SOME ASPECTS OF THE FOUR COLOR PROBLEM

A thesis presented

by

Walter Rees Stromquist

to

The Department of Mathematics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Mathematics

Harvard University

Cambridge, Massachusetts

May, 1975

Contents

The highest page numbers are x and 126. Figures are numbered consecutively throughout the chapters, but begin again with figure 1 in the appendix. The list of references on the last page refers only to the appendix.

INTRODUCTION

This thesis treats various aspects of the classical approach to the four-color problem, in which the goal is to prove that every map contains a reducible configuration. In this introduction, I will attempt to describe the relationship between my results and other current research in the same area.

In the last four years there has been a sudden increase in activity in the area of reducible configurations, sparked by the appearance of books by Ore [13, 1967] and Heesch [9, 1969] and by Shimamoto's false proof of four-color conjecture [7, 1971]. (Numbers refer to the bibliography which follows this introduction.) The results in this thesis were derived between June, 1971 and October, 1974. In a few cases, these results have now been overtaken by other research; in revisions, I have tried to reflect the new mathematics and new terminology only when it contributes directly to the ideas presented here.

This thesis contains new results in three areas. Reducible configurations -- or, more precisely, irreducible configurations -- are studied in Chapters I through III. Chapter IV presents a new existence theorem. In the appendix, a new lower bound is established for the size of any five-color map.

The first chapter contains preliminaries. Almost none of this material is new. It is included because it is hard to find in the literature, although many of the ideas can now be found in papers by Whitney and Tutte [17], Frank Bernhart [5], and E. R. Swart [16].

The key ideas in chapter I are "open sets" and "splicing." The idea of open sets has existed at least since G. D. Birkhoff's article [6, 1913], but the term was invented by Whitney and Tutte. The definition presented here, using "stand-ins," is my own. The "splicing" concept is also my own, but it has been simultaneously discovered by almost every other researcher in the field.

Before 1971, the only way to show that a configuration was not reducible was to analyze all of the potential proofs of reducibility, an impossible process for any but the smallest configurations. In 1971 the debunking of Shimamoto's "proof" [Whitney and Tutte, 17] required proving a certain form of irreducibility for a certain very large configuration (Chapter I of this thesis, figure 21). Since then, a body of irreducibility theorems has developed, discovered independently by enough researchers that they deserve to be called "folk theorems." For example, the proof of the "replacement theorem" is already implicit in Shimamoto's work; so that nobody writing later should claim priority.

My own version of these theorems, all of which are in the limited context of D-reducibility, are presented in Chapter II. The presentation in terms of "interior images" and the "replacement construction" is my own.

Chapter III develops the irreducibility theorems more fully by generalizing them to other forms of reducibility. These results are entirely my own. The useful rhetoric of "AB-diagrams" and "reducers" comes to me from Frank Bernhart.

Although the irreducibility theorems are interesting in themselves, their main value is to shed light on the question of which configurations are reducible. In this connection, the exhaustive computer analysis of 10-ring configurations by Frank Allaire and E. R. Swart [1], which has become known only in the last six months, is especially interesting. Among the configurations they have studied which are not covered by the irreducibility theorems, virtually all are reducible. This is true whether the concept used is D-reducibility or the broader 1-reducibility. In other words, the theorems of Chapters II and III represent a nearly complete answer to the question of which configurations are reducible. Allaire and Swart's analysis is summarized at the end of Chapter III.

Among the terms which I would change if I were starting again are the names of the reducibility concepts. These were selected to be consistent with Heesch [9], but they are now outmoded. The best synonym for "D-reducible" is probably "colourable," suggested by Allaire and Swart.

In Chapter IV the emphasis is shifted to the other side of the problem. Suppose we know exactly which configurations are reducible. How can we prove that every map contains one of them? In Chapter IV a class of configurations is defined which, based on the results of Chapters II and III, is an approximation to the class of reducible configurations. It is then proved, in Theorem 7 (which is the most important, as well as the most difficult result of the thesis) that every map contains a member of the new class.

This much of the idea of Chapter IV came to me through inspiring conversations with Wolfgang Haken, who, with Appel [2], was already following the same program. Beyond this point the development is my own.

The proof of Theorem 7 is extremely complicated, but it can become more manageable on the first reading if the reader is willing to make some simplifying assumptions. Specifically, he can assume that in the map under discussion, every vertex has valence 5, 6, or 7, and every third neighborhood is "orderly." One pleasant consequence is that "standard neighborhood" (as defined in the proof) means the same as "third neighborhood."

I believe the most important goal for further research in the four-color problem is to improve Theorem 7, by proving it for a more tightly drawn class of configurations providing a closer approximation to the class of reducible configurations. In this way, the entire problem could be reduced to verifying, say by computer, the reduction of a great many configurations which we would have every reason to believe could be reduced.

The appendix, entirely independent of the rest of the thesis, is a paper scheduled for publication in the <u>Journal of Combinatorial Theory, Series B</u>. It proves, by methods slightly resembling those of Chapter IV, that every map of fewer than 52 vertices may be four-colored. This result improves a recent theorem of the same kind of Jean Mayer [12] and a published result of Ore and Stemple [14].

I wish to thank Andrew Gleason, Wolfgang Haken, and Frank Bernhart for their careful reading and page-by-page critiques of earlier drafts of this material, and the editor and reviewers of the Journal of Combinatorial Theory for their advice regarding the appendix. I am also grateful to Judi Black, Janet Scruggs, and Martin Jacobs for typing drafts of various chapters, and to Debrah Johnson for typing the final copy.

BIBLIOGRAPHY

This bibliography includes only materials directly relevant to the aspects of the four color problem treated in this thesis. For a comprehensive bibliography covering other areas of the problem, see [15].

1. Allaire, F., and E. R. Swart, A systematic approach to the determination of reducible configurations in the four-colour conjecture, to appear.

2. Appel, K., and W. Haken, The Existence of Unavoidable Sets of Geographically Good Configurations, Department of Mathematics, University of Illinois, Urbana, Ill., 1974.

3. Bernhart, A., Six-rings in minimal five-color maps, Amer. J. Math. 69 (1947) 391.

4. Bernhart, F., Topics in Graph Theory Related to the Five-Color Conjecture, thesis, Kansas State University, 1973.

5. _____ Some remarks on reducibility in the five-color problem, Department of Mathematics, University of Waterloo, Waterloo, Ontario, 1974.

6. Birkhoff, G. D., The reducibility of maps, Amer. J. Math. 35 (1913) 115.

7. Haken, W., notes on Shimamoto's construction, from a seminar at the University of Illinois, 1971.

8. _____ An existence theorem for planar graphs, J. Combinatorial Theory 14 (1973) 180.

9. Heesch, H., Untersuchungen zum Vierfarbenproblem, Bibliographisches Institut, Mannheim/Vienna/Zurich, 1969.

10. _____ Chromatic reduction of the triangulations T_e, $e = e_5 + e_7$, J. Combinatorial Theory 13 (1972) 46.

11. Mayer, J., Nouvelles reductions dans le probleme des quatre couleurs, Chahiers mathematiques de Montpellier, No. 1, Universite des Sciences et Techniques du Languedoc, Montpellier, France, 1973.

12. _____ Inegalites nouvelles dans le probleme des quatre couleurs, to appear.

13. Ore, O., The Four Color Problem, Academic Press, New York, 1967.

14. _____ and J. G. Stemple, Numerical calculations on the four-color problem, J. Combinatorial Theory, 8 (1970) 65.

15. Saaty, T. L., Variations on Guthrie's four-color conjecture, Amer. Math. Monthly, 79 (1972) 2.

16. Swart, E. R., Some basic theorems on the abstract theory of Kempe-chain interchanges, mimeographed, 1975.

17. Whitney, H., and W. T. Tutte, Kempe chains and the four color problem, Utilitas Mathematica, 2 (1972) 241.

I. PRELIMINARIES

This chapter begins with our notions of maps, configurations, and colorings, and how to describe and construct them. The notion of reducibility is then introduced with an example, and we formalize the special property of D-reducibility. A key concept in this definition is the "open set" of colorings, and we finish the chapter with some fundamental lemmas on the construction of open sets.

All of our coloring problems will be in the dual form; that is, we will be coloring vertices. All of our graphs are assumed to be imbedded in a plane. We will use the standard definitions of adjacency, valence, face, subgraph, connectedness, circuit, and minimal circuit. (A minimal circuit, or ring, is a full subgraph which is a circuit. That is, no edges join the vertices of a ring except those which are necessary to make it a circuit.)

A map is a triangulation, a graph whose faces are all triangles. A configuration is a map, except that one face is not a triangle. The exceptional face is normally the outside face, and its boundary is called the boundary of the configuration. (Note that with this definition, a lone triangle is not a configuration.)

We will occasionally draw configurations in the same way as we have defined them, for example, as in figure 1. But usually, it will be easier to represent the same configuration by a "stick figure," as in figure 2. Here, only interior vertices are shown, and their valences are given. We use this notation only if the boundary may be reconstructed from these valences, so that figure 2 contains the same information as figure 1. Sometimes we will use a hybrid notation, as in figure 3.

Note: This particular configuration will reappear many times. It is called the 5-5-5 Triad, or just the Triad.

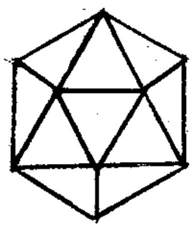


Figure 1

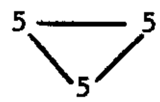


Figure 2

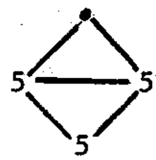


Figure 3

A proper configuration is one whose boundary is a minimal circuit. Examples of improper configurations are shown in figure 4. Note that we cannot represent these by stick figures.

Figure 4. Improper Configurations

The first neighborhood of a vertex in a graph is the subgraph consisting of the vertex and all of its neighbors (and the edges which join them). The second neighborhood is formed similarly from the first neighborhood and the neighbors of its vertices. The third neighborhood, etc., are defined similarly.

A function from the vertices and edges of one graph to those of another is a local isomorphism if, when restricted to the first neighborhood of any vertex, it is an isomorphism. A local isomorphism from a configuration to a map is a proper imbedding if it is actually an isomorphism and its image is a full subgraph; that is, no edges join the vertices of the image unless the edges are also in the image. The local isomorphism is an imbedding if its restriction to the interior of the configuration is a proper imbedding. We say that a configuration appears (properly) in a map if it may be imbedded (properly) in the map, and then we often identify the configuration with its image.

(Our definition of "imbedding" is somewhat nonstandard, in that we deliberately include situations of the type suggested by figure 5. An imbedding may map more than one boundary vertex onto the same image, if in all other respects it is a bijection on the vertices.)

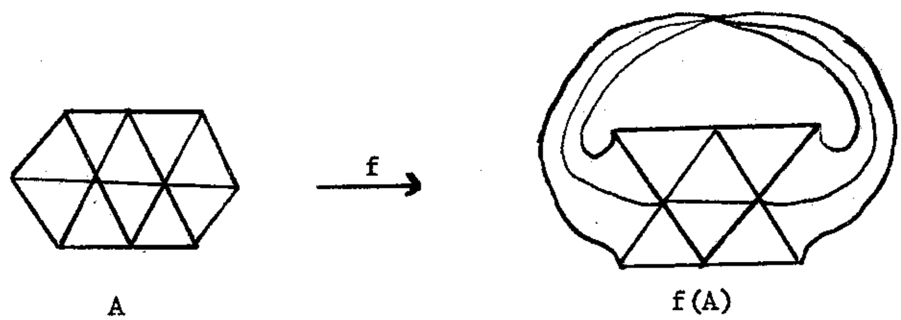


Figure 5

When a configuration A appears in a map M, there is another configuration with the same boundary, called the complementary configuration. The complement is proper if and only if A is properly imbedded in M.

Extensions. Given an abstract configuration, we may construct a new one by extension. It is necessary to designate a boundary vertex v, and an integer $n \geq 0$ called the order of the extension.

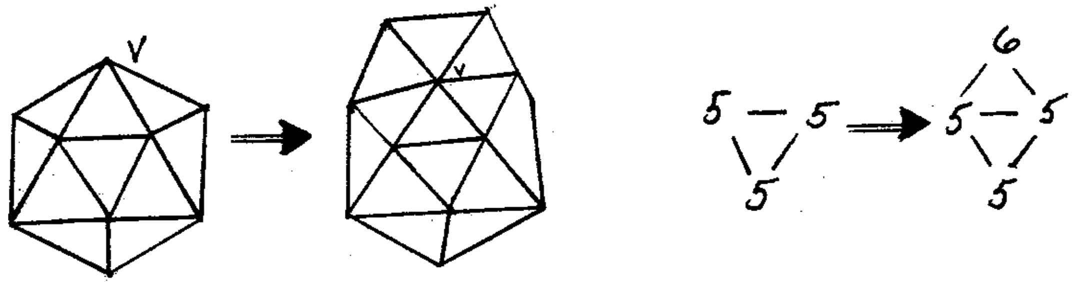
We simply add n new boundary vertices in the simplest way which converts v into an interior vertex. This process is illustrated in figure 6. An extension of order n is called an n-extension.

In the stick figures, an extension consists of adding a new interior vertex of valence k, and connecting it to ℓ consecutive vertices on the edge of the figure, in a manner consistent with the valences already shown. The order of the extension is $n = k-\ell-2$. This picture is so convenient that we will often describe an extension as "adding a vertex" of a given valence to a configuration, although what we are really doing is enclosing a vertex which was previously on the boundary of the configuration.

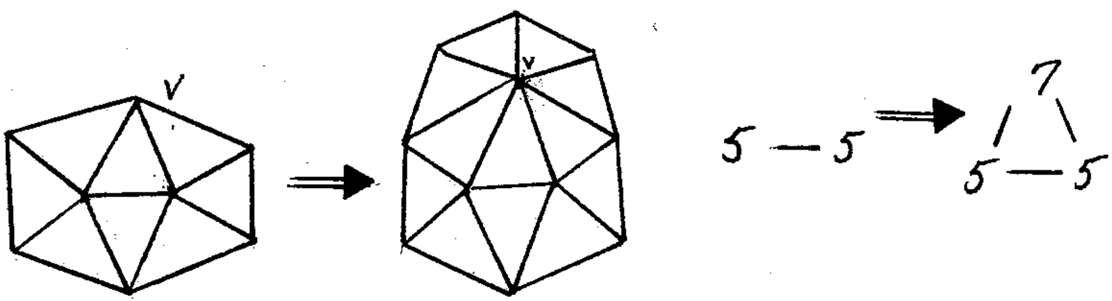
We will also use some terminology introduced by Haken and Heesch. An m-legger is an interior vertex in contact with exactly m boundary vertices. Thus, an n-extension "adds" an m-legger with $m = n+2$.

Stick figures give rise to a common shorthand by which some configurations can be described by listing the valences of their interior vertices. For example, "5:6755" describes a 5-vertex, four of whose neighbors have valences 6, 7, 5, 5 in order. Note that this representation is rarely unique; thus 5:565 and 6:555 are identical. See figure 7.

A 2-extension:



A 3-extension:



A 0-extension:

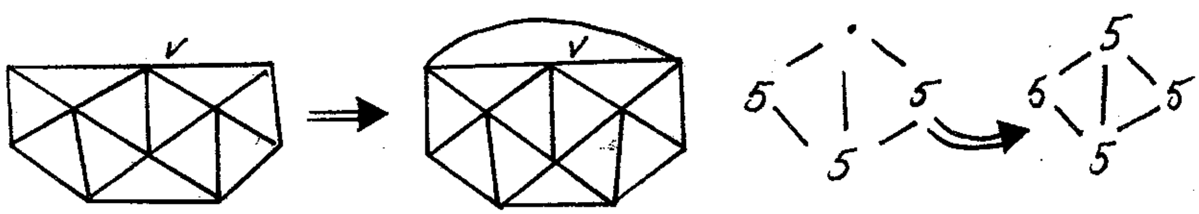
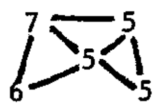
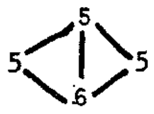


Figure 6. Extensions



5:6755



5:565
6:555

Figure 7

Articulation Points. An articulation point in a configuration A is an interior vertex whose removal disconnects the interior into two or more components. From each such component we may construct a new configuration including the component and the vertices of A adjacent to it. The new configuration is called a wing of A. This process is easiest to understand in terms of stick figures, since it amounts to just erasing the articulation point (figure 8).

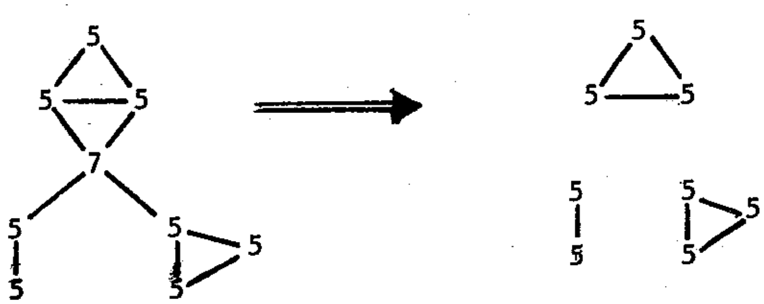


Figure 8. A configuration and its wings

An articulation may be classified as twofold, threefold, etc., according to the number of wings. It may also be called an m-legger. For example, figure 8 illustrates a "threefold 3-legger articulation point."

A coloring of a graph is a partition of its vertices into four parts, or colors, so that no two vertices of the same color are adjacent. The four-color conjecture is that all maps have colorings. We have the notions of restriction of a coloring to a subgraph, and extension of a coloring from a subgraph to a graph (although the latter need not exist nor be unique).

Given a graph, a coloring, and a pair of colors, say α and β, we can partition the graph into chains as follows. An $\alpha\beta$-chain is a connected subgraph which is maximal subject to all of its vertices having colors α or β. Defining $\gamma\delta$-chains similarly, the graph is partitioned into its $\alpha\beta$- and $\gamma\delta$-chains. This partition is the chain structure, which may of course be different if we start with a different coloring or a different pair of colors in place of α,β.

In this setup, suppose we select some of the $\alpha\beta$- and $\gamma\delta$-chains, and reverse the colors within each of the chains we have selected. Thus α becomes β, γ becomes δ, etc. The result is a new coloring with the same chain structure. This process is a simple chain-switch, and it is the only process we will use for making new colorings from old.

Reducibility. Roughly, a configuration is "reducible" if we can prove that it does not appear in any minimal five-color map. The configuration we will use to illustrate this idea is the "Birkhoff diamond", A, in figure 9.

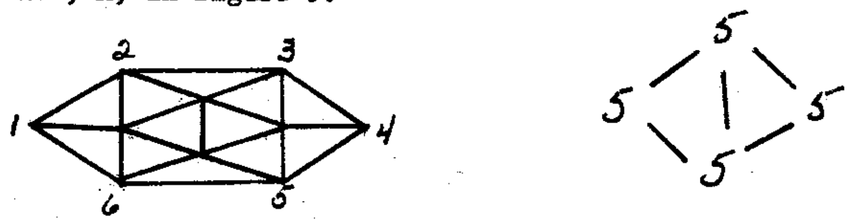


Figure 9. The Birkhoff Diamond

Suppose this configuration is properly imbedded in a minimal five-color map, M. Then the complement is a proper configuration, B. We form a new map M', consisting of the vertices of B but with vertices 1 and 5 (as numbered in figure 9) merged into one vertex, and vertices 2 and 4 merged into one vertex. Since M' has fewer vertices than M, it must have a coloring. This implies that B has a coloring in which vertices 1 and 5 have the same color, and vertices 2 and 4 have the same color. There are only five such colorings of the boundary ring; we know that one of them may be extended to B.

(This method, by which we limit our attention to a few colorings of the boundary, is called an "amalgamation argument." It would not have worked if we had not assumed that the original imbedding was proper. For example, if vertices 1 and 5 were joined by an edge within B, the graph M' would contain a loop and would not be a map.)

Now by experimenting, we find that all of these five colorings of the boundary extend to A except one, shown in figure 10. If M has no coloring, this must be the one which has an extension to E. We will modify this (extended) coloring by means of simple chain switches based on the color-pair $\alpha\gamma$, in such a way that the modified coloring may also be extended to A. This will provide a coloring for M, and hence, the desired result.

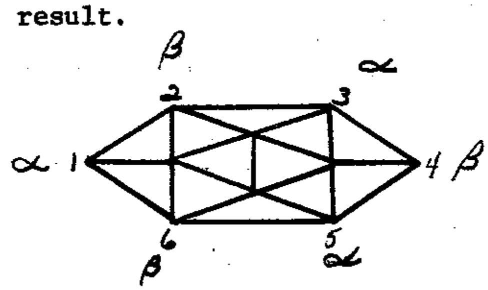


Figure 10

Case 1. There is no βδ-chain joining vertices 2 and 6. In this case vertex 6 may be changed to δ, possibly also changing 4 to δ. In either case, the resulting coloring may be extended to A.

Case 2. There is no αγ-chain joining vertices 3 and 5. This is symmetrical to Case 1.

Case 3. Both of these chains exist. In this case it is clear that 1 may not be in an αγ-chain with 3 and 5, for then an αγ-chain and a βδ-chain would cross within B, an impossibility. Similarly, 4 may not be in the βδ-chain of 2 and 6. Thus, we make the simple chain-switch involving the chains of 1 and 4, to obtain the coloring shown in figure 11, which extends to A.

Each case contradicts the assumption that M has no coloring.

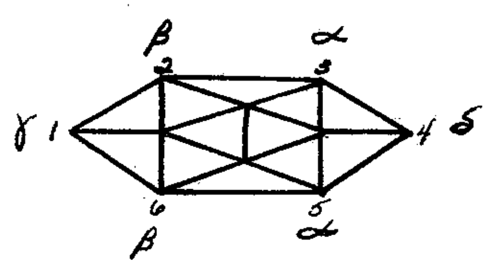


Figure 11

We have proved that A is "reducible"; it may not appear properly in a minimal five-color map.

