem. Quart. J. Pure Appl. Math.* **29** (1898), 270–285.
6. Kirkman T. P., *On the Representation of Polyhedra. Phil. Trans. Royal Soc. London* **146** (1856), 413–418.
7. Petersen J., *Die Theorie der regulären Graphs. Acta Math.* **15** (1891), 193–220.
8. Petersen J., *Sur le théorème de Tait. L'Intermédiarie des Mathem.* **5** (1898), 225–227.

9. Steinitz E., *Polyeder und Raumeinteilungen. Enzykl. Math. Wiss.* III AB **12** (1922), 1–139.
10. Tait P. G., *Note on a Theorem in the Geometry of Position. Trans. Royal Soc. Edinb.* **29** (1880), 657–660.
11. Tait P. G., *Listing's Topologie. Phil. Mag.* (5) **17** (1884), 30–46.
12. Veblen O., *An Application of Modular Equations in Analysis Situs. Ann. Math.* (2) **14** (1912-3), 86–94.
13. Whitney H., *The Coloring of Graphs. Ann. Math.* (2) **33** (1932), 688–718.
14. Whitney H., *Non-separable and Planar Graphs. Trans. Amer. Math. Soc.* **34** (1932), 339–362.

FOR CHAPTER 4

The book by White is a very readable presentation of embedding questions. Recent results on several topological parameters are contained in White-Beineke.

White A. T., *Graphs, Groups and Surfaces.* Amsterdam-London: North Holland 1973.

White A. T. and Beineke L. W., *Topological Graph Theory. Selected Topics in Graph Theory (Beineke;Wilson, eds.)* New York: Academic Press 1978, 15–49.

Specialized References:

1. Guy R., *The Decline and Fall of Zarankiewicz's Theorem. Proof Techniques in Graph Theory (Harary, ed.)*, New York: Academic Press 1969, 63–69.
2. Jungerman M., *The Non-Orientable Genus of the n-Cube. Pacific J. Math.* **76** (1978), 443–451.
3. Kleitman D. J., *The Crossing Number of* $K_{5,n}$. *J. Combinatorial Theory* **9** (1970), 315–323.
4. Kuratowski K., *Sur le problème des courbes gauches en topologie. Fund. Math.* **15** (1930), 271–283.
5. Mac Lane S., *A Structural Characterization of Planar Combinatorial Graphs. Duke Math. J.* **3** (1937), 340–372.

6. Menger K., *Zur allgemeinen Kurventheorie.* *Fund. Math.* **10** (1927), 96–115.

7. Ringel G., *Über drei kombinatorische Probleme am n-dimensionallen Würfel und Würfelgitter.* *Abh. Math. Sem. Univ. Hamburg* **20** (1955), 10–19.

8. Ringel G., *Das Geschlecht des vollständigen paaren Graphen.* *Abh. Math. Sem. Univ. Hamburg* **28** (1965), 139–150.

9. Ringel G., *Der vollständige paare Graph auf nichtorientierbaren Flächen.* *J. Reine Angew. Math.* **220** (1965), 89–93.

10. Wagner K., *Über eine Eigenschaft der ebenen Komplexe.* *Math. Ann.* **114** (1937), 570–590.

11. Whitney H., *Congruent Graphs and the Connectivity of Graphs.* *Amer. J. Math.* **54** (1932), 150–168.

12. Whitney H., *Planar Graphs.* *Fund. Math.* **21** (1933), 73–84.

FOR CHAPTER 5

An interesting survey of edge-colorings is given in Fiorini-Wilson. Golumbic's book is an excellent exposition of the very active field of perfect graphs.

Fiorini S. and Wilson R. J.,*Edge-Colorings of Graphs.* *Selected Topics in Graph Theory (Beineke; Wilson, eds.).* New York: Academic Press 1978, 103–126.

Golumbic M., *Algorithmic Graph Theory and Perfect Graphs.* New York: Academic Press 1980.

Specialized References:

1. Birkhoff G. D. and Lewis D., *Chromatic Polynomials.* *Trans. Amer. Math. Soc.* **60** (1946), 355–451.

2. Brooks R. L. *On Colouring the Nodes of a Network.* *Proc. Cambridge Phil. Soc.* **37** (1941), 194–197.

3. Catlin P. A., *Hajós' Graph-Coloring Conjecture: Variations and Counterexamples.* *J. Combinatorial Theory* **26B** (1979), 268–274.