The last step in this proof is called a "chain argument." One analysis of this argument is as follows: we took a hypothetical set of colorings of the boundary of A; and we assumed that included the coloring of figure 8 but did not include any coloring extendable to A. Then we proved that could not be precisely the set of colorings extendable to B. For this proof, B was an entirely arbitrary configuration. In other words, U lacked some property which it would need if it were the set of colorings extendable to a configuration. We will soon isolate this property in the concept of an "open set."

Note that the chain argument did not require that B be a proper configuration, a requirement that was imposed by the amalgamation argument. So we might try to dispense with the latter and rely only on the former; that is, prove that <u>every</u> coloring of the boundary of A can be extended to A, or can be altered so as to be extendable by using chain arguments. This is a big job even for such a small configuration as the Birkhoff Diamond, but it can be completed. It follows that this configuration may not appear in a minimal five-color map, properly or not.

When "reducibility" can be proved using only chain arguments, we call it "D-reducibility." This is the only form of reducibility which we will develop in this or the following chapter.

Definition of Open Set

Basically, an open set is a set of colorings of an abstract ring.

We begin with a special case. Given a configuration A, the <u>open set associated with A</u>, U_A, is the set of colorings of the boundary of A which can be extended to A. For example, if A is the "Pentagon" (figure 12), then U_A is the set of colorings of the 5-ring which use only three colors.

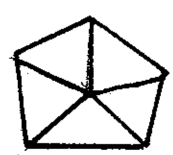


Figure 12. The Pentagon

Note: Suppose $c \epsilon U_A$ is a coloring. Then c may be extended to some coloring $\hat{c}$ of A. Suppose we choose a pair of colors and make a simple chain-switch in $\hat{c}$ to obtain a new coloring $\hat{d}$ of A. Let d be the restriction of $\hat{d}$ to ∂A. Then naturally, $d \epsilon U_A$. We will use this pattern as the framework for the general definition.

Let R be a ring and let U be a set of colorings of R. Then U is an <u>open set</u> if the following condition is met: For every coloring $c \epsilon U$, and every pair $\alpha\beta$ of colors, there is a configuration $A_{c,\alpha\beta}$ whose boundary may be identified with R, and an extension $\hat{c}$ of c to $A_{c,\alpha\beta}$, such that if we make any simple chain-switch (involving only $\alpha\beta$- and $\gamma\delta$-chains) in $\hat{c}$ to produce $\hat{d}$, then the restriction d of $\hat{d}$ to R is in U.

The configurations $\{A_{c,\alpha\beta}\}$ together play the same role as A did in the special case. Their existence suggests that for any open set U, there <u>might</u> be a configuration A which has U as the associated open set. In this case the $\{A_{c,\alpha\beta}\}$ may be viewed as stand-ins for A, helping to reveal its chain structures. It is useful to think of every open set as the associated open set for some hypothetical configuration.

(Most authors use some kind of "abstract chain structure" in place of the $\{A_{c,\alpha\beta}\}$. Our definition is an improvement for two reasons; it saves us from defining abstract chain structures, and it clarifies the upcoming Splicing Lemma.)

An open set which is associated with some configuration is called <u>realizable</u>.

Two nonempty, disjoint open sets on the same ring are called complementary. An open set is complementary to a configuration A if it is complementary to U_A. If U has no complement, it is called D-reducible. A configuration A is also D-reducible if it has no complementary open set. If A does have a complementary open set, it is D-irreducible.

Special Case. If A is an improper configuration, its boundary A is not a ring, but we may construct an abstract ring R which corresponds to ∂A in an obvious way. See figure 13. Now the open set U_A may be considered as an open set on R as well as on ∂A. Viewed as an open set on R, U_A has a nonempty complementary open set, consisting of all colorings of R which are not colorings of ∂A. For this reason, as a convention, we say that an improper configuration is never D-reducible.

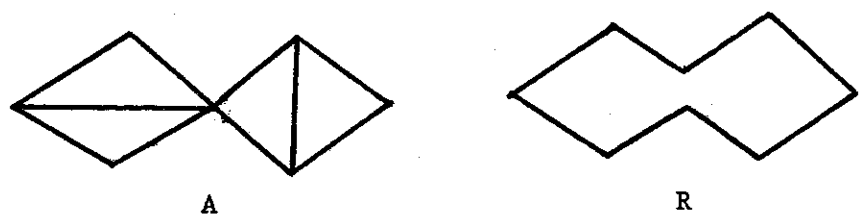


Figure 13. An abstract ring representing the boundary of an improper configuration

Suppose there exists a minimal five-color map, and let A and B be complementary configurations having common boundary R. Then U_A and U_B are complementary open sets on R. Thus neither A nor B is D-reducible. A D-reducible configuration may not appear in a minimal five-color map.

Examples of Open Sets. Let P be the Pentagon (figure 12). Then U_P is the set of colorings of the 5-ring which involve only three colors. Let W_P be the set of colorings of the 5-ring which involve exactly four colors. It can be shown that W_P is an open set, complementary to P. Therefore P must be D-irreducible.

Let T be the Triad (figure 2). Let W_T be the set of the colorings of the boundary of T shown in figure 14, and the 120° rotations of these colorings. All of the colorings which will not extend to T are in this set, except for the one shown in figure 15 and its rotations. Then W_T is an open set, the unique open set complementary to T; and T is D-irreducible. (The open set W_T was first reported by G. D. Birkhoff in 1913.)

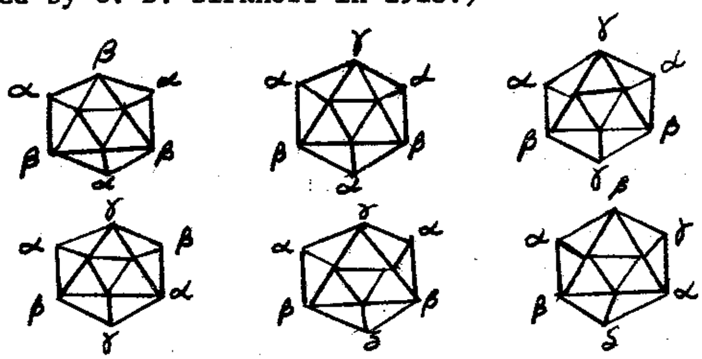


Figure 14. Colorings in W_T

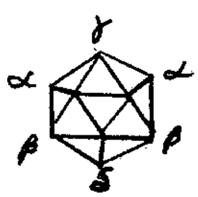


Figure 15. A coloring not in W_T

Almost all of the open sets we will meet in this paper are made from W_P, W_T, and various realizable open sets, using the Union Lemma and the Splicing Lemma.

<u>Union Lemma</u>. The union of open sets is an open set.

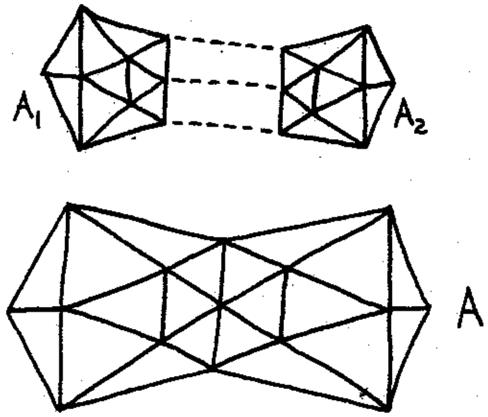
The proof is immediate from the definition. //

<u>Splicing</u>. We will define splicing for (1) configurations and their associated open sets, (2) colorings, and (3) arbitrary open sets.

Let A_1 and A_2 be configurations which have some consecutive boundary vertices in common, as in figure 16. Their union, the configuration A, is called a <u>splicing</u> of A_1 and A_2. Likewise, U_A is called a splicing of U_{A_1} and U_{A_2}. Note that U_{A_1} and U_{A_2} are defined on different rings, and U_A on a third ring.

Now suppose two rings R_1 and R_2 share some consecutive vertices, and that a third ring R is formed from some of their vertices, as shown in Figure 17. Let c_1 and c_2 be colorings of R_1 and R_2, respectively. If c_1 and c_2 agree on the intersection of R_1 and R_2, they define a coloring c of R, which we call a splicing of c_1 and c_2. (In one peculiar case, in which c_1 and c_2 have only two colors on $R_1 \cap R_2$, the coloring c is not uniquely defined. In this case both colorings c are called splicings of c_1 and c_2.)

Let U_1 be a set of colorings of R_1, and U_2 a set of colorings of R_2. The <u>splicing</u> U of U_1 and U_2 is a set



<u>Figure 16. Splicing of Configurations</u>

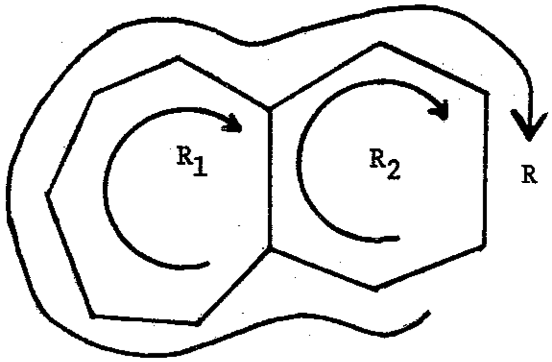


Figure 17. Ring setup for Splicing

of colorings of R, defined as follows: A coloring c is in U if it is a splicing of some $c_1 \varepsilon U_1$ and $c_2 \varepsilon U_2$.

Now U is always uniquely defined by U_1 and U_2, and the arrangement of the various rings. All of our splicings will be made with reference to diagrams.

The following lemma follows directly from our definition of open set.

Splicing Lemma. A splicing of open sets is an open set.

Proof: Let R_1, R_2, and R be rings arranged as in figure 17. Let U_1 and U_2 be open sets on R_1 and R_2, and let U be their splicing. Let c be a coloring in U, obtained by splicing $c_1 \varepsilon U_1$ and $c_2 \varepsilon U_2$. Let $\alpha\beta$ be a color-pair.

We must find a configuration $A = A_{c,\alpha\beta}$ whose boundary may be identified with R, and an extension of c to a coloring $\hat{c}$ of A, such that any simple chain-switch in $\hat{c}$ produces a coloring $\hat{c}'$ whose restriction to R, c', is in U.

Since U_1 and U_2 are open sets, we can find configurations A_1 and A_2, and colorings $\hat{c}_1$ and $\hat{c}_2$ extending c_1 and c_2, having the same property with respect to R_1 and R_2. Let A be the splicing of the configurations A_1 and A_2. Since $\hat{c}_1$ and $\hat{c}_2$ agree on the intersection of R_1 and R_2, we may use $\hat{c}_1$ and $\hat{c}_2$ to define the coloring $\hat{c}$ of A which extends c.

Now any simple chain switch in $\hat{c}$ (producing $\hat{c}'$, which restricts to c') induces simple chain-switches in $\hat{c}_1$ and $\hat{c}_2$ (producing $\hat{c}_1'$ and $\hat{c}_2'$, which restrict to c_1' and c_2'). Now $c_1' \varepsilon U_1$, and $c_2' \varepsilon U_2$, by definition, and c' is a splicing of c_1' and c_2'. Therefore c' is in U as desired.//

An example will show the usefulness of splicing, and at the same time introduce the irreducibility theorems of the next chapter.

Let R_1 and R_2 be 5-rings, sharing three vertices as shown in figure 18. Let W_1 and W_2 be open sets on these rings, each identical to the open set W_P from the previous example. Then their splicing, W, is an open set on the 6-ring R.

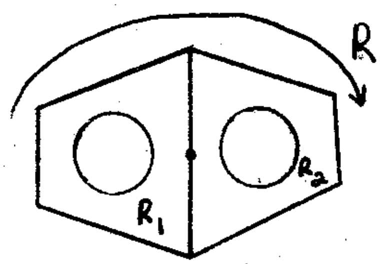


Figure 18

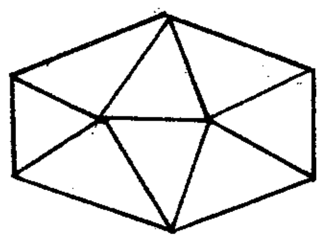


Figure 19

It is an easy exercise to show that none of the colorings in W will extend to the configuration A shown in the similarly-oriented figure 19. Thus, W is an open set complementary to A, and A, the double pentagon, is D-irreducible.

We can get more mileage from the splicing lemma if we define open sets by "diagrams." A <u>diagram</u>, D, is a graph with a designated boundary, and for each interior face which is not a triangle, a designated open set on that face. Figure 18 is an example of a diagram; so is figure 20, below. (In each case the symbol  indicates the open set W_P.)

A coloring of the boundary ∂D <u>extends</u> to D if it extends to the graph D in such a way that the coloirng of each face is included in the open set assigned to that face. From repeated application of the splicing lemma, we know that the set of colorings which extend to D is an open set, the <u>associated open set</u> of D.

The power of diagrams is shown by an example. Consider the configuration H of figure 21, below. This is the "Shimamoto Horseshoe," whose supposed D-reducibility caused great excitement recently. But we can show that it is not D-reducible by a reference to the diagram D of figure 20. It is a tedious exercise to show that no coloring extends to both D and H; but once this is done, we know that D defines an open set complementary to H. (This proof appeared in a paper by Whitney and Tutte.)

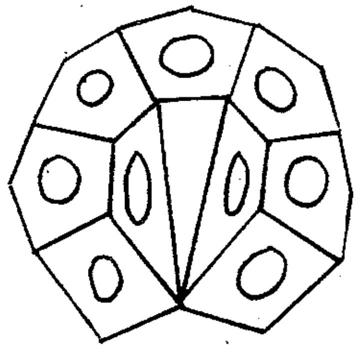


Figure 20

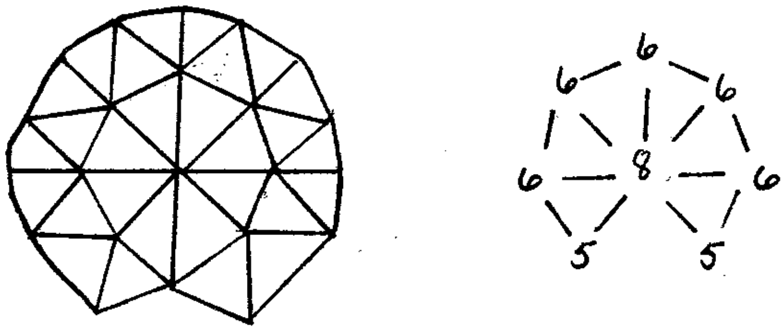


Figure 21

Improper Diagrams. A diagram, like a configuration, is improper if the boundary is not a ring. In particular, a diagram may be used to define an open set on a ring even if two vertices in the ring correspond to a single vertex in the diagram. For example, the diagram in figure 22 defines an open set on a 9-ring. When this occurs, we use a special convention: a double arc (or "elongated equal sign") between boundary vertices of a diagram implies that the vertices are merged into one. Thus, figure 23 represents the same diagram as figure 22.

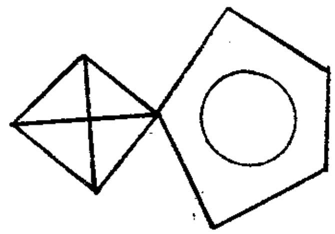


Figure 22

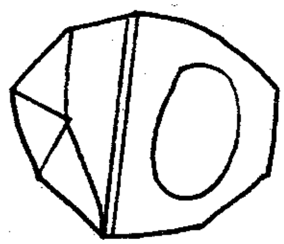


Figure 23

II. THEOREMS ON D-IRREDUCIBILITY

The object of this chapter is to find configurations which are not D-reducible. We already know a few, such as the Pentagon and the 5-5-5 Triad, whose D-irreducibility was established in the last chapter by listing their complementary open sets. Now we will use these open sets, and the splicing lemma, to construct open sets complementary to other configurations.

To apply the splicing lemma we will use many diagrams, for which we will need some standard symbols, shown in figures 24 and 25. Figure 24 represents the open set W_P complementary to the Pentagon, which consists of all colorings of the 5-ring which use all four colors. Figure 25 shows the 5-5-5 Triad and the open set W_T complementary to it; they are shown side by side in order to fix the orientation. An empty space in a face of a diagram will represent the universal open set, consisting of all the colorings of the ring.

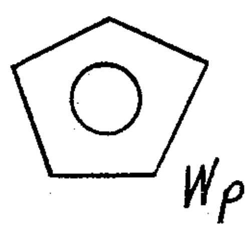


Figure 24

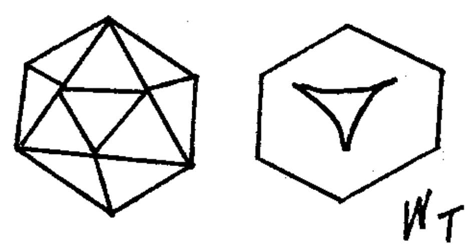


Figure 25

Interior images. Sometimes a large configuration contains within it an "image" of a smaller configuration, and we can deduce the D-irreducibility of the larger configuration from the known D-irreducibility of the smaller one.

Let B and A be proper configurations, and let C be a subconfiguration of B. Suppose that there is a local isomorphism $f:C\to A$ such that:

(1) if f maps v onto a boundary vertex A, then v is a boundary vertex of B (not only of C);

(2) f is surjective except that some boundary edges (and the corresponding faces) of A may not be in the image of f; and

(3) f is injective except that two boundary vertices of C (which are not adjacent in B) may be mapped onto any one boundary vertex of A.

It follows that f is an isomorphism on the interior of C, and each interior vertex of C adjoins as many boundary vertices in B as its image does in A. See figure 26. In this case we say that C is an interior image of A in B.

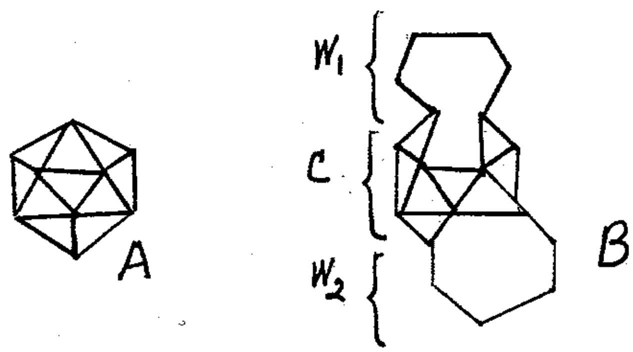


Figure 26

Removing C divides B into one or more components. For each such component, consider its "first neighborhood" in B -- that is, the full subconfiguration formed from the vertices of the component and their neighbors in B. The subconfigurations obtained in this way are called the wings of B (with respect to the interior image C). Now B may be formed by successive splicings of C with the wings.

Several examples of interior images are shown in figure 27. Note that an interior image of a Pentagon is just a 5-legger. An interior image of a 5-5-5 Triad is an interior triad.

Lemma. Suppose that B contains an interior image of A, and let B' be a proper configuration contained in B. Then either B' is a subconfiguration of one of the wings of B, or B' contains an interior image of some proper configuration A' contained in A.

Proof: Let C be the image and let $f:C\to A$ be as above. Suppose B' is not contained in any of the wings. Then let $C' = B' \cap C$, and let $f':C'\to A$ be the restriction of f. Let A' be the image $f'(C')$, together with any edges of A which were left out of $f(C)$ but whose

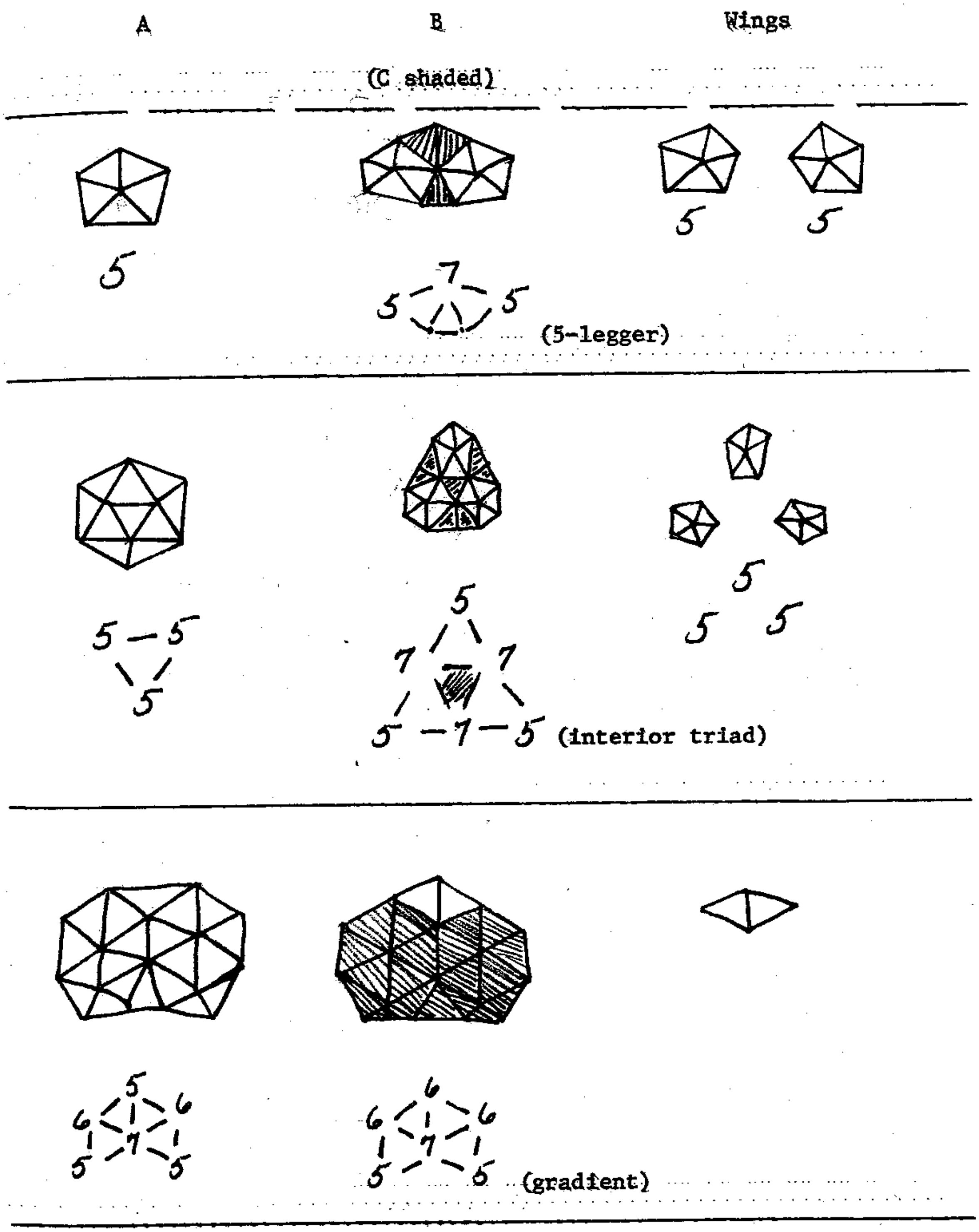


Figure 27. Interior images

endpoints are in $f'(C')$. Then A' must be a proper subconfiguration of A, and $f':C' \to A'$ must meet all of the requirements of the definition, making B' contain an interior image of A'. //

We are now ready for the first of the irreducibility theorems.