4. Dirac G. A., *A Property of 4-Chromatic Graphs and some Remarks on Critical Graphs.* *J. London Math. Soc.* **27** (1952), 85–92.

5. Dirac G. A., *The Structure of k-Chromatic Graphs.* *Fund. Math.* **40** (1953), 42–55.

6. Hadwiger H., *Über eine Klassifikation der Streckenkomplexe.* *Viertelj. Schr. Naturforsch. Ges. Zürich* **88** (1943), 133–142.

7. Lovász L., *Normal Hypergraphs and the Perfect Graph Conjecture.* *Discrete Math.* **2** (1972), 253–267.

8. Mycielski J., *Sur le coloriage des graphes.* *Colloq. Math.* **3** (1955), 161–162.

9. Shannon C. E., *A Theorem on Coloring the Lines of a Network.* *J. Math. Phys.* **28** (1949), 148–151.

10. Tutte W. T., *More about Chromatic Polynomials and the Golden Ratio.* *Combinatorial Structures and Their Applications (Guy; Hanani; Sauer; Schönheim, eds.).* New York-London-Paris: Gordon Breach 1970, 439–453.

11. Tutte W. T., *The Golden Ratio in the Theory of Chromatic Polynomials.* *Ann, N.Y. Acad. Sci.* **175** (1970), 391–402.

12. Vizing V. G., *On an Estimate of the Chromatic Class of a p-Graph (Russian).* *Diskret. Analiz.* (1964), 25–30.

13. Wagner K., *Bemerkungen zu Hadwigers Vermutung.* *Math. Ann.* **141** (1960), 433–451.

14. Whitney H., *A Logical Expansion in Mathematics.* *Bull. Amer. Math. Soc.* **38** (1932), 572–579.

FOR CHAPTER 6

Ford-Fulkerson gives a classical treatment of maximum-minimum theorems and flows in networks. Of the vast literature on transversal theory, thee book by Minsky may still be the most comprehensive. For the reader who is interested in the matrix viewpoint and Latin squares, we warmly recommend the book by Ryser (which still represents the most beautiful seduction to combinatorics).

Ford L. R. and Fulkerson D. R., *Flows in Networks.* Princeton: Princeton Univ. Press 1962.

Minsky L., *Transversal Theory.* New York, London: Akacademic Press 1971.

Ryser H. J., *Combinatorial Mathematics. Carus Math. Monographs* **14**, New York: Math. Assoc. Amer. 1963.

Specialized References:

1. Gale D., *A Theorem on Flows in Networks. Pacific J. Math.* **7** (1957), 1073–1082.
2. Hall M., *Distinct Representatives of Subsets. Bull. Amer. Math. Soc.* **54** (1948), 922–926.
3. Hall M., *On Representatives of Subsets. J. London Math. Soc.* **10** (1935), 26–30.
4. Knuth D. E., *A Permanent Inequality. Amer. Math. Monthly* **88** (1981), 731–740, 798.
5. König D., *Graphen und Matrizen. Mat. Fiz. Lapok* **38** (1931), 116–119.
6. Mirsky L. and Perfect H., *Systems of Representatives. J. Math. Anal. Appl.* **15** (1966), 520–568.
7. Nash-Williams C. St. J. A., *Edge-Disjoint Trees of Finite Graphs. J. London Math. Soc.* **36** (1961), 445–450.
8. Ore Ø., *Graphs and Matching Theorems. Duke Math. J.* **22** (1955), 625–639.
9. Ryser H. J., *Combinatorial Properties of Matrices of Zeros and Ones. Can. J. Math.* **9** (1957),371–377.
10. Tutte W. T., *The Factorization of Linear Graphs. J. London Math. Soc.* **22** (1947), 107–111.
11. Tutte W. T., *A Short Proof of the Factor Theorem for Finite Graphs. Can. J. Math.* **6** (1954), 347–352.

FOR CHAPTER 7

A stimulating summary of recent results is contained in Bermond. For results on plane graphs and polytopes, we refer the reader to the corresponding chapter in Grünbaum's book (see Chapter 3). The reader will find a wealth of information about Ramsey theory, which has developed enormously in the last few years, in the book by Graham-Rothschild-Spencer.

Bermond J. C., *Hamiltonian Graphs. Selected Topics in Graph Theory (Beineke; Wilson, eds.)*, New York: Academic Press 1978, 127–167.

Graham R. L., Rothschild B. L. and Spencer J. H., *Ramsey Theory*. New York: Wiley 1980.

Specialized References:

1. Bondy J. A. and Chvátal V., *A Method in Graph Theory. Discrete Math.* **15** (1967), 111–135.
2. Chvátal V., *On Hamilton's Ideals. J. Combinatorial Theory* **12B** (1972), 163–168.
3. Chvátal V., *New Directions in Hamiltonian Graph Theory. New Directions in the Theory of Graphs (Harary, ed.)*, New York: Academic Press 1973, 65–95.
4. Chvátal V. and Erdős P., *A Note on Hamiltonian Circuits. Discrete Math.* **2** (1972), 111–113.
5. Dirac G. A., *Some Theorems on Abstract Graphs. Proc. London Math. Soc.*, Ser. 3, **2**(1952), 69–81.
6. Eberhard V., *Zur Morphologie der Polyheder*. Leipzig 1891.
7. Fleischner H., *The Square of Every Two-Connected Graph is Hamiltonian. J. Combinatorial Theory* **16B** (1974), 29–34.
8. Nash-Williams C. St. J. A., *Hamilton Arcs and Circuits. Recent Trends in Graph Theory (Capobianco et al, eds.)*, Lecture Notes in Math. **186**, Berlin-Heidelberg-New York: Springer 1971, 197–210.
9. Ore Ø., *Note on Hamilton Circuits. Amer. Math. Monthly* **67** (1960), 55.
10. Ramsey F. P., *On a Problem on Formal Logic. Proc. London Math. Soc.* (2) **30** (1930), 264–286.
11. Turán P., *Eine Extremalaufgabe aus der Graphentheorie. Mat. Fiz. Lapok* **48** (1941), 436–452.
12. Tutte W. T., *On Hamilton Circuits. J. London Math. Soc.* **21** (1946), 98–101.
13. Tutte W. T., *A Theorem on Planar Graphs. Trans. Amer. Math. Soc.* **82** (1956), 99–116.

14. Whitney H., *A Theorem on Graphs*. *Ann. Math.* **32** (1931), 378–390.

FOR CHAPTER 8

For an introduction to the theory of matroids, the corresponding chapters in Aigner and the book by Welsh are recommended. A development from the concept of a chain group can be examined in the, already classic, paper by Tutte.

Aigner M., *Combinatorial Theory*. Berlin-Heidelberg-New York: Springer 1979.

Tutte W. T., *Lectures on Matroids*. *J. Res. Nat. Bur. Stand.* **69B** (1965), 1–48.

Welsh D. J. A., *Matroid Theory*. London-New York-San Francisco: Academic Press 1976.

Specialized References:

1. Edmonds J., *Minimum Partition of a Matroid into Independent Subsets*. *J. Res. Nat. Bur. Starnd.* **69B** (1965), 67–72.
2. Edmonds J., *Submodular Functions, Matroids and Certain Polyhedra*. *Combinatorial Structures and Their Applications (Guy; Hanani; Sauer; Schñheim, eds.)*, New York-London-Paris: Gordon Breach 1970, 69–87.
3. Edmonds J. and Fulkerson D. R., *Transversals and Matroid Partition*. *J. Res. Nat. Bur. Stand.* **69B** (1965), 147–153.
4. Lehman A., *A Solution of the Shannon Switching Game*. *J, Soc. Indust. Appl, Math.* **12** (1964), 687–725.
5. Minty G. J., *On the Axiomatic Foundations of Theories of Directed Linear Graphs, Electrical Networks and Network Programming*. *J. Math. Mech.* **15** (1966), 485–520.
6. Mirsky L. and Perfect H., *Applications of the Notion of Independence to Combinatorial Analysis*. *J. Combinatorial Theory* **2** (1967), 327–357.
7. Nash-Williams C. St. J. A., *An Application of Matroids to Graph Theory*. *Theory of Graphs, International Symposium (Rome)*, Paris: Dunod 1966, 263–265.
8. Rado R., *A Theorem on Independence Relations*. *Quart. J. Math. (Oxford)* **13** (1942), 83–89.