Theorem 1. Let B contain an interior image of A, and suppose that A is D-irreducible. Then so is B. Furthermore, if neither A nor any of the wings of B contains a D-reducible subconfiguration, then neither does B.

Proof: Let C be the image, and let $f:C \to A$ be as above. Let W_A be an open set complementary to A. We will construct an open set W_B complementary to B, and conclude that B is D-irreducible.

The construction of W_B is shown by example in figure 28, which is a continuation of figure 26. W_B is defined by an improper diagram, or equivalently, by splicing W_A with some universal open sets. It is clearly a non-empty open set. It is complementary to B because any coloring of W_B extended to B would define, via the correspondence established by f, an extension of a coloring of W_A to A.

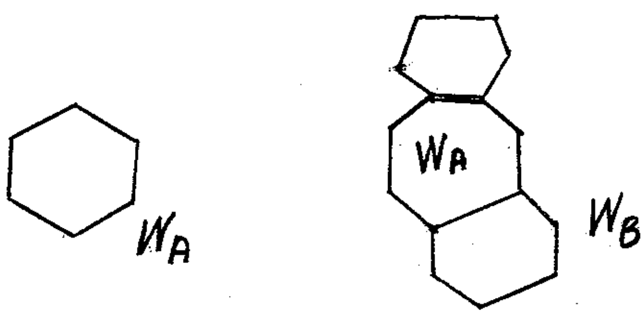
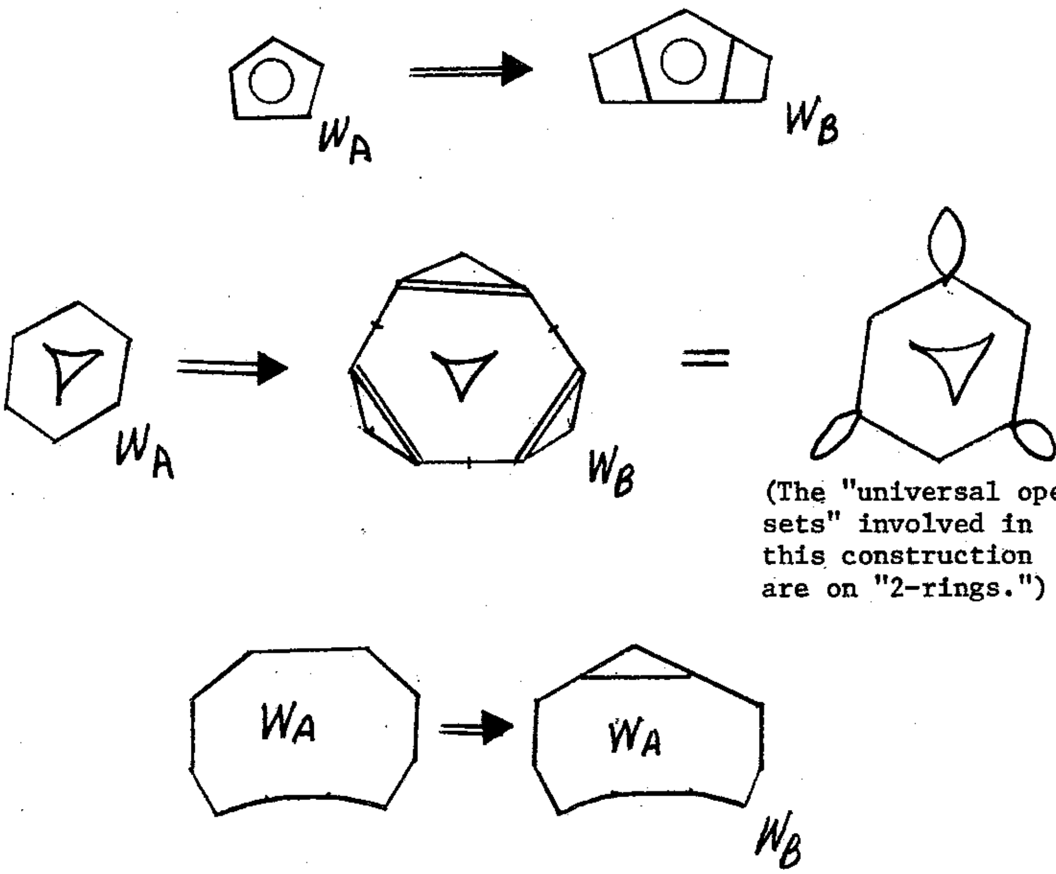


Figure 28

For the second assertion of the theorem, let B' be any subconfiguration of B. Every improper configuration is necessarily D-irreducible. If B' is proper, its D-irreducibility follows from what we have done by the previous lemma. //

For specific examples of this construction, see figure 29, which is a continuation of figure 27 and also illustrates the following corollaries.



(The "universal open sets" involved in this construction are on "2-rings.")

Figure 29

<u>Corollary (5-legger theorem)</u>: Any configuration containing a 5-(or more-) legger is D-irreducible. //

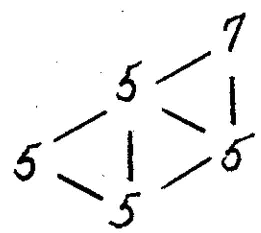
Recall that a 3-extension creates a 5-legger.

<u>Corollary (interior triad theorem)</u>: Any configuration containing an interior triad is D-irreducible. //

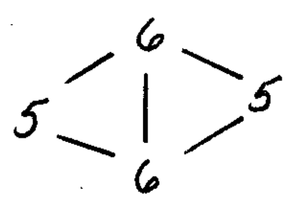
Suppose we pick an interior vertex of A which is at least a 2-legger, and increase its valence by giving it a new boundary neighbor. The result is called a <u>gradient</u> of A. See, for example, the last case in figure 27. The gradient contains an interior image of the original configuration.

<u>Corollary (gradient theorem)</u>: A gradient of a D-irreducible configuration is D-irreducible. //

All of the useful applications of Theorem 1 are covered by its corollaries. A few of these applications are shown in the next three figures. Figure 30 contains a 3-extension and is therefore D-irreducible, even though it contains a D-reducible subconfiguration. Figure 31 is D-irreducible because of its two "interior triads," but it is known to be reducible in another sense, as will be seen in the next chapter. Figure 32 actually contains four interior triads.



<u>Figure 30. D-irreducible</u>



<u>Figure 31. D-irreducible (C*-reducible)</u>

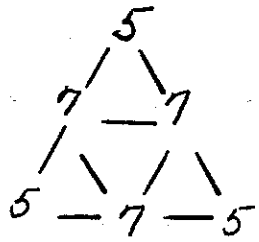


Figure 32. D-irreducible

We can now come to the most important theorem of the chapter.

Theorem 2. (The 2-extension theorem.) Suppose that a configuration B is obtained by a 2-extension of A. Then B is D-reducible if and only if A is D-reducible.

Proof: First we will assume that A is D-irreducible, and from an open set complementary to A, we will construct one complementary to B. Let B and A be as shown in figure 33, and let W_A be an open set complementary to A. Then define W_B by the splicing diagram of figure 34.

We will show by contradiction that W_B is complementary to B. Suppose that some coloring of the boundary is extendable both to B and to the diagram defining W_B. Then exactly three colors must appear on v_1, v_2, v_3, v_4 -- if there were four, v_0 could not be colored in B; and if there were two, v_5 could not be colored in figure 34.

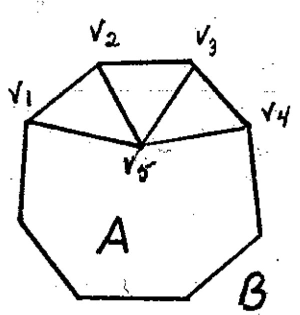


Figure 33

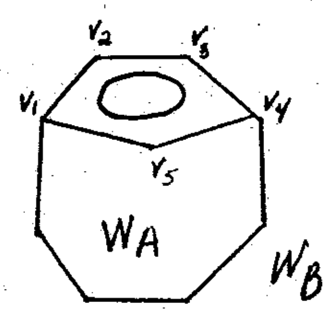


Figure 34

The fourth color must be assigned to both v_0 and v_5. But now to extend the coloring further to both B and W_B would require extending identical colorings to both A and W_A, an impossibility. Therefore W_B is complementary to B, and B is D-irreducible.

To prove the converse, suppose W_B is given, and use the splicing diagram in figure 31 to construct W_A (the face marked "W_A" is considered the "outside face" in this diagram).

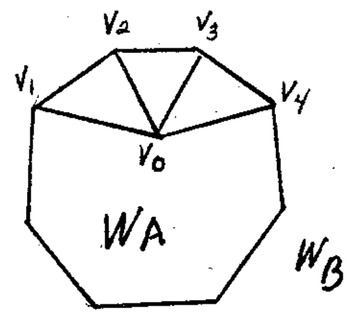


Figure 35

Clearly W_A contains no colorings which extend to A, so to show that W_A is a complementary open set to A requires only that W_A not be empty. To show this, it suffices to show that W_B must contain at least one coloring with less than four colors on v_1, v_2, v_3, v_4. This follows from a lemma, which will complete the proof of the theorem.

Lemma. If v_1, v_2, v_3, v_4 are any four vertices on a ring R and U is a nonempty open set of colorings of this ring, then U contains a coloring in which v_1, v_2, v_3, v_4 are not all colored differently.

Proof: Let $c \epsilon U$. We may assume that v_1, v_2, v_3, v_4 appear in that cyclical order on R, and have the colors α, β, γ, δ, respectively in the coloring c. According to the definition of open set, choose a configuration $A = A_{c,\alpha\beta}$ whose boundary may be identified with R, and an extension $\hat{c}$ of c to A. In A, v_1 and v_3 may be joined by an $\alpha\gamma$-chain, or v_2 and v_4 may be joined by a $\beta\delta$-chain, but not both. Suppose the former does not occur. Then we may change the color of v_1 to γ by a simple chain-switch in $\hat{c}$ which does not affect v_3; and the resulting coloring of R must be in U but does not have four different colors on v_1, v_2, v_3, v_4. //

(The constructions used to prove the two parts of Theorem 2 are not quite inverses of each other.)

Note: The construction in figure 18, used there as an example of splicing, is also an example of the 2-extension theorem.

Applications of the 2-extension theorem are shown in figures 36-40. Each of these is built up from a Pentagon or Triad by a succession of 2-extensions, and is therefore D-irreducible.

We note in passing that if a 1-extension is made at any point in figures 38-40, the result is D-reducible. This suggests that not only is the quality of D-reducibility preserved by 2-extensions, but also some quality of "almost-reducibility" which may someday be made more precise.

Figure 36

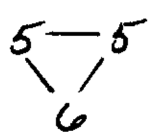


Figure 37

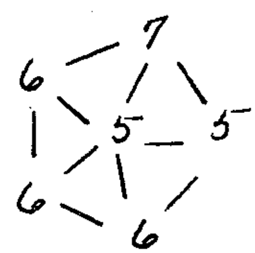


Figure 38

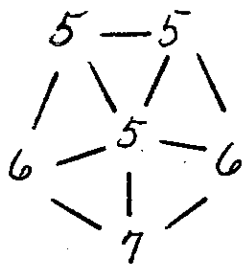


Figure 39

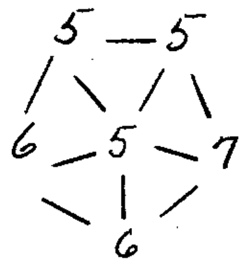


Figure 40

The replacement construction. Let C and A be configurations, and let v_0 be an interior vertex of C adjoining three consecutive boundary vertices v_1, v_2, v_3. Remove v_2 (and the edges incident to it) from C, and construct a new configuration B by splicing A with the remainder of C, in such a way that the common boundary in the splicing consists of v_1, v_2, and v_3. See figure 41. We call this the replacement construction, and we say B is obtained by replacing a boundary vertex of C with a copy of A.

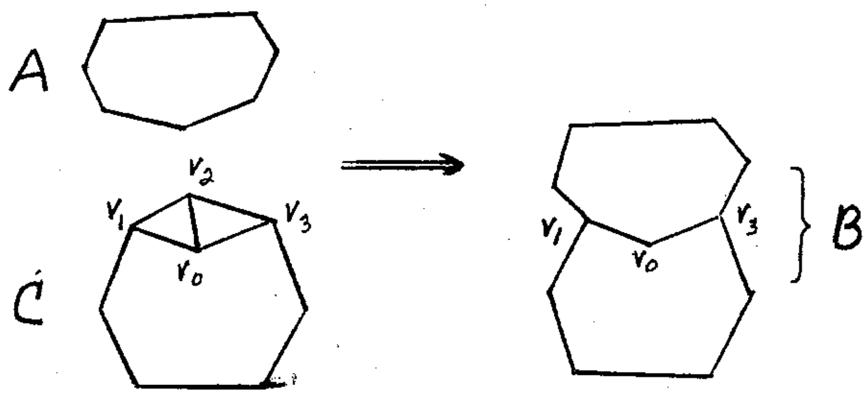


Figure 41. The replacement construction

Note that every interior vertex in B corresponds to exactly one interior vertex of C or A. This correspondence (which we will regard as an identification) preserves valences, except for v_0, which may have a higher valence in B than in C. We may also consider v_0 as a boundary vertex of A.

If C has interior vertices other than v_0 and A has interior vertices, then v_0 is an articulation point of B. If C is a Pentagon and v_0 is the unique interior vertex of C, then B is just a 2-extension of A; so the replacement construction is a generalization of extensions. Examples are in figures 42 and 43.

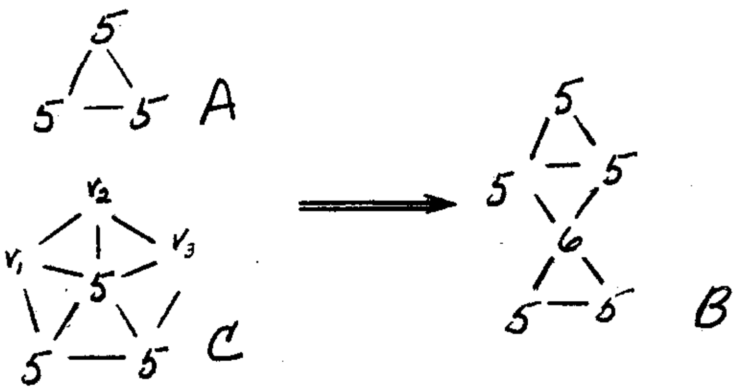


Figure 42. A replacement creating an articulation point

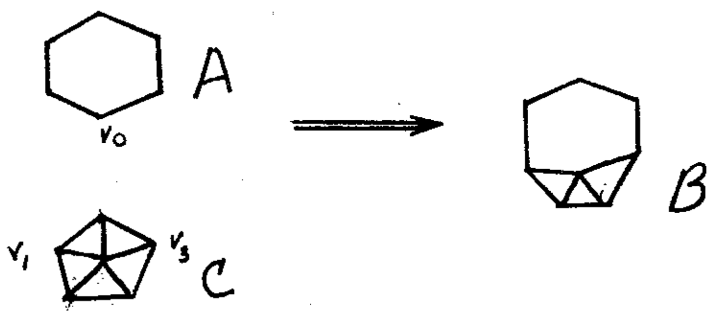


Figure 43. A 2-extension as a replacement construction

Lemma. Let B be obtained by replacing a boundary vertex of C with a copy of A, and let B' be a proper configuration contained in B. Then either B' is a subconfiguration of C or A, or B' is obtained by replacing a boundary vertex of some sub-configuration C' of C with a copy of some subconfiguration A' of A.

Proof: If B' is proper and is not contained entirely in C or A, then B' must include v_1, v_0, and v_3, and at least one other vertex in each of C and A. Let $A' = B' \cap A$, and let C' be formed from $B' \cap C$ by replacing v_2 in the way that it appears in C. Then B' is formed from C' and A'. //

The next theorem generalizes part of Theorem 2, and yields two important corollaries on articulation points.

Theorem 3 (replacement theorem). Let B be obtained by replacing a boundary vertex of C with a copy of A. If C and A are D-irreducible, then so is B. Furthermore, if neither C nor A contains a D-reducible subconfiguration, neither does B.

Proof: The general situation is pictured in figure 44. Here we assume that C and A are D-irreducible; so we may let W_C and W_A be their respective complementary open sets. Figure 45 defines a splicing of these open sets to form W_B, an open set which we assert must be complementary to B. For suppose some coloring extends to both B and W_C. If v_0 has the same color as v_4, we are extending the same coloring to both A and W_A, an impossibility. But if v_4 is colored differently from v_0, then we can color C, by coloring those vertices which are in B in the same way they are colored in B, and assigning to v_2 the color of v_4. The result is to extend the same coloring to C and to W_C.

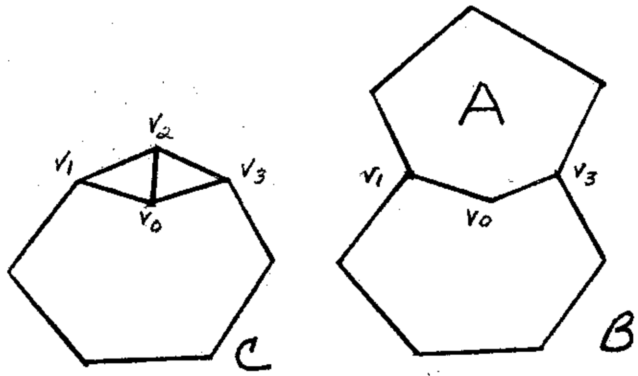


Figure 44

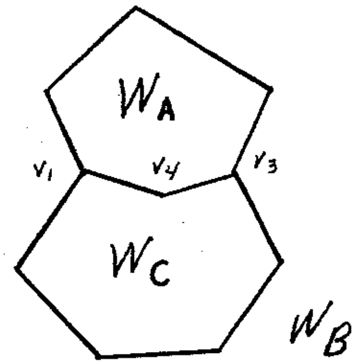


Figure 45

This completes the proof if $W_B \neq \emptyset$. This is clear if W_A includes at least one coloring for which v_1 and v_3 agree, and one coloring for which they disagree; for then any coloring in W_C may be spliced (possibly after permuting the colors) with one of these two colorings in W_A. This is the normal case.

But suppose, on the contrary, that every coloring in W_A assigns the same color to v_1 and v_3. Then it is possible for W_B to be void; so we form another open set V_B by replacing W_C in figure 45 with a universal open set. Now V_B is obviously nonempty; the problem is to show that it is complementary to B. For this we will need a chain argument in the open set W_A, which we will indicate breifly: for every coloring in W_A, v_1 and v_3 are joined by every kind of chain. Thus, the color of v_4 may be changed at will, and the coloring will still be in W_A. So extending a coloring to B and V_B always requires extending identical colorings to A and W_A.

Similarly, if every coloring of W_A assigns different colors to v_1 and v_3, then the color of v_4 is irrelevant to a determination of whether a coloring is in W_A; so the same open set V_B is a nonempty complementary open set to B.

This proves the first conclusion of Theorem 3; the second now follows as well, because of the previous lemma. //

We now apply Theorem 3 to the problem of articulation points.

Corollary 1 (3-legger articulation points): Suppose a configuration A has an articulation point that adjoins three or more boundary vertices, and suppose that two of the wings are D-irreducible. Then so is A.

Proof: If the two irreducible wings are each replaced by single boundary vertices, the resulting configuration is necessarily D-irreducible, since it contains a 5-(or more-) legger. Reversing this process amounts to two applications of Theorem 3, so A itself must be D-irreducible. //

Since each wing of a configuration includes the original articulation point as a boundary vertex, it makes sense to consider a 1-extension of the wing at the articulation point. Frequently, even if the wing is irreducible, it becomes reducible after the 1-extension. When this is not the case, we can improve on Corollary 1. The following is a direct translation of Theorem 3:

Corollary 2 (2-legger articulation points): If an articulation point divides a configuration A into two wings, both of which are D-irreducible and one of which remains so after a 1-extension at the articulation point, then A is D-irreducible. //

The nicest application of this corollary is to the case of "hanging pairs" of 5-vertices. Consider figure 46. The lower wing is D-irreducible, and the upper wing remains D-irreducible after the prescribed 1-extension. Therefore Corollary 2 applies and the configuration is D-irreducible. The upper wing of this configuration is called a "hanging pair."

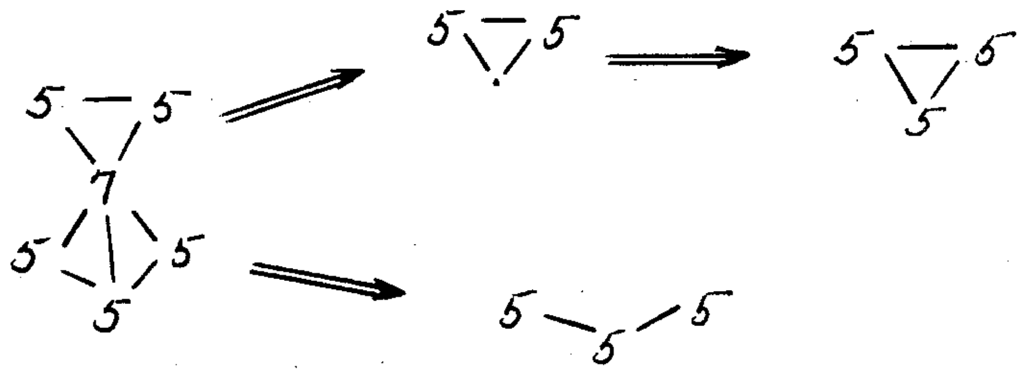


Figure 46. D-irreducible (Corollary 2)

More examples appear in figures 47-49. Figure 47 requires two applications of the replacement theorem, and contains two articulation points.

Each of the configurations in figure 48 is obtained by replacing a boundary vertex of 6:565 with a Triad. Since 6:565 contains two interior triads, it must be D-irreducible; therefore, so are the configurations in the figure.

Figure 49 illustrates the limits of the articulation point corollaries. This configuration barely fails to meet the hypotheses of either corollary; and in fact, it is D-reducible.

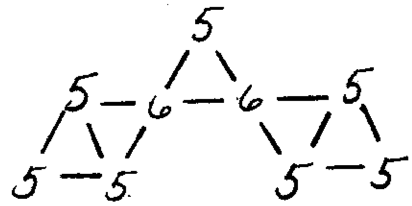


Figure 47. D-irreducible

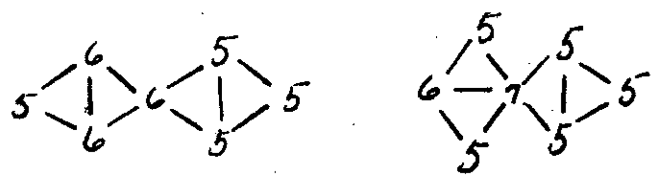


Figure 48. D-irreducible

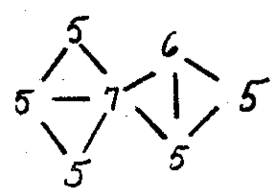


Figure 49. D-reducible

Examples. How complete are the D-irreducibility theorems? One way to decide is to examine small configurations, whose D-reducibility or -irreducibility have recently been resolved by computer. We will find that in most cases, if a configuration is not covered by one of the D-irreducibility theorems in this chapter, it is actually D-reducible.

Our body of examples will be those listed by Allaire and Swart in their comprehensive computer study of 10-ring configurations. These include configurations whose boundaries have no more than 10 vertices, and which can be built from very small configurations by extensions without ever letting the boundary exceed 10 vertices. From this set we exclude the following:

(a) any configuration with an interior vertex of valence less than 5;

(b) the Pentagon and Triad;

(c) any configuration with a D-reducible proper subconfiguration;

(d) any configuration with a 4-legger, or a 3-legger articulation point, or a "hanging pair" of 5 vertices (since no such configuration may be minimally D-reducible); and

(e) the following specific configurations, and any which contain them.

5:56565	6:567(6)5(5)5
6:565(5)6	8:6(5)5(5)655
6:6(5)57(5)5	5:55(5)8(5)66
5:5(5)57(5)5	6:5(5)5x6((5)6)5 (see figure 50)

(Each of these is probably D-reducible, but they are not listed by Allaire and Swart in the early draft of their report.)

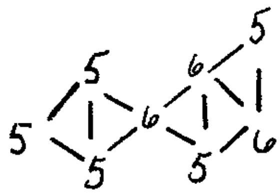


Figure 50. 6:5(5)5x6((5)6)5

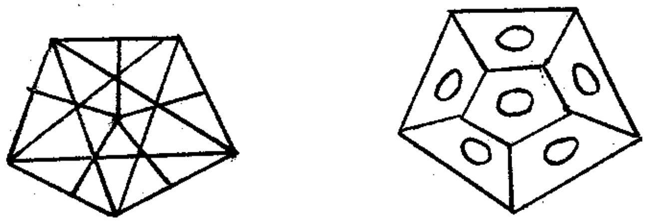
The resulting list consists of 104 configurations. Twelve are D-irreducible because of the interior triad theorem, and in nine of these cases, the complementary open set constructed according to the proof of that theorem is the largest complementary open set. (The exceptions are 6:565, 7:5665, and 7:566(5)5.) Three more cases are D-irreducible because of the theorem on 2-legger articulation points.

Of the remaining 89 candidates for D-reducibility, 83 are D-reducible. Only six are D-irreducible without being covered by any of the theorems.

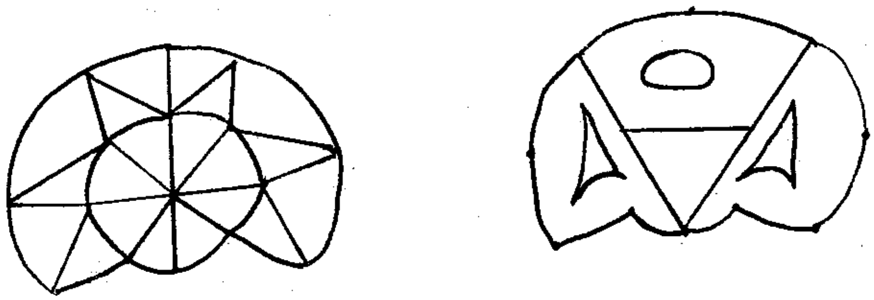
Furthermore, in each of these six cases, complementary open sets may be obtained by splicings of W_P and W_T. Thus, while the theorems do not cover every case, their underlying technique does.

The six exceptional configurations, and their complements, appear in figure 51. A list of all 104 examples appears at the end of the next chapter.

5:66666



8:55655



This open set, used again below, is analogous to Shimamoto's "second construction," which was used to construct the example at the end of Chapter I (figure 16).

7:5(5)66(5)5

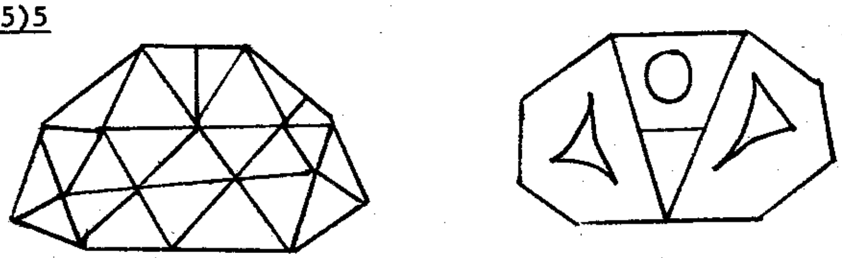
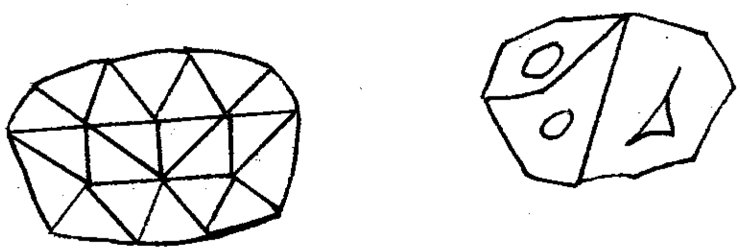


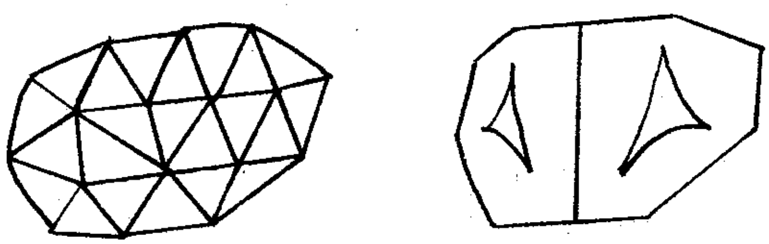
Figure 51. Six exceptional D-irreducible configurations and their complements. (These complements may not be the largest.)

7:56565



(Consider also the symmetrical version.)

7:5656(5)5



Allaire and Swart derive this open set, used again below, from a slight generalization of the interior image theorem. The above open set may, in turn, be derived from this one.

7:556(5)6

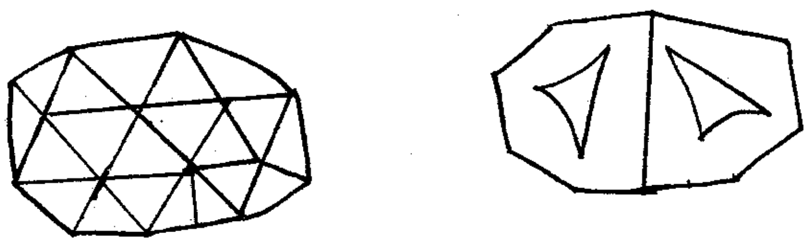


Figure 51, concluded

III. OTHER FORMS OF REDUCIBILITY

The notion of D-reducibility is not the broadest form of reducibility, since there are configurations which are not D-reducible but which are known not to appear in any minimal five-color map. In this chapter we will develop some broader concepts of reducibility, and we will see what can be done to generalize the irreducibility theorems to these new concepts.

We first introduce some new terminology, taking advantage of a superficial analogy to topological open sets. If R is a ring, let $\mathcal{U}_R$ (or, just $\mathcal{U}$) represent the set of all colorings of R. If $U \subseteq \mathcal{U}$, its complement U^C is the set $\mathcal{U}-U$. We call U closed if U^C is open. In any case, we let $\overline{U}$, the closure of U, be the smallest closed set containing U. (There is a smallest one, because of the union lemma for open sets.)

To restate a definition: a configuration A is D-reducible if $\overline{U}_A = \mathcal{U}$.

Let A be a proper configuration. A reducer for A is a (usually improper) configuration B, whose boundary may be identified with that of A but which has fewer interior vertices, such that $\overline{U}_B \subseteq \overline{U}_A$. This means that no coloring in U_B may also be in a complementary open set to U_A; or, that every coloring extendable to B may also be extended to A, possibly after modification by chain arguments.

If A is D-reducible, any smaller configuration is a reducer for A.

We call A C-reducible if it has a reducer. Otherwise, it is C-irreducible.

Lemma. A C-reducible configuration may not appear properly in a minimal five-color map.

Proof: Suppose a C-reducible configuration A appears properly in a minimal five-color map, M, and let B be a reducer for A. Since the boundary of B may be identified with that of A, we may construct a new map M' by putting B in place of A in M. (Since B may be improper, we need to assume that A appears properly in M to conclude that M' contains no loops.) Let C be the complementary configuration to A in M (and to B in M'). Since M is a five-color map, we know that $U_C \cap U_A = \emptyset$, and hence that $U_C \cap \overline{U}_A = \emptyset$. Since $U_B \subseteq \overline{U}_A$, it follows that $U_C \cap U_B = \emptyset$, and M' can have no coloring. Since M' is smaller than M, this is a contradiction. //

Lemma on subconfigurations. If A' contains a C-reducible subconfiguration A, then A' is also C-reducible.

Proof: If B is a reducer for A, a reducer for A' can be made by substituting B for A within A'. //

Note that the C-reducibility of a configuration cannot be defined entirely in terms of the associated open set, but also depends on the number of interior vertices. This motivates the following definitions.

A configuration is C*-reducible if it has a reducer with no interior vertices. Otherwise, it is C*-irreducible.

A configuration with no interior vertices may be drawn in such a way as to emphasize its boundary, by using the double-edge convention introduced at the end of Chapter I. For example, the two illustrations in figure 52 represent the same configuration. Frank Bernhart calls these representations AB-diagrams, a term which we will also apply to the configurations themselves. We will also use his term totally degenerate open set for the associated open set of an AB-diagram.

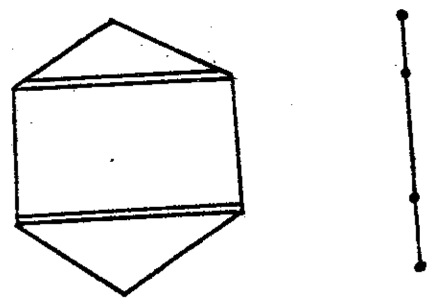


Figure 52

An open set U is complete, or C*-complete, if it meets every totally degenerate open set; that is, if it contains some coloring which will extend to each AB-diagram. By tracing the definitions we obtain the following equivalent definition:

Lemma. A configuration is C*-irreducible if and only if it has a complementary open set which is C*-complete. //

Examples. W_P and W_T are C*-complete. Thus the Pentagon and 5-5-5 Triad are C*-irreducible.

The point of this lemma is that the C*-reducibility of a configuration can be defined entirely in terms of the associated open set, so that the problem of C*-reducibility, like that of D-reducibility, is reduced to the study of abstract open sets. This is impossible for C-reducibility.

We will use one more reducibility concept.

An AB-diagram is simple if it has a representation in which every single and double interior edge adjoins a single vertex. For examples, see figure 53. The last AB-diagram in the figure is simple because it is really the same configuration as the first one.

An open set is <u>1-complete</u> if it contains a coloring which extends to each simple AB-diagram. A configuration is <u>1-irreducible</u> if it has a 1-complete complementary open set. Otherwise, it is <u>1-reducible</u>. A configuration is 1-reducible if and only if it has a reducer which is a simple AB-diagram.

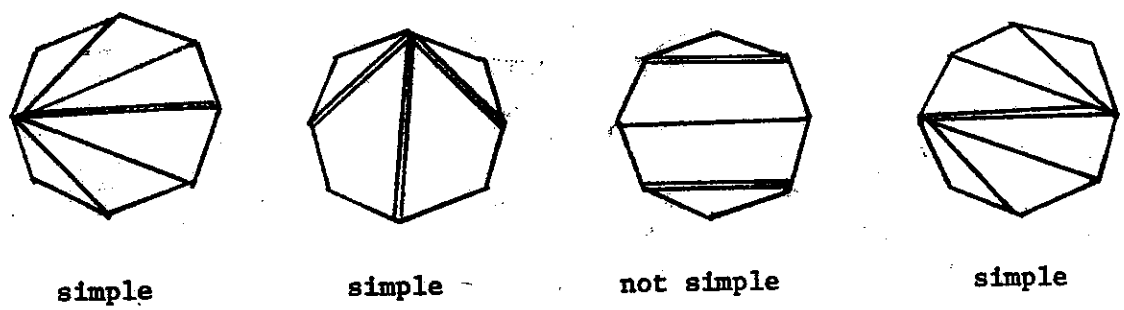


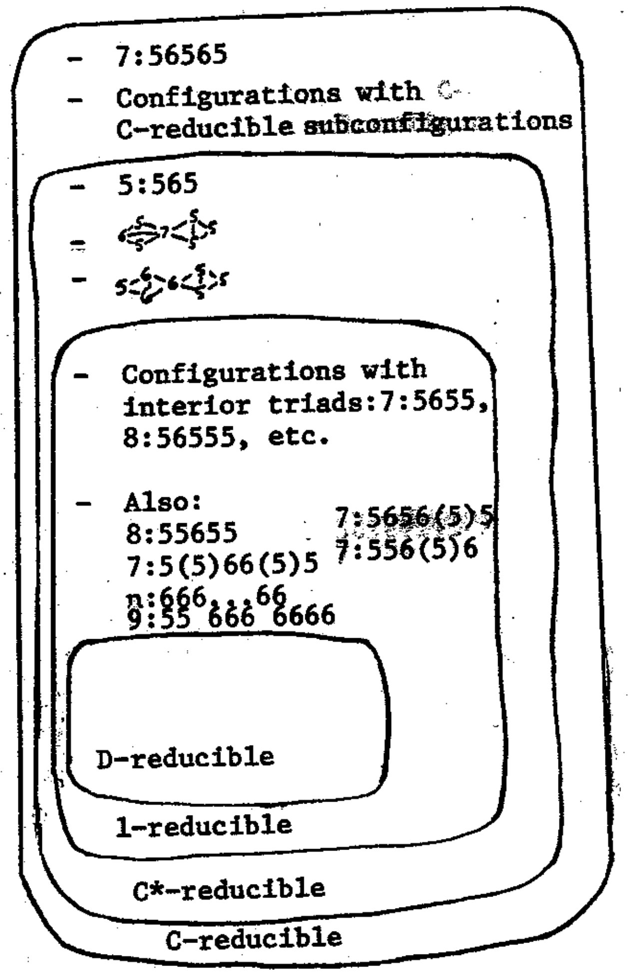
Figure 53. AB-diagrams

Equivalently, an open set U on a ring R is 1-complete if, for any nonempty set Q of vertices of R such that no two vertices of Q are adjacent, there is a coloring in U in which the vertices of Q have one color, and none of the other vertices of R have that color.

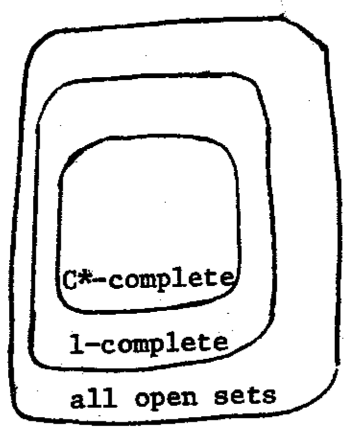
<u>Relationships among the types of reducibility</u>. A configuration is C-reducible if it has a reducer. It is C*-reducible if the reducer is an AB-diagram, and 1-reducible if the reducer is a simple AB-diagram. It is D-reducible if any smaller configuration is a reducer.

It follows that our four concepts may be ranked by implication: D-reducibility implies 1-reducibility implies C*-reducibility implies

C-reducibility. The relationships are illustrated in figure 54. C-reducibility is the broadest of our four concepts; in fact, every configuration which is known not to appear in a minimal five-color map is C-reducible.



Types of reducibility



Types of open sets

<u>Figure 54</u>

Apparently, none of the narrower concepts satisfy the lemma on subconfigurations. It is therefore easy to find configurations, such as the one shown in figure 55, which are C-reducible without (it seems) being C*-reducible. But if we limit our attention for a moment to minimally reducible configurations, the difference between these two concepts is mostly theoretical. Allaire and Swart found that the configuration 7:56565 is C-reducible but is not C*-reducible, and has no C*-reducible subconfiguration; however, it is the only such example known.

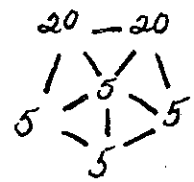


Figure 55

There are also some configurations which are C*-reducible without being 1-reducible. The leading example is the configuration 6:565, and other examples can be constructed from that one by the replacement construction (see figure 48, Chapter II). With these exceptions, all the published proofs of well-known reductions are actually proofs of 1-reducibility. In short, C-reducibility and 1-reducibility are very close.

The gap between 1-reducibility and D-reducibility is wider. Several examples arise from the interior triad theorem, which can be used to prove D-irreducibility but fails completely for 1-irreducibility. Five of the special cases in figure 51, Chapter II (all but 7:56565) are also 1-reducible without being D-reducible, as are the larger examples shown in figure 56.

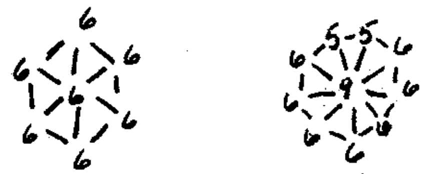


Figure 56. 1-reducible but not D-reducible

The types of reducibility can also be compared in more conventional terms, by describing the arguments used to establish reducibility of each type. To prove D-reducibility, one may use only chain arguments. For C-reducibility, one has also the full force of amalgamation arguments. For C*-reducibility, the amalgamation must involve merging all interior vertices with the boundary vertices; and for 1-reducibility, the amalgamation must consist of merging all of the interior vertices, together with one or more boundary vertices, with a single boundary vertex.

It is a philosophical question whether our chief interest should be in D-reducibility or C-reducibility. Perhaps the four-color conjecture can still be resolved by compiling a longer list of reductions. If so, the broadest possible concept is the only useful one. But if a more subtle approach is to be used, the theoretically more accessible D-reducibility is the concept of choice; especially since it is quite likely (as we will argue in the next chapter) that every map contains a D-reducible configuration.

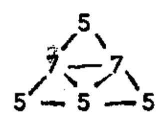
Generalization of the Irreducibility Theorems

Each of the irreducibility theorems in Chapter 2 gives rise to a variety of conjectures, by substituting "C-irreducible," "C*-irreducible," or "1-irreducible" for "D-irreducible." Now we will investigate which of these conjectures are plausible, and which can be proved.

<u>The interior image theorem</u>. Theorem 1 is false for 1-, C*-, and C-reducibility. In proving it for D-irreducibility, it was only necessary to construct an open set, and this was done by reference to an improper diagram, so that the resulting open set was manifestly incomplete. For other forms of irreducibility, it would be necessary to construct a complete open set, and in general there is no reason to think that one would exist.

In fact, there are many counterexamples to the "interior image conjectures." Among others, the configurations 6:565, 7:5655, and 5:57(5)75 (see figure 57) all have interior triads but are nevertheless C*-reducible.

There may even be an example of the failure of the "gradient conjecture" for C*-irreducibility, if it is true, as reported, that 7:56565 is C*-irreducible while its gradient 7:56665 is C*-reducible.