9. Tutte W. T., *On the Algebraic Theory of Graph Colorings. J. Combinatorial Theory* **1** (1966), 15–50.

10. Whitney H., *On the Abstract Properties of Linear Dependence. Amer. J. Math.* **57** (1935), 509–533.

FOR CHAPTER 9

Ore's book gives a survey of the 4-color problem up to the middle of the 60's. The first detailed loading algorithms are contained in Heesch. One also finds results on reducibility and unavoidability in Saaty-Kainen.

Heesch H., *Untersuchungen zum Vierfarbenproblem. BI-Hochschulskripten* **810/810a/810b**. Mannheim-Wien-Zürich: Bibliographisches Institut 1969.

Ore Ø., *The Four Color Problem.* New York: Academic Press 1967.

Saaty T. L. and Kainen P. C., *The Four-Color Problem.* New York: McGraw-Hill 1977.

Specialized References:

1. Allaire F., *A Minimal 5-Chromatic Planar Graph Contains a 6-Valent Vertex. Proc. Seventh Southeastern Conf. Comb., Graph Theory, Comp.* Baton Rouge: Louisiana State Univ. 1976, 61–78.

2. Bernhart A., *Six Rings in Minimal Five Color Maps. Amer. J. Math.* **69** (1947), 391–412.

3. Bernhart F., *Topics in Graph Theory Related to the Five Color Conjecture.* Doctoral Diss. Kansas State Univ. 1974.

4. Chojnacki C. A., *A Contribution to the Four Color Problem. Amer. J. Math.* **64** (1942), 36–54.

5. Errera A., *Une contribution au problème des quatres couleurs. Bull. Soc. Math. France* **53** (1925), 42–55.

6. Franklin P., *Note on the Four Color Problem. J, Math. Phys.* **16** (1938), 172–184.

7. Heesch H., *Chromatic Reduction of the Triangulations* T_e, $e = e_5 + e_7$. *J. Combinatorial Theory* **13B** (1972), 46–55.

8. Mayer J., *Inégalités nouvelles dans le problème des quatre couleurs.* *J. Combinatorial Theory* **19B** (1975), 119–149.

9. Wernicke P., *Über den kartographischen Vierfarbensatz.* *Math. Ann.* **58** (1904), 413–426.

10. Winn C., *A Case of Coloration in the Four Color Problem.* *Amer. J. Math.* **59** (1937), 515–528.

11. Winn C., *On Certain Reductions in the Four Color Problem.* *J. Math. Phys.* **16** (1938), 159–171.

FOR CHAPTER 10

The reader who has made it to here, can should venture into the original paper of Appel-Haken-Koch. An extremely interesting description of the proof and its consequences is contained in Appel-Haken.

Appel K. and Haken W., *The Four Color Problem Mathematics Today (Steene, ed.)*, Berlin-Heidelberg-New York: Springer 1978, 153–180.

Appel K. and Haken W., *Every Planar Map is Four Colorable, Part I: Discharging.* *Ill. J. Math.* **21** (1977), 429–490.

Appel K., Haken W. and Koch J., *Every Planar Map is Four Colorable, Part II: Reducibility.* *Ill. J. Math.* **21** (1977), 491–567.

Specialized References:

1. Allaire F. *Another Proof of the Four Colour Theorem I.* *Proc. Seventh Manitoba Conf. on Numerical Math. and Comp.* Univ. of Manitoba, Winnipeg 1977, 3–72.

2. Allaire F. and Swart E. R., *A Systematic Approach to the Determination of Reducible Configurations in the Four-Colour Conjecture.* *J. Combinatorial Theory* **25B** (1978), 349–362.

3. Appel K. and Haken W., *The Existence of Unavoidable Sets of Geographically Good Configurations.* *Ill. J. Math.* **20** (1976), 218–297.

4. Bernhart F., *A Digest of the Four Color Theorem.* *J. Graph Theory* **1** (1977), 207–225.

5. Dürre K., Heesch H. and Miehe F., *Eine Figurenliste zur chromatischen Reduktion. Preprint Nr.* **73**, Inst. Math. TU Hannover 1977.