<u>Figure 57. 5:57(5)75</u>

The 5-legger conjuecture is true for C*-reducibility, with an additional assumption. If B contains an interior image of a Pentagon and the wings are C*-irreducible, so is B. We will prove one case of this.

Theorem 4. If a configuration A is C*-irreducible, and B is obtained from A by a 3-extension (or higher order), then B is C*-irreducible.

Proof: We will prove the case of a 3-extension as shown in figure 58. We define two open sets W_1 and W_2, complementary to B, in figure 59. W_1, is familiar from Theorem 1. W_2 is formed by splicing an open set W_A, a C*-complete open set complementary to A, with the open set W_H () on the six-ring consisting of all colorings which use all four colors. (W_H is built from W_P as shown in figure 60.)

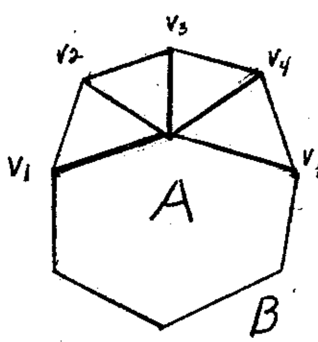


Figure 58

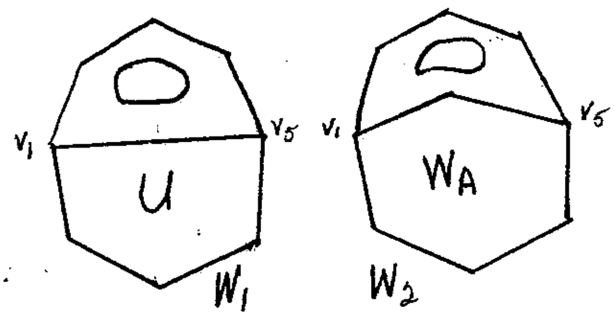


Figure 59

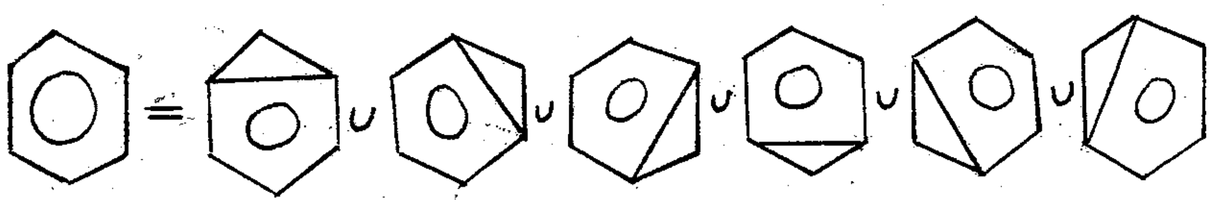


Figure 60

Let $W = W_1 \cup W_2$. Clearly W is an open set complementary to B; we need to prove that W is C*-complete. To do this, let X be any AB-diagram whose boundary may be identified with that of B.

If v_1 and v_5 are not joined with a double line in X (i.e., if they are not identified in the configuration represented by X) then it is easy to check that some coloring in W_1 extends to X. (This is ultimately a consequence of the C*-completeness of W_P.)

If v_1 and v_5 are joined by a double line in X, then X really is two AB-diagrams, one belonging on part of the boundary of A and one belonging on the vertices v_1, v_2, v_3, v_4, v_5. We may find colorings in W_A and W_H which extend to the respective parts of X. Clearly, these two colorings may be spliced; and their splicing, which is a member of W_2, is extendable to X.

Therefore W is C*-complete and B is C*-irreducible. //

<u>The 2-extension theorem</u>. We are concerned only with the part of theorem 2 which deduces the irreducibility of the larger configuration from that of the smaller. The converse may be extended to 1-irreducibility and C*-irreducibility, but is of little interest.

We will prove the 2-extension theorem for 1-irreducibility, as theorem 5.

It is unfortunate that the 2-extension theorem cannot be extended to C*-irreducibility. But it cannot, unless we first prove the four-color conjecture; for from the hypothesis of a five-color map, we can construct a counterexample to the extended version. This "counter-example" follows the proof of theorem 5.

No "real" examples are known of the failure of the 2-extension conjecture for C*-irreducibility or C-irreducibility. But so few examples are known of configurations which are C-reducible without also being 1-reducible, that it is hardly surprising that none of them happen to be 2-extensions of C*-irreducible configurations. This probably cannot be taken as strong evidence for the truth of the conjecture.

In specific cases, the C*-irreducibility of 2-extensions may be verified directly by constructing a complementary open set in the usual way and checking it against all AB-diagrams. The analysis is practical when the number of interior vertices is less than six, and in all such cases, the 2-extension conjecture holds for C*-irreducibility.

Theorem 5. If a configuration A is 1-irreducible, and B is obtained from A by a 2-extension (or higher order), then B is 1-irreducible.

Proof: We will show that a splicing of order 3 between any 1-complete open set and the set W_P is 1-complete. The desired result will then follow from the proof of Theorem 2.

Let W_1 be an open set on a ring R_1, let W be the splicing of W_1 and W_P, and let R be the boundary of W, as shown in figure 61. We will use our alternate definition of 1-reducibility; we will let Q be a nonempty set of vertices of R, and we will find a coloring in W such that precisely the vertices in Q have the color α.

Case 1. v_1 and v_4 are in Q. Let $Q_1 = R_1 \cap Q$, the set of vertices of R_1 which are in Q. Since W_1 is 1-complete, there is a coloring $c_1 \epsilon W_1$ for which only the vertices of Q_1, including v_1 and

v_4, have the color α. We may assume that v_0 has color β.

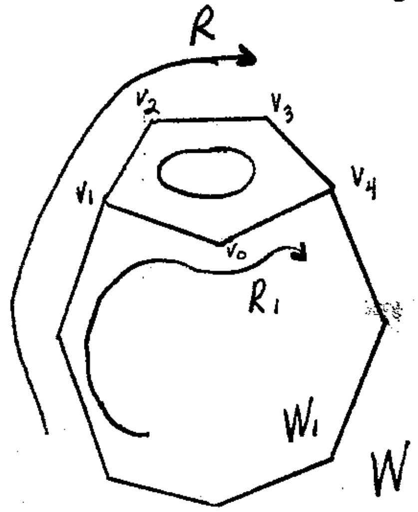


Figure 61

Now figure 62 shows how to splice c_1 with a coloring in W_P to obtain a coloring in W in which only the vertices of Q have color α.

Case 2. $v_1 \varepsilon Q$, $v_4 \notin Q$. Construct Q_1 and c_1 as before. We may assume that v_1 has color α, v_0 has color β, and v_4 has color γ in Q_1. The case is completed in figure 63, using α or β for v_3 according to whether or not $v_3 \varepsilon Q$.

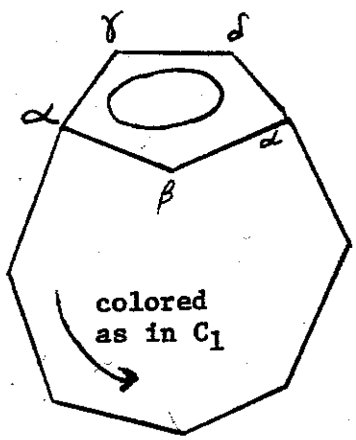


Figure 62

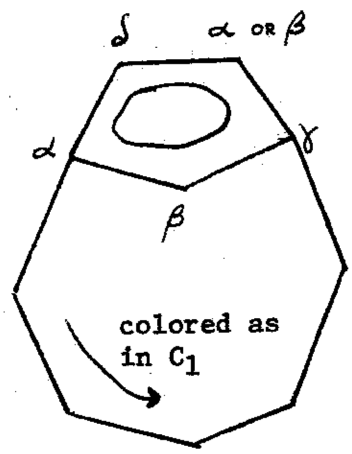


Figure 63

Case 3. None of v_1, v_2, v_3 or v_4 is in Q. Now define $Q_1 = (Q \; R_1) \; \{v_0\}$, and select c_1 as before. We may assume that in c_1, v_1 has color β, v_0 has color α, and v_4 has color β or γ. The case is completed in figure 64.

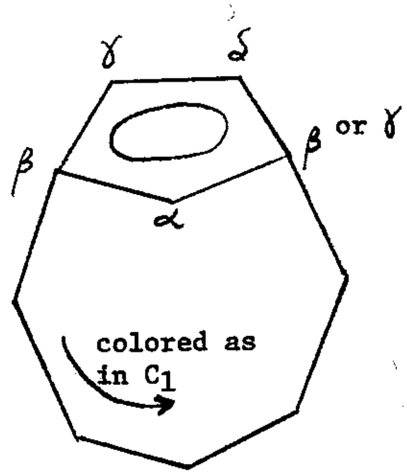


Figure 64

Case 4. of v_1, v_2, v_3, v_4; only v_2 Q. In this case set $Q_1 = (Q \; R_1) \; \{v_0\}$, and choose c_1 as before. (The case would be easier if we could set $Q_1 = Q \; R_1$, but there would be no assurance that $Q_1 \neq \emptyset$.) If it is possible to choose c_1 so that v_1 and v_4 have different colors, do so; the case can then be completed as in figure 65.

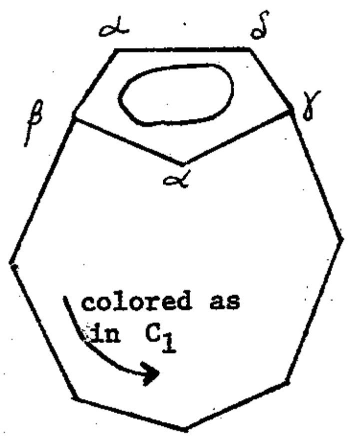


Figure 65

If no such choice of c_1 is possible, we must use a chain argument. Choose c_1 so that v_1, v_0, v_4 have colors β, α, β; and fill in R_1 by a configuration $A_{c_1, \beta\gamma}$ as prescribed in the definition of open set W_1. Now v_1 and v_4 must be joined by a $\beta\gamma$-chain; otherwise, we could alter c_1 so that v_4 had the color γ but Q_1 was still satisfied. Therefore we can alter c_1 so that v_0 has color δ, without affecting the other vertices on R_1.

From this new coloring we can complete the case by the splicing in figure 66. //

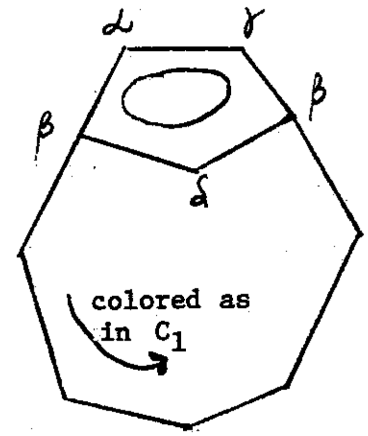


Figure 66

Following is the counterexample that prevents the extension of the 2-extension theorem to C*-reducibility.

Counterexample. If there is a minimal five-color map, it contains a pentagon. Let A be the complementary configuration, for which we use a symbol in figure 67. We know nothing about A, except that its associated open set is W_p; a coloring extends to A if it involves exactly four colors. Clearly A may not be reducible in any sense, being contained in a minimal five-color map.

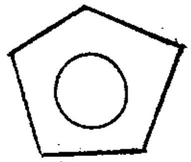


Figure 67

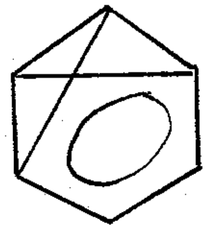


Figure 68

Let B be a 2-extension of A (figure 68). Now B may not be D-reducible or 1-reducible by theorems 2 and 5. But B is C*-reducible. Its reducer is the AB-diagram shown in figure 52. That is, any coloring extendable to this AB-diagram may also be extended to B, possibly with the help of chain arguments.

(We will not give the chain arguments here, except to note that by coincidence, they are nearly a direct transcription of the arguments used to reduce the "Birkhoff Diamond" in Chapter 1.)//

The replacement theorem (Theorem 3). Let U be a splicing of order 3 of two open sets U_1 and U_2. We showed in Theorem 5 that if U_1 is 1-complete and U_2 is W_P, then U is 1-complete. It is not generally true that if U_1 and U_2 are both 1-complete, so is U; or that if U_1 and U_2 are both C*-complete, so is U; or even that if U_1 is 1-complete and U_2 is C*-complete then U is 1-complete. It is true that U is 1-complete if U_1 and U_2 are both C*-complete, or if U_1 is 1-complete and U_2 is W_T. These results can be proven by the technique of Theorem 5, and they yield the following generalization.

Theorem 6. Let B be obtained by replacing a boundary vertex of C with a copy of A. Suppose that either

(1) C and A are both C*-irreducible, or

(2) C and A are both 1-irreducible and one of them is a Pentagon or 5-5-5 Triad.

Then B must be 1-irreducible. //

Specific Configurations. The union and splicing lemmas can be used to prove C*-irreducibility for some specific configurations not covered by the theorems. The method is to find several complementary open sets to the same configuration, by using splicing constructions such as the interior triad theorem; and then to show by exhaustion that the union of all these open sets is a C*-complete, complementary open set.

Two examples of this method appear in figures 69-70 and figures 71-72. Figure 69 shows a configuration, and figure 70 shows splicing constructions for several complementary open sets whose union is C*-complete. Similarly, figure 71 shows a configuration whose C*-irreducibility is demonstrated in figure 72. In each case, checking for C*-completeness is a tedious exercise; it is necessary to examine each AB-diagram to see that it can be colored by some coloring in the open set.

Both of these configurations are significant in that they are "geographically good," a quality (defined in the next chapter) which would normally make them likely candidates for reduction.

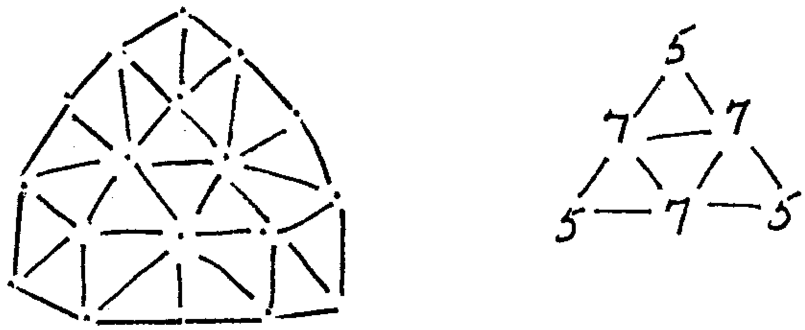
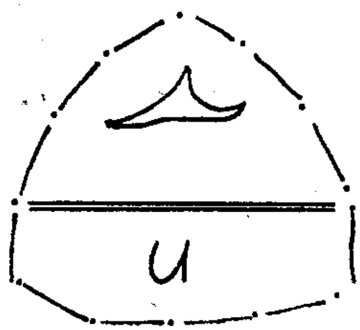
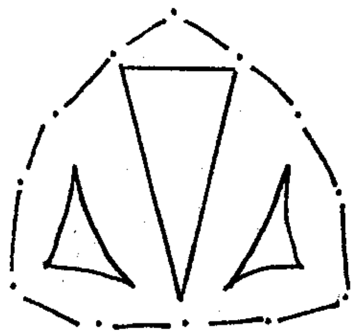


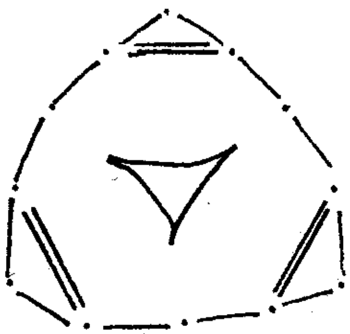
Figure 69

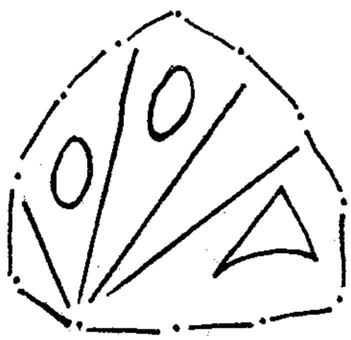


(3 symmetries)



(3 symmetries)





(6 symmetries)

Figure 70. Complements to 7:57(5)75

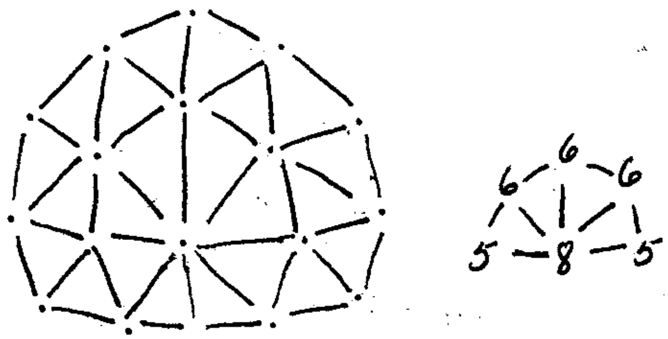
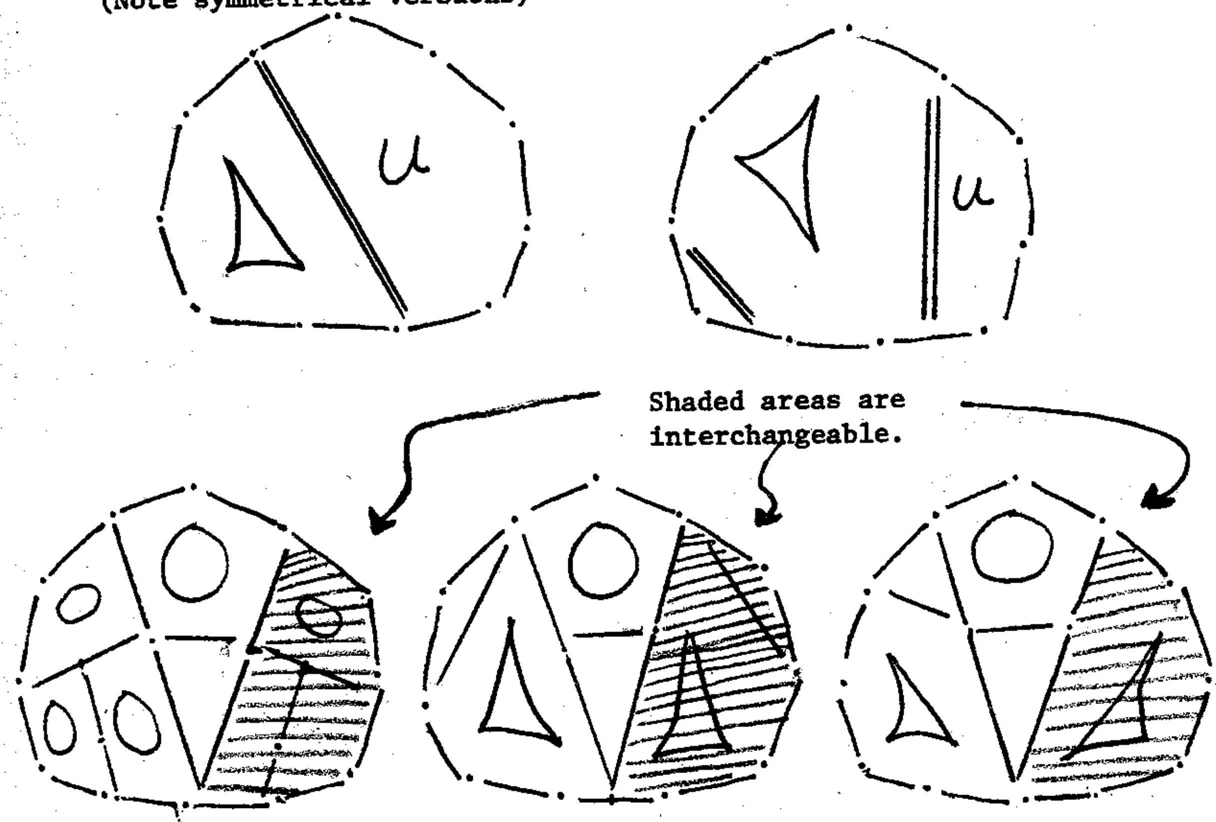


Figure 71

Complementary Open Sets
(Note symmetrical versions)



As Frank Bernhart points out, these open sets are related by

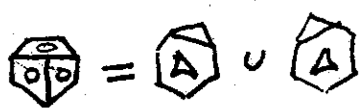


Figure 72. Complements to 8:56665

Bernhart's Thesis. In his thesis, Frank Bernhart has introduced a new necessary condition for an open set to be the associated open set of a configuration. Almost all open sets meet the condition, but some are known which do not. We can use the term "legitimate open set" for one which satisfies Bernhart's condition.

Since open sets are supposed to be associated open sets of hypothetical configurations, there is no interest at all in illegitimate open sets. Therefore it is good to know that all of the theory of these three chapters remains valid when the term "legitimate open set" is substituted for "open set" throughout. In particular, the splicing lemma, the irreducibility theorems, and all of our examples survive.

It is possible that a configuration could be reducible (under some concept) in the new context which was not reducible before. This would happen if its only complementary open set (or complementary complete open set) is illegitimate. No such examples are known.

Examples: As in the case of D-reducibility, we can test the completeness of our 1-irreducibility theorems by applying them to the 104 configurations we have selected from the Allaire-Swart list.

(It is still appropriate to omit configurations with 4-leggers and hanging pairs, because either of these would prevent a configuration from being minimally 1-reducible, by Theorems 5 and 6. The same is true for 3-legger articulation points, provided there are only two wings and one of them is a Pentagon or Triad -- and this is guaranteed when the boundary cannot exceed 10 vertices. Thus, the examples we are using are the only interesting ones.)

One of the 104 examples (the one in figure 47, Chapter II) is 1-irreducible because of the replacement theorem. Two more (figure 48) can also be proved 1-irreducible as soon as we know that 6:565 is 1-irreducible. Of the remaining 101 candidates for 1-reducibility, 98 are 1-reducible. The only 1-irreducible configurations not covered by the theorems are 6:565, 7:56565, and 7:5665. The first is C*-reducible, the second is C-reducible, and the last is (apparently) C-irreducible.

These examples support the generalization that if a relatively small or low-valued configuration cannot be proved 1-irreducible using the theorems of this chapter, then it is usually reducible.

The 104 examples are listed below.

104 n-Ring Configurations (with $n \leq 10$) Whose Reducibility was Tested by Allaire and Swart

I. D-reducible configurations (83)

6 boundary vertices

5:555

7 boundary vertices

6:555

8 boundary vertices

7:5555
6:5655
5:5665
5:66655
6:5(5)5x5(5)5

9 boundary vertices

6:5665
8:55555
7:56555
7:55655
5:66665

7:6(5)555
7:5(5)655
7:56(5)55
5:7(5)6655
5:76655(5)

5:6(5)7655
7:555x5(5)5
6:6(5)5x5(5)5

10 boundary, 6 interior

7:56655
6:56665
7:5(5)755
7:56(5)65
6:6(5)665

10 boundary, 7 interior

9:555 555
8:565 555
8:556 555
7:565 655
7:565 565

6:666 655
8:6(5)5555
8:5(5)6555
8:56(5)555
8:5(5)5655

8:55(5)655
7:6(5)6555
7:6(5)5655
7:6(5)5565
7:57(5)555

7:56(5)565
6:5(5)7655
6:5(5)7565
6:57(5)565
5:7(5)6665

continued

10 boundary, 7 interior

continued

5:76665(5)
5:7(5)6656
5:76656(5)
5:6(5)6666
7:6(5)6(5)55

7:6(5)65(5)5
7:6(5)56(5)5
7:6(5)55(5)6
6:5(5)76(5)5
5:5(5)77(5)5

8:555x555
7:6(5)5x555
6:6(5)5x6(5)5
6:6(5)5x5(5)6

10 boundary, 8 interior

7:6(6)5(7)5(5)5
7:6(6)5(5)5(6)6
8:6(6)5(5)555
7:7(6)5(5)555
7:6(6)5(5)565

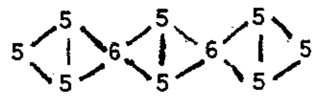
7:5(5)656(5)5
7:5((5)5)6565
7:5((5)5)66(5)5
5:6(5)8(5)655
5:7(5)7(5)655
5:7(5)7655(5)
5:77(5)655(5)
5:7(5)6(5)755
5:7(5)67(5)55
5:767(5)55(5)

8:5(5)5 x 5555
7:6(5)55 x 5(5)5
7:5(5)65 x 5(5)5
6:7((5)5)5 x 5(5)5
6:6((5)5)6 x 5(5)5
6:5((5)6)6 x 5(5)5

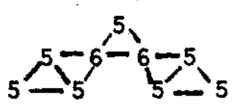
10 boundary, 9 interior

7:6(6)5(5)5 x 5(5)5
7:6(5)5(5)6 x 5(5)5

10 boundary, 10 interior



II. D-irreducible configurations (21)

	Proof of D-irreducibility	1-reducible?	
6:565	int. triad	no	C*-reducible
7:5655	int. triad	yes	
7:565(5)5	int. triad	yes	
7:5665	int. triad	no	not C-reducible*
8:56555	int. triad	yes	
8:55655	Figure 47	yes	
7:56565	Figure 47	no	C-reducible but not C*-reducible
5:66666	Figure 47	yes	
7:57(5)55	int. triad	yes	
7:6(5)655	Figure 47	yes	
7:6(5)565	int. triad	yes	
7:5(5)665	int. triad	yes	
8:565(5)55	int. triad	yes	
8:5655(5)5	int. triad	yes	
5:7(5)7655	int. triad	yes	
7:5(5)6565	Figure 47	yes	
7:5(5)66(5)5	Figure 47	yes	
6:6(5)75(5)5	int. triad	yes	
7:565 x 5(5)5	replacement	no	C*-irreducible; 1-irreducibility follows from that of 6:565
6:6(5)6x5(5)5	replacement	no	(same as above)
	replacement	no	not C-reducible*

*In the early version of their paper, Allaire and Swart do not guarantee that they have tested all possible reducers. Therefore, these assertions of C-irreducibility should be taken as experimental observations rather than as theorems.

IV. REDUCIBILITY AND VALUE

The value of a configuration is the number of boundary vertices minus the number of interior vertices.

Examples of value. The Pentagon has value 4. The Triad has value 3. The Shimamoto Horseshoe (figure 73) has value 6.

Let $\nu(A)$ represent the value of A.

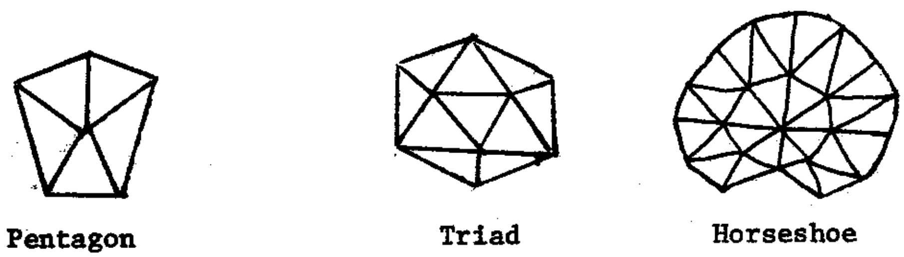


Figure 73

Making a 2-extension does not change the value of a configuration. More generally:

Lemma 1. If B is obtained from A by a k-extension, then

$$\nu(B) = \nu(A) + (k-2) \ .$$

Proof: A k-extension adds k new boundary vertices, and converts one boundary vertex into an interior vertex. //

Lemma 2. If B contains an n-fold, m-legger articulation point, and the wings are $W_1,\ldots,W_n$, then

$$\nu(B) = (\nu(W_1) - 3) + \ldots + (\nu(W_n) - 3) + (m-1) \ .$$

Proof: See figure 74. The number of boundary vertices of B is the total number of boundary vertices of the wings, minus 3 for each wing, plus m. The number of interior vertices of B is the total number of interior vertices in the wings, plus 1 (the articulation point itself). //

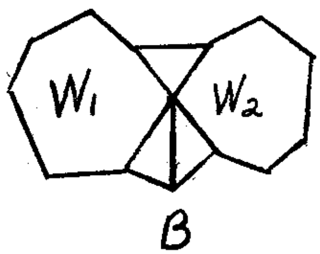


Figure 74

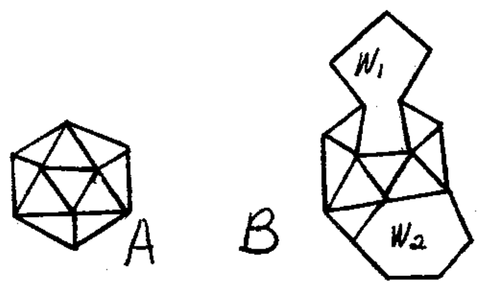


Figure 75

Lemma 3. If B contains an interior image of A, and the wings are $W_1,\ldots,W_n$, then

$$\nu(B) = \nu(A) + \sum_{i=1}^{n} (\nu(W_i) - 3) .$$

Proof: See figure 75. B has as many interior vertices as A and $W_1,\ldots,W_n$ combined, and also as many boundary vertices as A and $W_1,\ldots,W_n$ combined minus 3 for each wing. (If 4 boundary vertices of W_i are in the image of A, then one boundary vertex of A appears twice in B.)//

Lemma 4. If B is obtained from C by replacing a boundary vertex of C with a copy of A, then

$$\nu(B) = \nu(C) + \nu(A) - 4 .$$

Proof: B has as many interior vertices as C and A combined, and 4 fewer boundary vertices. //

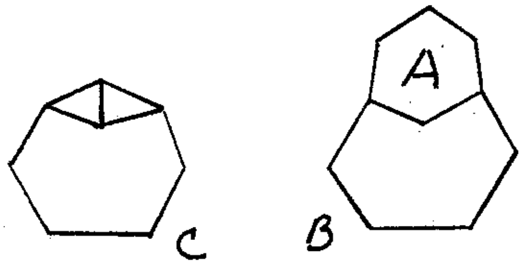


Figure 76

We can already state the main result of this chapter:

Theorem 7. If M is a minimal five-color map, there is a configuration having value less than 3, which may be imbedded in the third neighborhood of some vertex of M.

(The imbedding may be improper.)

The significance of this theorem is that low values are related to D-reducibility. We will describe this relationship in the first half of the chapter, and defer the proof of the theorem until the second half.

Reducibility and extensions. As we have seen, a 2-extension does not affect either the value of a configuration or its D-reducibility. Intuitively, a 1-extension tends to make a configuration "more reducible," while higher-order extensions make it "less reducible."

Suppose that A, a configuration having value ν, is obtained from a Pentagon by making a series of extensions, say n of them. From Lemma 1 we can derive the average order of the extensions, which is $2 + \frac{\nu-4}{n}$.

If ν is less than 4, the average order of the extensions is less than 2, and intuitively, A ought to be reducible.

This intuition is supported by two more concrete arguments, based on geographical goodness, and constructibility.

Geographically good configurations. A configuration is geographically good (GG) if it contains no 4- (or more-) leggers. Equivalently, it cannot be obtained from a smaller configuration by a 2-extension (or higher order), and it does not contain an articulation point, except possibly 2- or 3-legger articulation points.

Geographical goodness is useful as a first approximation to reducibility. It is likely that every minimal reducible configuration is GG (this is at least true for D-reducibility, mainly because of the 2-extension theorem). The converse, that GG configurations tend to be reducible, holds only very roughly. The most conspicuous exceptions are:

(1) the triad,

(2) some configurations with articulation points (some researchers define GG to exclude some or all articulation points, thus producing a closer approximation to reducibility), and

(3) GG configurations with interior triads. See figure 77. These must be D-irreducible by the interior triad theorem, but many of them, especially those with low values, turn out to be C*-reducible. The last two examples in figure 77 are the first known examples of GG configurations, without articulation points,

other than the Triad, which are C*-irreducible. (See Chapter III, figures 62-65.) Both have value 6.

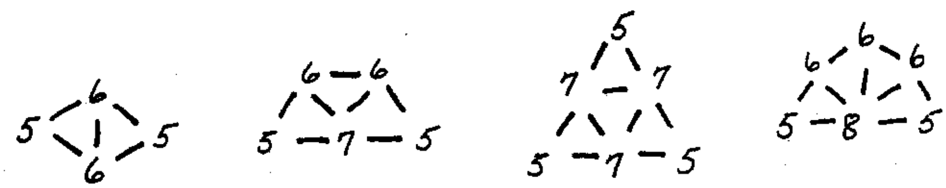


Figure 77. GG configurations with interior triads

The following result is important in spite of the distance between GG and reducibility.

Result of Appel and Haken. There is a finite list of GG configurations such that every map contains one of them.

Appel and Haken are compiling, by computer, an explicit list consisting of several thousand specific configurations. To show that every map contains a configuration from the list, they use "discharging" methods, similar to those used to show that every map with fewer than some stated number of vertices may be four-colored.

Theorem 7 leads directly to an improvement of the above result.

Lemma. Every configuration with value less than 3 contains a GG configuration with value less than 3.

Proof: Let A be a minimal counterexample. Since A is not GG, it must be obtainable by a 2-extension (or higher order), or contain a 4- (or more-) legger articulation point.

In the first case, remove the extension. The result is a subconfiguration of A, whose value may not exceed that of A (Lemma 1); so it contains a GG configuration with value less than 3.

In the second case, one of the wings must have value less than 3 (Lemma 2), so must contain a GG configuration with value less than 3. //

Corollary (to Theorem 7 and its proof): There is a finite list of GG configurations such that every minimal five-color map contains one of them. Each configuration on the list has value less than 3, and is small enough to be imbedded in the third neighborhood of a vertex.

Proof: In proving Theorem 7 we will show that the configuration which it asserts can be imbedded in M may be chosen so that it has no interior vertex of valence greater than 30. There are only finitely many configurations meeting this requirement which are small enough to fit within a third neighborhood. Of those with value less than 3, each contains a GG configuration whose value is also less than 3, and these comprise the list called for in the corollary. //

The Triad, which is on Appel and Haken's list, is not on ours. Unfortunately, ours still contains a few configurations which are known not to be reducible. Also, Appel and Haken define GG so as to exclude three-legger articulation points, a refinement not permitted by our methods unless we replace the value 3 in the corollary with 4, which would admit the Triad to our list.

Constructibility. If C is a class of configurations, and A is a configuration, we say that A is constructible from C if

(1) A$\varepsilon\mathcal{C}$, or

(2) A is obtained by replacing a boundary vertex of B with a copy of C, and B and C are constructible from $\mathcal{C}$, or

(3) A contains an interior image of B, and B and all of the wings are constructible from $\mathcal{C}$.

(This definition is not really circular, since the constructibility of A is made to depend on the constructibility of smaller configurations only.)

The irreducibility theorems may be summarized by saying that if $\mathcal{C}$ is a class of D-irreducible configurations, then any configuration which is constructible from $\mathcal{C}$ is also D-irreducible. Furthermore, if none of the configurations of $\mathcal{C}$ contain D-reducible subconfigurations, the same can be said of the configurations which are constructible from $\mathcal{C}$.

If A is constructible from the class consisting of the Pentagon and the Triad, we say simply that A is constructible. Every constructible configuration is D-irreducible (and free of D-reducible subconfigurations); and furthermore, nearly every configuration which is known to be D-irreducible (and free of D-reducible subconfigurations) is constructible. In fact, the only exception we have seen so far (there are others) is the Shimamoto Horseshoe. Among large configurations, there is a large middle ground for which D-reducibility is unresolved, but it is still a good rule that non-constructible configurations tend to be D-reducible. The lower the values, the better the rule.

Theorem 8. Let C be the class of configurations having value at least 4. If A is not in C, then A is not constructible from C.

Proof: If A is the smallest counterexample, then A must be obtained form configurations having value at least 4, using one of the processes listed in the definition of constructibility. One of lemmas 3 and 4 must apply, and in either case, it is clear that $\nu(A) \geq 4$. //

One way to interpret Theorem 8 is that if a person chooses to believe that all configurations with value less than 4 are D-reducible (or have D-reducible subconfigurations), then we cannot prove him wrong using the irreducibility theorems. Unfortunately, the D-irreducibility of the Triad (value 3) destroys this conjecture. In view of theorem 7, it would be nice if Theorem 8 could be proved with value "3" in place of "4"; it would then follow that configurations with value less than 3 are not constructible, and every minimal five-color map would contain a non-constructible configuration. But this will not work either, since it is possible for configurations with value less than 3 to be constructed from Triads.

Examples: A configuration of the type shown in figure 78 is constructible, but can have arbitrarily small value - specifically, (4-k) where k is the number of Triads it contains. A configuration such as that in figure 79 contains no Triads, but is still constructible from Triads, and its value is (4-k) where k is the number of its 6-vertices.

Note: The word "constructible" was introduced by Frank Bernhart. His term "S-constructible" means the same as our "constructible (from the Pentagon and Triad)", except that part (3) is omitted from the definition. This sacrifices the strong correlation with D-reducibility, but provides a tolerable approximation to C*-reducibility, which is his concern.

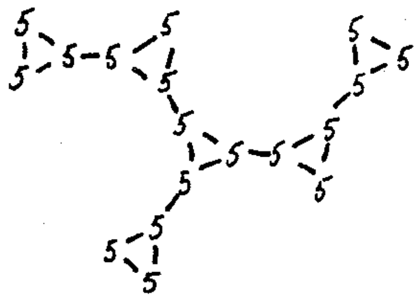


Figure 78

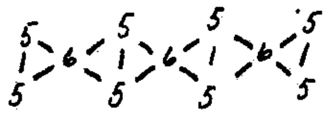


Figure 79

Constructible configurations with arbitrarily small value

Also note: A few specialized constructions have been discovered for producing D-irreducible configurations which are not merely special cases of the theorems in Chapter II. Among these are Bernhart's "triple fusion" (applicable, for example, to the configuration in figure 80, which has value 9) and a process derived from Shimamoto's "second construction," which produced the Horseshoe. Each of these applies only to a few large, high-valued configurations, and if they were all included in our definition of constructibility, Theorem 8 would still hold.

To summarize our philosophical arguments: most reasonably small configurations with value less than 3 are D-reducible because

(1) they may be obtainable from Pentagons by a series of extensions having average order less than 2,

(2) they contain GG configurations which also have value less than 3, and

(3) they are not constructible, except from other configurations having value 3 or less.

We can now begin the proof of Theorem 7.

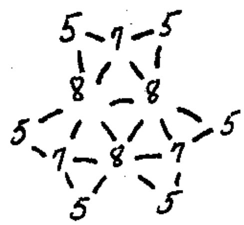


Figure 80. D-irreducible

Values of neighborhoods. For configurations which appear in maps as the second or third neighborhoods of vertices, there is an alternate formulation of value, in terms of the charges of the vertices in the neighborhood. By the charge on a vertex v, which we denote by $c(v)$, we mean 6 less than the valence. (This is the negative of the usual definition.)

Let v be a vertex in a minimal five-color map. Denote by R_k the subgraph of k-th neighbors of $v (k = 1,2,...)$ and by m_k the number of k-th neighbors. We say that the k-th neighborhood is orderly if $R_1,...,R_k$ are all rings. If the k-th neighborhood of v is orderly, then it is a configuration, its inclusion in the map is a proper imbedding, its boundary is precisely R_k, and its value is $(m_k - m_{k-1} - ... - m_1 - 1)$.

In an orderly neighborhood, the edges joining R_k to R_{k+1} divide the region between these rings into triangles, and there is one triangle for each edge in either ring. Therefore, the number of edges between R_k and R_{k+1} is $(m_k + m_{k+1})$. This result holds whenever R_k is a ring and the "inner boundary" of R_{k+1} is a ring; that is, even if there are extraneous edges between vertices of R_{k+1}.

Lemma (second neighborhood formula): Let A be the second neighborhood of a vertex in a minimal five-color map. If A is orderly, the value of A is given by

$$\nu(A) = \sum_{v' \varepsilon R_1} c(v') + m_1 - 1 .$$

Proof: The value of A is $(m_2 - m_1 - 1)$. To evaluate m_2, we will calculate the sum S of the numbers of neighbors of the vertices of R_1. This is the sum of their valences,

$$S = \sum_{v' \varepsilon R_1} (c(v') + 6) = \sum_{v' \varepsilon R_1} c(v') + 6m_1 \qquad (1)$$

S can also be determined by counting the neighbors themselves. Each $v' \varepsilon R_1$ is a neighbor to two other first neighbors, and v itself is a neighbor to all m_1 of them. The number of appearances of second neighbors as neighbors to vertices in R_1 is the number of edges between R_1 and R_2, which is $(m_1 + m_2)$. Therefore

$$S = 2m_1 + m_1 + (m_1 + m_2) = m_2 + 4m_1 . \qquad (2)$$

Combining equations (1) and (2) gives

$$m_2 + 4m_1 = \sum_{v' \varepsilon R_1} c(v') + 6m_1$$

$$m_2 = \sum_{v' \varepsilon R_1} c(v') + 2m_1 \qquad (3)$$

so that the value of A is

$$m_2 - m_1 - 1 = \sum_{v' \epsilon R_1} c(v') + m_1 - 1 . \quad //$$

Lemma (third neighborhood formula for orderly neighborhoods). Let A be the third neighborhood of a vertex v in a minimal five-color map. If A is orderly, its value is given by

$$\nu(A) = \sum_{v' \epsilon R_1 \cup R_2} c(v') - 1 . \tag{4}$$

Proof: The value of A is $(m_3 - m_2 - m_1 - 1)$. Equation (3) is still valid for m_2. This time we will determine m_3 by counting the sum S of the valences of second neighbors,

$$S = \sum_{v' \epsilon R_2} (c(v') + 6) = \sum_{v' \epsilon R_2} c(v') + 6m_2 . \tag{5}$$

Alternatively, we count the neighbors themselves. Each second neighbor is itself a neighbor to two others. The number of appearances of first neighbors as neighbors to vertices in R_2 is $(m_1 + m_2)$, as before; and the number of appearances of third neighbors is $(m_2 + m_3)$. From this point of view,

$$\begin{aligned} S &= 2m_2 + (m_1 + m_2) + (m_2 + m_3) \\ &= m_3 + 4m_2 + m_1 . \end{aligned} \tag{6}$$

Combining equations (5) and (6) gives us the desired value.

$$m_3 + 4m_2 + m_1 = \sum_{v' \varepsilon R_2} c(v') + 6m_2$$

$$m_3 = \sum_{v' \varepsilon R_2} c(v') + 2m_2 - m_1 \qquad (7)$$

$$(m_3 - m_2 - m_1 - 1) = \sum_{v' \varepsilon R_2} c(v') + m_2 - 2m_1 - 1$$

$$= \sum_{v' \varepsilon R_2} c(v') + (\sum_{v' \varepsilon R_1} c(v') + 2m_1) - 2m_1 - 1$$

$$= \sum_{v' \varepsilon R_1 \quad R_2} c(v') - 1 . \quad //$$

In the last lemma, suppose that A is a configuration, but is not orderly, failing only by having some extraneous edges among the third neighbors. Then the above calculation is still valid for $(m_3 - m_2 - m_1 - 1)$. However, this may not be the value of A, since some third neighbors may be in the interior of A rather than on the boundary. If x is the number of interior third neighbors, the value of A is given by this version of the third neighborhood formula:

$$\nu(A) = \sum_{v' \varepsilon R_1 \quad R_2} c(v') - 1 - 2x \ . \qquad (8)$$

Note: It is generally true in orderly neighborhoods, as in (7), that

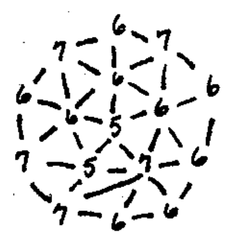
$$m_k = \sum_{v' \varepsilon R_{k-1}} c(v') + 2m_{k-1} - m_{k-2} \quad (k \geq 3) .$$

The formulas (which we will not use) for the values of orderly k-th neighborhoods, k = 4 or 5, are

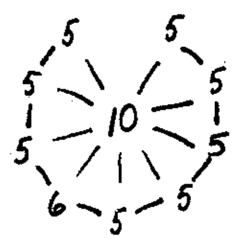
$$(k = 4) \qquad \nu(A) = \sum_{v' \in R_2 \cup R_3} c(v') - 2m_1 - 1$$

$$(k = 5) \qquad \nu(A) = \sum_{v' \in R_3 \cup R_4} c(v') - 2 \sum_{v' \in R_1} c(v') - 5m_1 - 1.$$

The neighborhood formulas are useful for determining the values of configurations illustrated by stick figures. We can tell at a glance, from the third neighborhood formula, that the configuration in figure 81 has value 3. The formulas may sometimes even be applied to configurations which are not second or third neighborhoods, such as the one in figure 82. Here the quickest way to find the value is to image adding a 7-vertex to the gap in the first neighborhood. This is a 2-extension, so does not change the value; and by the second neighborhood formula, the value of the extended configuration is -7 + 10 - 1 = +2



<u>Figure 81</u>



<u>Figure 82</u>

<u>Imbedding of neighborhoods in maps</u>. For the rest of this chapter, let M be a minimal five-color map; v will always represent a vertex of M. We know that M has no vertex of valence less than 5, and we can also rule out certain kinds of rings. A <u>Birkhoff Ring</u> is any ring of five or fewer vertices whose removal would divide M into two components of at least two vertices each. A <u>Bernhart Ring</u> is any ring of six vertices,

one of which is a 5-vertex, whose removal would divide M into two components of at least four vertices each. It is known that M may not contain either a Birkhoff Ring or a Bernhart Ring.