6. Koch J., *Computation of Four Color Reducibility.* Ph.D. Thesis, Dept. Comp. Science, Univ. Ill. 1976.

7. Stromquist W., *Some Aspects of the Four Color Problem.* Ph.D. Thesis, Harvard Univ. 1975.

8. Tutte W. T. and Whitney H., *Kempe Chains and the Four Colour Problem. Utilitas Math.* **2** (1972), 241–281.

LIST OF SYMBOLS

The most frequently used symbols in the text are listed here in alphabetical order.

$\hat{A}$	Subspace spanned by A
$\hat{A}^{\perp}$	Orthogonal subspace
C_n	Circuit of length n
$C + D$	Sum of two sets
E	Edge set
$F, G, H, \ldots$	Countries
$G,\ G(V, E)$	Graph
$G^{\star}$	Dual graph
$\overline{G}$	Complement of G
G^m	Power of G
$G - A$	Removal of A
$G.A$	Restriction
$G \times A$	Contraction
G/e	Contraction through e
$G(S \bigcup T, E)$	Bipartite graph
$J(H)$	Directly extendable colorings
K_p	Complete graph
$K_{m,n}$	Complete bipartite graph
$K(H, C)$	Configuration
$L(G)$	Line graph
N_k	Nonorientable surface
$N(v)$	Neighborhood of v
$(N: A_1, \ldots, A_m)$,	
$(N: \mathcal{A})$,	Set system
P	Petersen graph
P_n	Path of length n
PG_m	Projective space
Q_n	The cube graph
R	Set of countries
$R(A)$	Vertices accessible from A
$R(k, l)$	Ramsey number

$R(\pi)$	π-residue
S_h	Orientable surface
V	Vertex set
W_n	The wheel
$X(\pi)$	Chromodendron
$a(G)$	Arboricity
b	Birkhoff number
$c(G)$	Number of components
$c_u(G)$	Number of odd components
$\mathrm{cr}(G)$	Crossing number
$d(v),\ d_H(v)$	Degree of v
$e(S)$	Euler characteristic
$\mathrm{ex}(H, P)$	Extremal number
$e, f, g, \ldots$	Edges
$\lVert f \rVert$	Carrier
$m(G)$	Matching number
p	Number of vertices
$p(G: \lambda)$,	
$p(\mathcal{L}: \lambda)$	Chromatic Polynomial
$\mathrm{per}(M)$	Permanent
q	Number of edges
$r(A),\ r(\Gamma)$	Rank
$u, v, w, \ldots$	Vertices
v_n	n-vertex
Γ	Chain group
$\Gamma^{\star}$	Dual chain group
$\Gamma.A$	Restriction
$\Gamma \times A$	Contraction
$\Gamma(n, k)$	k-regular bipartite graph
$\Delta(G)$	Maximal degree

$\alpha(G)$	Independence number
$\beta(\mathcal{M})$	Covering number
$\gamma(G)$	Genus
$\overline{\gamma}(G)$	Nonorientable genus
$\delta(G)$	Minimal degree
δf	Coboundary
$\kappa(G)$	Connectivity number
$\kappa_G(u,v)$	Local connectivity number
$\mu_G(u,v)$	Connecting number
$\pi(\mathcal{M})$	Packing number
$\chi(G)$, $\chi(\mathcal{L})$	Chromatic number
$\chi'(G)$	Chromatic index
$\omega(G)$	Clique number
$\mathcal{B}$	Basis
$\mathcal{B}(G(V,E))$	Bond matroid
$\mathcal{I}$	Independent set
$\mathcal{K}$	Circuit
$\mathcal{K}_0$	0-chain
$\mathcal{K}_1$	1-chain
$\mathcal{M}, \mathcal{M}(S)$	Matroid
$\mathcal{M}^*$	Dual matroid
$\mathcal{M}(\Gamma)$	Matroid of a chain group
$\mathcal{M}.A$	Restriction
$\mathcal{M} \times A$	Contraction
$\mathcal{P}(G(V,E))$	Polygon matroid
$\mathcal{V}_q$	Vector space of all 0,1-chains
∂	Boundary
$C(G)$	Cocycle group
$\mathcal{L}, \mathcal{L}(V,E,L)$	Map
$\mathcal{L}^*$	Dual map
$Z(G)$	Cycle group

INDEX

GRAPH THEORY

A Development from the 4-Color Problem

M. Aigner

BCS Associates

Moscow, Idaho, U.S.A.