Lemma. The second neighborhood of v is orderly, and is a configuration whose inclusion in M is a proper imbedding.

Proof: There are several ways this lemma might fail, and each would cause M to contain a Birkhoff Ring (or a vertex with valence less than 5). For example, suppose there are two second neighbors of v, adjacent to each other but not adjacent to the same first neighbor. Call them v_1 and v_2. This would cause the second neighborhood not to be orderly, or even not to be a configuration. But in this case, there must be some first neighbor adjacent to v_1, say v_3; and some first neighbor adjacent to v_2, say v_4. It is easy to see that v_1, v_3, v_4, v_2, and (usually) v form a Birkhoff Ring. See figure 83. //

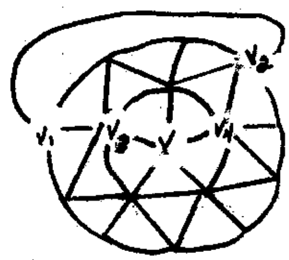


Figure 83

The same cannot be said of the third neighborhood. In fact, the third neighborhood of v in M may fail to be a configuration. Intuitively, the third neighborhood may fail to be "simply connected," so it will be useful to have something like a covering space. We will define a configuration which can be imbedded in the third neighborhood of v, and which illustrates its local properties.

The abstract third neighborhood A of v is the largest configuration such that

(1) A is itself the third neighborhood of some vertex v', and

(2) there is a local isomorphism $f:A \to M$ mapping v' onto v.

It follows from the definition and the previous lemma that f is an isomorphism when restricted to the second neighborhood of v'. We will use this isomorphism to identify the vertices in that second neighborhood with their images in M. Thus, we will refer to A as a neighborhood of v.

If the second neighborhood of v in A is the entire interior of A, we can conclude that A is orderly and that $f:A \to M$ is a (perhaps improper) imbedding of A in M. However, it is possible that A may contain some third neighbors in its interior. One way this can happen is illustrated in figure 84. In the figure, one of the second neighbors of v is a 5-vertex and is in contact with exactly one third neighbor, which is also a 5-vertex. The third neighbor is interior to A. Because of situations like this, we are adding some new information by stating the next lemma.

Lemma. In the above context, $f:A \to M$ is an imbedding.

Proof: Any failure of this lemma would cause M to contain either a Birkhoff Ring or a Bernhart Ring, or a vertex with valence less than 5. //

(Remember that our definition of "imbedding" is non-standard; see page 3.)

We are now able to identify all of the interior vertices of A with their images in M. We can classify the interior vertices as follows:

(1) v itself

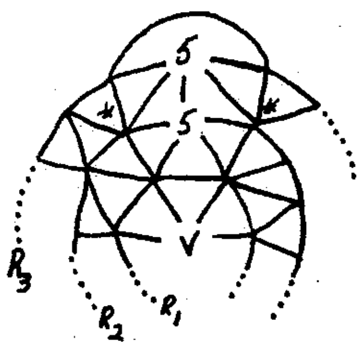
(2) first neighbors of v

(3) "inner" second neighbors, those in contact with two first neighbors

(4) "outer" second neighbors, those in contact with only one first neighbor

(5) interior third neighbors

By a <u>typical</u> interior third neighbor, we mean an interior third neighbor which is itself a 5-vertex, and which is adjacent with three second neighbors, having valences (in order) x, 5, y, with $x \leq 7$ and $y \leq 7$. This is exactly the situation in figure 84, if the valences of the vertices marked by * do not exceed 7. All other interior third neighbors are <u>atypical</u>.



<u>Figure 84</u>

Standard neighborhoods. We will now go one step farther from the original neighborhood of v in M. The "standard neighborhood" of v is a configuration which does not actually appear in M, but whose interior vertices are still roughly in correspondence with vertices near v in M.

The standard neighborhood is obtained by a series of alterations of the abstract third neighborhood A. The procedure is easiest to visualize in terms of stick figures, since we will be performing simple operations on the interior of A, which involve more complicated operations on the boundary of A.

The standard neighborhood C of v is obtained from the abstract third neighborhood A by performing these operations on the interior of A:

(1) Remove all interior third neighbors, except typical ones.

(2) Remove any outer second neighbor with valence greater than 7, and replace it with a 7-vertex.

(3) Remove any inner second neighbor with a valence greater than 8, and replace it with an 8-vertex.

(4) Remove any first neighbor with valence greater than 9, together with the outer second neighbors which adjoin it. Replace the first neighbor with a 9-vertex, and enclose it with four outer second neighbors. Of these four, the ones on the end are 7-vertices if they adjoin interior

third neighbors, and otherwise 6-vertices;
the middle two are a 6-vertex and a 7-vertex
(in either order).

Some examples of this construction appear in figures 85-86.

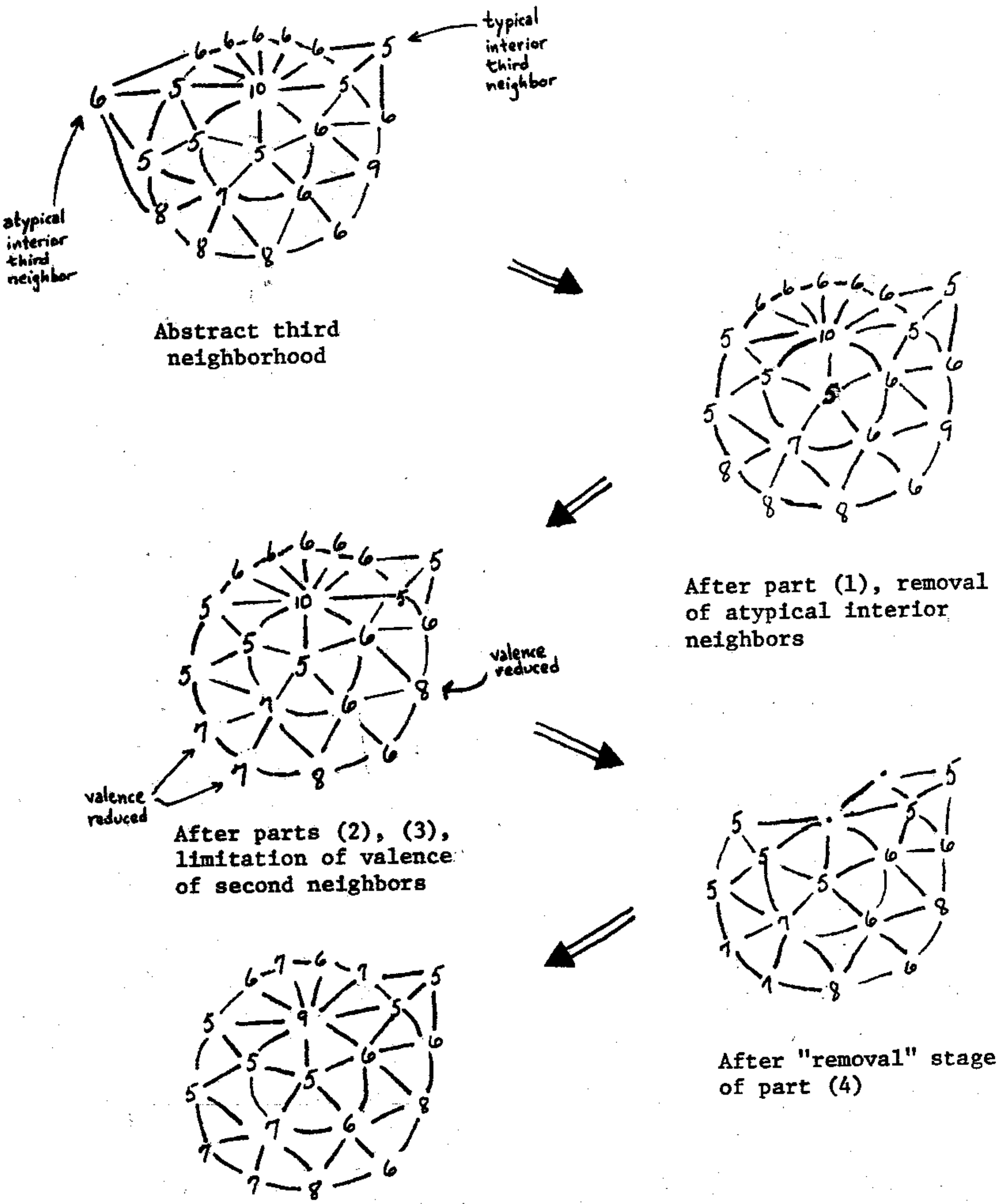
The interior vertices of C, except for the outer second neighbors added in the last part of the construction, may still be identified with vertices in the actual neighborhood of v in M. This correspondence is not valence-preserving. Still, it makes sense to ask, given a vertex v' in M, whether v' appears in the standard neighborhood of v, and what its valence is in the standard neighborhood:

(1) v appears in C with its original valence.

(2) If v' is a first neighbor of v, then v' appears in C with its original valence, but not exceeding 9. If the original valence of v' exceeds 9, there are some 6- and 7-vertices in C which do not correspond to vertices in M.

(3) If v' is an inner second neighbor of v in M, then v' appears in C with its original valence, but not exceeding 8.

(4) If v' is an outer second neighbor of v in M, and is connected to v by a first neighbor having valence no greater than 9, then v' appears in C with its original valence, but not exceeding 7.



Abstract third neighborhood

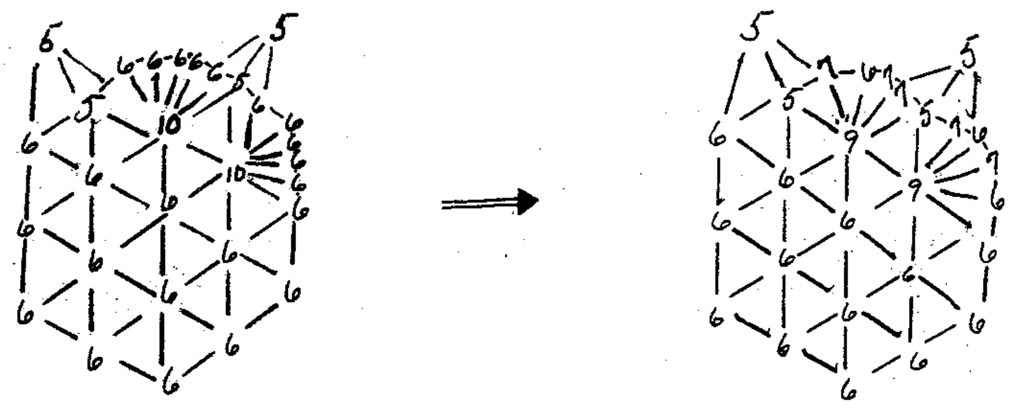
After part (1), removal of atypical interior neighbors

After parts (2), (3), limitation of valence of second neighbors

After "removal" stage of part (4)

The standard neighborhood

Figure 85. Step-by-step construction of a standard neighborhood



Abstract third neighborhood Standard neighborhood

Figure 86

(5) If v' is a typical interior third neighbor in the abstract third neighborhood of v, then it is also an interior third neighbor in the standard neighborhood.

(6) This list accounts for all of the interior vertices of C.

The key fact about the construction of the standard neighborhood is that the replacement steps, done sequentially, are all 2-extensions, and do not affect the value of the configuration. The replacements are made only to enable us to measure the value of standard neighborhoods with the third neighborhood formula. So, if we let B be the configuration obtained by following all of the "removal" steps of the construction, and none of the "replacement" steps, then B has the same value as C. But B is a configuration which appears (perhaps improperly) in M. This leads to another lemma:

<u>Lemma</u>: If the standard neighborhood of v has value less than 3, then there is a configuration with value less than 3 which can be imbedded in the third neighborhood of v in M. //

Theorem 7 will be proved, when we show that some vertex in M has a standard neighborhood with value less than 3. We will do this by computing the total value of the standard neighborhoods of all vertices in M. Letting T be this total, and N the number of vertices in M, we will show that $T - 3N < 0$, and the theorem will follow.

$$T - 3N = \sum_{v \in M} \left\{ \sum_{v' \in R_1 \cup R_2} c(v') - 1 \right\} - 2X - 3N$$

$$= \sum_{v \in M} \sum_{v' \in R_1 \cup R_2} c(v') - 4N - 2X \ . \qquad (9)$$

In each of these equations, X represents the total number of interior third neighbors in all the standard neighborhoods in M. The outer sum is taken over all vertices v in M. The inner sum is taken over all vertices v' which appear as first or second neighbors in the standard neighborhood of v; the charge $c(v')$ is to be measured in the standard neighborhood.

We will evaluate the terms of (9) in terms of "incidence numbers." If A is any configuration, let m_A be the number of distinct imbeddings of A in M. (When we use this notation, A is always small enough that "imbedding," "proper imbedding," and "local isomorphism" all mean the same thing.) For example, $m_{5:665}$ is the number of appearances in M of the configuration in figure 87. Becuase of automorphisms of A, the number of imbeddings m_A may be some multiple of the number of distinct "images" of A in M; thus, $m_{5:55}$ is six times the number of distinct

5-5-5 Triads in M.

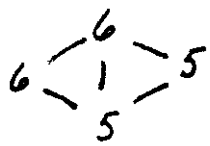


Figure 87. 5:665

We may generalize by letting A stand for a set of similar configurations. For example, we may write $A = 5{:}6H5$ $(H \geq 10)$; then

$$m_A = m_{5:6H5} = m_{5:6,10,5} + m_{5:6,11,5} + m_{5:6,12,5} + \\ + m_{5:6,13,5} + \cdots .$$

In this context we reserve special meaning for the letters L and H: we always assume that $L \leq 7$ and $H \geq 10$. Also, Y represents any valence, without restriction.

When A is an extremely simple configuration, we emphasize the special case by using a different letter:

$$n_k = m_k = \text{number of k-vertices in M } (k \geq 5)$$
$$\ell_{jk} = m_{j:k} = \text{number of j-k edges (if } k \neq j)$$
$$\ell_{kk} = m_{k:k} = \text{twice the number of k-k edges in M.}$$

Note that whether $k = j$ or not, ℓ_{jk} is the number of appearances of a j-vertex as a neighbor to a k-vertex.

We can note several relationships among the incidence numbers:

$$\ell_{jk} = \ell_{kj} \quad \text{(all j,k)}$$

$$N = n_5 + n_6 + n_7 + n_8 + \cdots \qquad (10)$$

(Recall that N is the total number of vertices of all valences)

$$12 = n_5 \quad - n_7 - 2n_8 - 3n_9 - 4n_{10} - \cdots \qquad (11)$$

(A well-known consequence of Euler's Formula)

$$kn_k = \ell_{5k} + \ell_{6k} + \ell_{7k} + \ell_{8k} + \cdots \qquad (12)$$

(Since kn_k is the total number of neighbors of k-vertices).

Counting. We will denote the sum in (9) by $\sum$. To evaluate $\sum$, we will not sum directly over v, but we will consider each vertex v' in M, and ask how much it contributes to $\sum$ by its appearances in $R_1 \cup R_2$ for various vertices v. We will also have to count the contribution to $\sum$ from vertices in $R_1 \cup R_2$ which do not correspond to vertices in M.

First: given v' in M, how often does it appear as a first neighbor in standard neighborhoods? Once for each neighbor of v'. How much does it contribute to $\sum$ each time? Its charge, measured in M, but not more than +3, since as a first neighbor in a standard neighborhood its valence is limited to 9. Thus, if v' is a 5-vertex, its contribution to $\sum$ is (5) (-1); if v' is a 7-vertex, its contribution is (7) (+1), etc. Taking the sum over all $v' \in M$ gives us this total contribution to $\sum$ from first neighbors:

$$A = -5n_5 + 7n_7 + 16n_8 + 27n_9 + \sum_{k \geq 10} 3kn_k \qquad (13)$$

Now how often does v' appear as an inner second neighbor in standard neighborhoods? Again, the number of times is equal to the valence of v'. Each time, v' contributes to $\sum$ its original charge, limited by +2. So the contribution to $\sum$ from inner second neighbors must be

$$B = -5n_5 + 7n_7 + 16n_8 + 18n_9 + \sum_{k \geq 10} 2kn_k \qquad (14)$$

How many times does v' appear as an outer second neighbor in standard neighborhoods? The answer depends on the valences of the neighbors of v'. It is an outer second neighbor once for each 6-neighbor of v', twice for each 7-neighbor, three times for each 8-neighbor, and four times for each 9-neighbor. Neighbors of higher order should not be counted, because if v' appeared as a second neighbor in the abstract third neighborhood of v through the intermediary of a higher-order neighbor, it would be removed from the standard neighborhood of v.

If we add this up for all 5-vertices v', and multiply by -1, the charge of a 5-vertex, we get this contribution from 5-vertices appearing as outer second neighbors:

$$C_5 = -\ell_{56} - 2\ell_{57} - 3\ell_{58} - 4\ell_{59} \tag{15}$$

We can ignore the case of 6-vertices v', since their charge is 0. As outer second neighbors in standard neighborhoods, all other vertices have charge +1. Thus, we have these additional contributions from outer second neighbors:

$$\left.\begin{aligned} C_7 &= +\ell_{76} + 2\ell_{77} + 3\ell_{78} + 4\ell_{79} \\ C_8 &= +\ell_{86} + 2\ell_{87} + 3\ell_{88} + 4\ell_{89} \\ C_9 &= +\ell_{96} + 2\ell_{97} + 3\ell_{98} + 4\ell_{99} \\ &\cdots \end{aligned}\right\} \tag{16}$$

(Each line contains only four terms, but there may be arbitrarily many lines C_7, C_8, C_9, C_{10}, ...)

We still need to count the contribution from vertices v' in standard neighborhoods which do not correspond to vertices in M. These are the outer second neighbors added in part (4) of the construction of standard neighborhoods. Most of them are 6-vertices and can be ignored. There is one 7-vertex, contributing a charge of +1, for each time the construction occurs, which is once for each first neighbor of any vertex having valence greater than 9. Thus the contribution

$$D_1 = \sum_{k \geq 10} kn_k \tag{17}$$

There is also a 7-vertex for each of these new outer second neighbors which adjoins a typical interior third neighbor. On inspection we can see that this contribution is exactly

$$D_2 = m_{5:HL5L} \tag{18}$$

(Remember $H \geq 10$ and $L \leq 7$. This incidence number counts the configuration appearing in figure 88. Each time this configuration occurs, we find the vertex indicated by v in the figure. In the standard neighborhood of v, the top 5-vertex is a typical interior third neighbor, and the leftmost L-vertex in the figure, an outer second neighbor, adjoins it.)

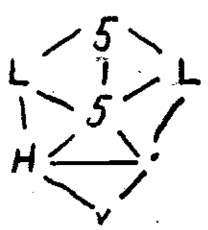


Figure 88

Before summarizing the contributions to $\sum$, we will evaluate X. How many interior third neighbors are there in standard neighborhoods? On inspection of figure 89 we see that the answer is twice the number of images of the configuration 5:L5L. Because of symmetries, $m_{5:L5L}$ is four times the number of such images. Therefore,

$$2X = m_{5:L5L} \tag{19}$$

Figure 89

We can now add the contributions in (13) through (19) to obtain a horrendous formula for T-3N in terms of incidence numbers.

$$\left.\begin{aligned}
T - 3N &= (A + B + \sum_{k\geq 5} C_k + D_1 + D_2) - 4N - 2X \\
&= -5n_5 \quad + 7n_7 + 16n_8 + 27n_9 + \sum_{k\geq 10} 3kn_k \\
&\quad -5n_5 \quad + 7n_7 + 16n_8 + 18n_9 + \sum_{k\geq 10} 2kn_k \\
&\quad -\ell_{56} - 2\ell_{57} - 3\ell_{58} - 4\ell_{59} \\
&\quad +\ell_{76} + 2\ell_{77} + 3\ell_{78} + 4\ell_{79} \\
&\quad +\ell_{86} + 2\ell_{87} + 3\ell_{88} + 4\ell_{89} \\
&\quad \vdots \\
&\quad + \sum_{k\geq 10} kn_k + m_{HL5L} \\
&\quad - 4n_5 - 4n_6 - 4n_7 - 4n_8 - 4n_9 - \sum_{k\geq 10} 4n_k
\end{aligned}\right\} \quad (20)$$

$$- m_{5:L5L} .$$

Rearranging freely, we obtain

$$\left.\begin{aligned} T - 3N = &- 14n_5 - 4n_6 + 10n_7 + 28n_8 + 41n_9 + \sum_{k\geq 10}(6k-4)n_k \\ &+ (\ell_{56} + \ell_{66} + \ell_{76} + \ldots\ ) - 2\ell_{56} - \ell_{66} \\ &+ 2(\ell_{57} + \ell_{67} + \ell_{77} + \ldots\ ) - 4\ell_{57} - 2\ell_{67} \\ &+ 3(\ell_{58} + \ell_{68} + \ell_{78} + \ldots\ ) - 6\ell_{58} - 3\ell_{68} \\ &+ 4(\ell_{59} + \ell_{69} + \ell_{79} + \ldots\ ) - 8\ell_{59} - 4\ell_{69} \\ &+ m_{5:HL5L} - m_{5:L5L} \ . \end{aligned}\right\} \quad (21)$$

Now use the identity (12) to replace the terms in parentheses. They become $6n_6$, $14n_7$, $24n_8$, and $36n_9$. Therefore

$$\left.\begin{aligned} T - 3N = &- 14n_5 + 2n_6 + 24n_7 + 52n_8 + 77n_9 + \sum_{k\geq 10}(6k-4)n_k \\ &- 2\ell_{56} - 4\ell_{57} - 6\ell_{58} - 8\ell_{59} \\ &- \ell_{66} - 2\ell_{67} - 3\ell_{68} - 4\ell_{69} \\ &+ m_{5:HL5L} - m_{5:L5L} \ . \end{aligned}\right\} \quad (22)$$

The identity (11) can be multiplied by 26, and turned into an inequality:

$$0 < 26n_5 \qquad - 26n_7 - 52n_8 - 78n_9 - \sum_{k\geq 10} 26(k-6)n_k$$

which can be added to (22) to give

$$\left.\begin{aligned} T - 3N < &+ 12n_5 + 2n_6 - 2n_7 \qquad - n_9 - \sum_{k\geq 10}(20k-152)n_k \\ &- 2\ell_{56} - 4\ell_{57} - 6\ell_{58} - 8\ell_{59} \\ &- \ell_{66} - 2\ell_{67} - 3\ell_{68} - 4\ell_{69} \\ &+ m_{5:HL5L} - m_{5:L5L} \ . \end{aligned}\right\} \quad (23)$$

<u>Lemma</u>. We may assume that, for $10 \leq k \leq 30$,

$$0 \leq (6k - 12)n_k - 6\ell_{5k} - 3\ell_{6k}, \tag{24}$$

and that, for $k > 30$,

$$0 \leq (6k)n_k \quad - 6\ell_{5k} - 3\ell_{6k}. \tag{25}$$

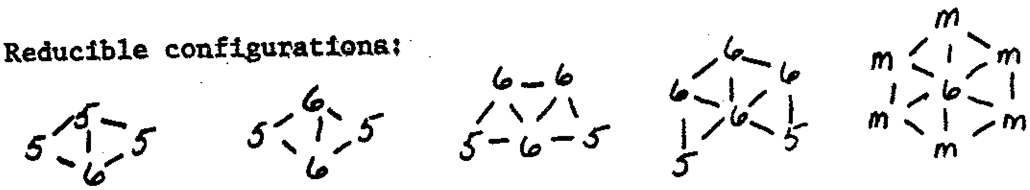
Proof: First assume $10 \leq k \leq 30$. The result follows immediately if we can show that, for each k-vertex, 6 times the number of 5-k edges incident to it, plus 3 times the number of 6-k edges incident to it, does not exceed (6k-12). But if this were false, we would have a k-vertex with either (k-1) 5-neighbors, or (k-2) 5-neighbors and a 6-neighbor, or (k-3) 5-neighbors and three 6-neighbors. These are configurations with values +1, +2, and +2 respectively (as can most easily be determined by using the second neighborhood formula as in figure 82). Therefore, should one of these configurations occur, the theorem is proved already.

This reasoning would also apply to values of k greater than 30, except that we would like to avoid using configurations whose vertices have such high valences. Still, the second inequality in the lemma is obvious. It becomes an equality only if each k-vertex is surrounded by 5-vertices. //

<u>Lemma</u>. $$0 \leq - 2n_6 - \ell_{56} + \ell_{66} + 2\ell_{67} + 3\ell_{68} + 3\ell_{69} + 3\ell_{6,10} + \dots + 1/2m_{6:55755}. \tag{26}$$

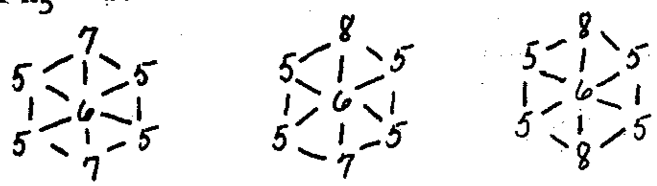
Proof: As in the previous lemma, we group each 6-vertex with the edges incident to it. We will group each occurrance of the configuration 6:55755 with its included 6-vertex. Let v be a 6-vertex. Let $x_k (k = 5,\dots)$ be the number of k-neighbors of v, and let $x_{6:55755}$

1) **Reducible configurations:**



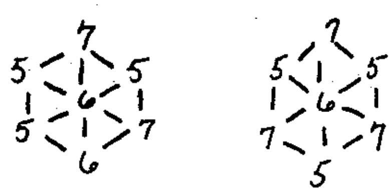
each $m \leq 6$.

2) **Cases with $x_5 = 4$:**



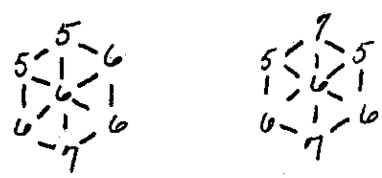
$x_5 = 4$	$x_5 = 4$	$x_5 = 4$
$x_7 = 2$	$x_7 = 1$	$x_8 = 2$
$x_{6:55755} = 4$	$x_8 = 1$	$\Sigma = 0$
$\Sigma = 0$	$x_{6:55755} = 2$	
	$\Sigma = 0$	

3) **Cases with $x_5 = 3$:**



$x_5 = 3$	$x_5 = 3$
$x_6 = 1$	$x_7 = 3$
$x_7 = 2$	$\Sigma = +1$
$\Sigma = 0$	

4) **Cases with $x_5 = 2$:**



$x_5 = 2$	$x_5 = 2$
$x_6 = 3$	$x_6 = 2$
$x_7 = 1$	$x_7 = 2$
$\Sigma = +1$	$\Sigma = +2$

In all cases with $x_5 \leq 1$, Σ is at least +3.

Figure 90

be the number of imbeddings of 6:55755 which involve v. We will show that

$$0 \leq - 2 - x_5 + x_6 + 2x_7 + 3x_8 + 3x_9 + 3x_{10} + + \ldots + 1/2x_{6:55755} \qquad (27)$$

and the lemma will follow by adding up the corresponding inequalities for all 6-vertices in M.

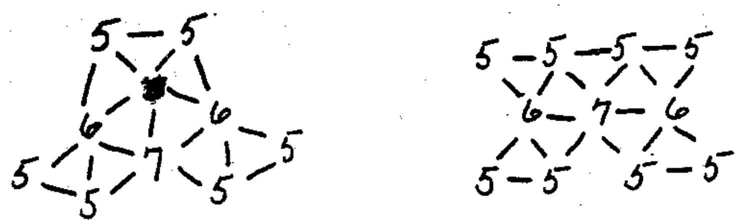
Two cautions are necessary. First, each 6-6 edge is counted in two groupings; but this is appropriate, since ℓ_{66} is twice the number of 6-6 edges. Second, 6:55755 contains an internal symmetry, so that $x_{6:55755}$ is always either 0, or 2, or 4.

To prove (27) we have to consider all possible neighborhoods of v. These are done in figure 90. Some are reducible, and may be ignored; in all of the others, (27) holds.

(In the figure, $\sum$ represents the right side of (27). Cases are omitted which so resemble other cases as to be redundant.) //

<u>Lemma</u>. $$0 \leq - 1/2m_{6:55755} + n_7 \ . \qquad (28)$$

Proof: Each 7-vertex may be included in only one image (i.e., two imbeddings) of 6:55755. If it were involved in two images, one of the configurations in figure 91 would appear in M; and these contain the reducible configurations 5:556(5)76 and 7:5565 respectively.//



<u>Figure 91</u>

<u>Lemma.</u>

$$0 \leq -12n_5 + 3\ell_{56} + 4\ell_{57} + 6\ell_{58} + 6\ell_{59} + 6\ell_{5,10} + \dots + m_{5:L5L} - m_5\mathrm{HL5L} + m_{5:L5L} \quad (29)$$

Proof: This time we will group the objects being counted around the 5-vertices. Given a 5-vertex v, let x_k be the number of k-neighbors of v, and let x_{5LL5L} (etc.) be the number of imbeddings of $x_{5:L5L}$ (etc.) which include v as the central 5-vertex. We show that for each v,

$$0 \leq -12 + 3x_6 + 4x_7 + 6x_8 + 6x_9 + 6x_{10} + \dots + x_{5:L5L} - x_{5:HL5L} + x_{5:5H557} \quad (30)$$

-- and the result follows.

Neighborhoods of 5-vertices appear in figure 92. It is pretended that the only possible neighbors of the 5-vertex are 5-, 6-, 7-, and 10 vertices, since replacing any other neighbor with a 10-neighbor could not increase the sum in (30).

Each neighborhood is found either to be reducible, or to satisfy (30). //

<u>Lemma.</u>

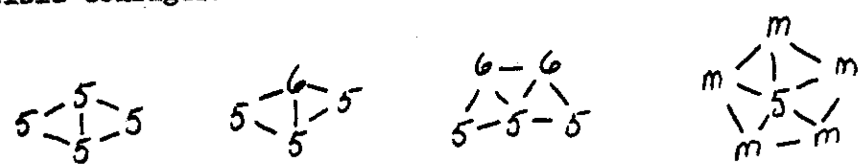
$$0 \leq -m_{5:5H557} + n_7 \quad (31)$$

Proof: Each 7-vertex may be included in only one image of 5:5H557. Otherwise the reducible configuration 7:5555 would occur. //

Adding inequalities (23), (24) and (25) for all values of k, (26), (28), (29), and (31) gives

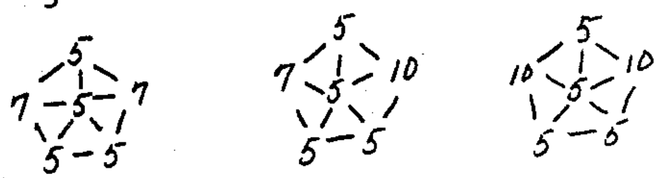
$$T - 3N < -n_9 - \sum_{k=10}^{30} (14k-140)n_k - \sum_{k>30} (14k-152)n_k - 2\ell_{59} - \ell_{69} \quad (32)$$

1) **Reducible configurations:**



each $m \leq 6$.

2) **Cases with $x_5 = 3$:**

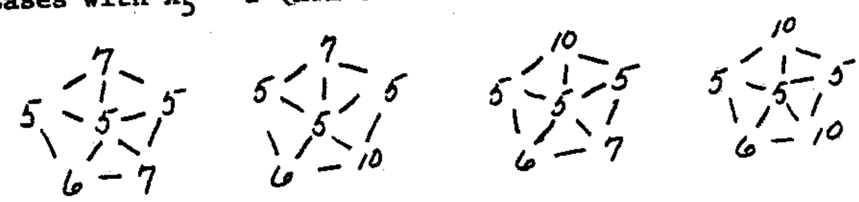


$x_7 = 2$
$x_{5:L5L} = 6$
$\Sigma = +2$

$x_7 = 1$
$x_{5:L5L} = 2$
$x_{10} = 1$
$x_{5:5H557} = 1$
$x_{5:HL5L} = 1$
$\Sigma = 0$

$x_{10} = 2$
$\Sigma = 0$

3) **Cases with $x_5 = 2$ (non-consecutive 5-neighbor):**



$x_6 = 1$
$x_7 = 2$
$x_{L5L} = 4$
$\Sigma = +3$

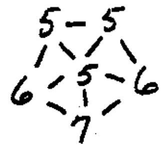
$x_6 = 1$
$x_7 = 1$
$x_{10} = 1$
$x_{L5L} = 2$
$x_{HL5L} = 1$
$\Sigma = +2$

$x_6 = 1$
$x_7 = 1$
$x_{10} = 1$
$\Sigma = +1$

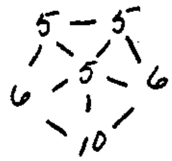
$x_6 = 1$
$x_{10} = 2$
$\Sigma = +3$

Figure 92

4) Cases with $x_5 = 2$ (consecutive 5-neighbors):



$x_6 = 2$
$x_7 = 1$
$x_{5:L5L} = 4$
$\Sigma = +2$



$x_6 = 2$
$x_{10} = 1$
$x_{5:L5L} = 4$
$x_{5:HL5L} = 2$
$\Sigma = +2$

$x_6 = 2$
$x_{10} = 1$
$x_{5:L5L} = 2$
$x_{5:HL5L} = 1$
$\Sigma = +1$

In all cases with $x_5 \leq 1$, Σ is at least +3.

Figure 92
(continued)

Clearly this implies that $T - 3N < 0$, and Theorem 7 is proved. //

There remains a detail necessary for the corollary to Theorem 7, as stated on page 69: we must show that the configuration claimed by Theorem 7 may be chosen with no vertex of valence greater than 30. It will suffice that some standard neighborhood containing no vertex of valence greater than 30 has value less than 3.

(Actually, a much smaller number could be used here in place of 30. The proof below will apply without change if the valence 21 is used, and more careful arguments culd justify a still smaller number. But the exact value of the upper bound does not matter, since it is too large to be practical in any case; only the fact that there is an upper bound is important.)

The only vertex in a standard neighborhood which could have such a large valence is the central one. Let T' be the total value of standard neighborhoods of vertices having valence not greater than 30, and let N' be the number of such vertices. We will show that $T' - 3N' < 0$.

__Lemma__. A vertex v of valence $k > 30$ cannot have a standard neighborhood with value less than $-4k-1$.

Proof: Assume that v has k first neighbors and k second neighbors, each of valence 5, and that for each second neighbor, there is an interior third neighbor. These assumptions are inconsistent, but clearly establish a lower bound for the value of the standard neighborhood, which according to the third neighborhood formula, is $-4k-1$. //

It follows from the lemma that

$$(T - T') \geq (-4k - 1)(N - N')$$

or

$$(T' - T) \leq (+4k + 1)(N - N') \qquad (33)$$

Since

$$(N - N') = \sum_{k>30} n_k \qquad (34)$$

we also have

$$(T' - T) \leq (+4k + 1) \sum_{k>30} n_k \ . \qquad (35)$$

We repeat the inequality (32), dropping some irrelevant terms:

$$(T - 3N) < - \sum_{k>30} (14k - 152)n_k \qquad (36)$$

Multiply (34) by 3 and add the result to (35) and (36):

$$3N - 3N' = \sum_{k>30} 3n_k \qquad (37)$$

$$T' - 3N' < \sum_{k>30} (-7k + 153)n_k \qquad (38)$$

and since $7k > 153$ for $k > 30$, this means that $T' < 3N'$ and the last detail is finished. //

APPENDIX

"The Four-Color Theorem for Small Maps"

This article is presented exactly as submitted to the _Journal of Combinatorial Theory (Series B)_ where it is scheduled to appear in mid-1975. Its only theorem is that every map with fewer than 52 vertices may be four-colored.

Although the article is self-contained, its idea is closely related to that of Chapter IV. It contains a definition of "value" -- in fact, two definitions -- and it is worth showing the connection between these and the definition in Chapter IV.

Let v be a vertex in a map, and let the _immediate neighborhood_ of v be the configuration containing in its interior v, and those first neighbors of v having valence less than 8. In the first part of "Small Maps," the "value of v" is defined, roughly, to be the Chapter IV-value of the immediate neighborhood of v. (This definition is exactly equivalent if v does not have consecutive neighbors of valence at least 8.)

Later in the article a number of adjustments are made to the "value of v." These are mostly arbitrary, but they still have a common theme: generally, the value of v is decreased only if v has another neighborhood, slightly larger than the immediate neighborhood but with a smaller Chapter IV-value.

THE FOUR-COLOR THEOREM

FOR SMALL MAPS

by Walter Stromquist

Department of the Treasury
Washington, D. C.

Revised April 29, 1975

ABSTRACT:

Any map with fewer than 52 vertices contains a "reducible configuration;" therefore, any such map may be vertex-colored in four colors. This is proved by defining the "value" of each vertex, according to the valences of its neighbors, in such a way that low values lead to reducible configurations, and high values lead to large maps.

ACKNOWLEDGEMENT:

The author wishes to acknowledge the helpful suggestions of Professor Andrew M. Gleason.

THE FOUR-COLOR THEOREM FOR SMALL MAPS

The four-color conjecture is true for all maps of fewer than 52 regions. To prove this, we will show that every such map contains a reducible configuration; that is, one which is known not to exist in a minimal five-color map.

At this writing, the last published result of this type was for maps of fewer than 40 regions, announced by Ore and Stemple in 1968. Since then, the number has been raised to 45 by this author, and to 48 by Jean Mayer. The force behind these increases has been the discovery of new reducible configurations, especially by H. Heesch, F. Bernhart, and Mayer.

Several new reductions have been discovered recently by F. Allaire and E. R. Swart; however, these reductions became available in time for only minor revisions in this paper. The paper by Allaire and Swart will appear in this Journal.

We will treat the problem in the dual form; that is, we will be coloring vertices. The four-color conjecture is that the vertices of any planar graph may be colored, using four colors, so that neighboring vertices have different colors. We will usually be considering a minimal counterexample, so we will assume certain well-known properties of any such map: (1) its faces are all triangles; (2) no vertex has valence less than 5; and (3) there are no minimal circuits of length less than 6 except those which surround single vertices.

We will describe configurations by a common shorthand involving the valences of their vertices. For example, "5:66755" refers to a 5-vertex, whose neighbors, in cyclical order, have valences 6, 6, 7, 5, 5.

We also use parentheses in the manner introduced by Ore and Stemple: e.g., "5:6(5)6755" is the same configuration as before but with an additional 5-vertex adjoining the two 6-vertices. Note that a configuration may be described in different ways; e.g., 5:565 and 6:555, depending on which vertex we want to think of as the "center". Examples of this notation are shown in figure 1.

We will use the following symbols to represent undetermined valences:

X - any valence

m - (minor) valence 5 or 6

G - greater than 5

M - (major) valence greater than 6

E - eight or greater

The <u>first neighborhood</u> of a vertex consists of the vertex and its neighbors. The <u>charge</u> on a vertex is six less than its valence. Charges are sometimes easier to discuss than valences, since the total charge in every map is -12, and hence the average charge is close to zero. (This is a result of Euler's Formula. Our definition of charge is the negative of the usual definition.)

For the background of reducible configurations, see, for example, Ore's book [1]. For our purposes, a reducible configuration is one which cannot appear in a map which is minimal counterexample to the four-color conjecture. The reductions used in this paper, including the new ones, are listed in a brief appendix.

To introduce the method of this paper, we will apply it to a simpler case: that of graphs of less than 28 vertices.

§1. The Theorem for Graphs of Fewer Than 28 Vertices

Given a map with fewer than 28 vertices, we will show that it contains a reducible configuration.

First, define the value of a vertex to be the sum of the charges of its neighbors, except that neighbors of valence greater than 7 will be counted as having charge +1. (The term "value" will be given a new definition when we develop the proof for larger maps.)

It is easy to compute the total value of all the vertices in the map. For example, a 5-vertex is a neighbor to exactly five vertices, and has charge -1, so it contributes -5 to the total. If V is the total value, and n_k is the numbers of k-vertices in the map, we find that

$$V = -5n_5 \qquad + 7n_7 + 8n_8 + 9n_9 + \dots \quad . \tag{1}$$

By Euler's Formula we can show that

$$84 = 7n_5 \qquad - 7n_7 - 14n_8 - 21n_9 - \dots \tag{2}$$

and from our assumption about the size of our map we know that

$$-84 < - 3n_5 - 3n_6 - 3n_7 - 3n_8 - 3n_9 - \dots \quad . \tag{3}$$

Adding formulae (1), (2), and (3), we get

$$V < - 1n_5 - 3n_6 - 3n_7 - 9n_8 - 15n_9 - \dots \quad . \tag{4}$$

It follows that our map contains a 5-vertex of value less than -1, or a 6-vertex of value less than -3, or a 7-vertex of value less than -3. (It is impossible to have an 8-vertex of value less than -9, etc.)

But all such structures include well-known reducible configurations -- in fact, one of the following must be present:

5:555	6:565
5:565	7:5555
5:55666	7:5655

Any 7-vertex surrounded by three 6-vertices and four 5-vertices. (One of these, 7:5655, is not really well-known; it is recent, due to Frank Bernhart. For this particular application, 7:556555 would suffice.)

For example, see figure 2, which shows the ways that a 5-vertex can have value -2.

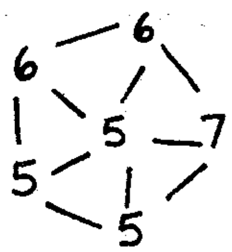
Since our map contains a reduction, it can't be a minimal 5-color map. We have therefore proved the four-color theorem for maps of fewer than 28 vertices.

§2. Introduction to the Full Proof

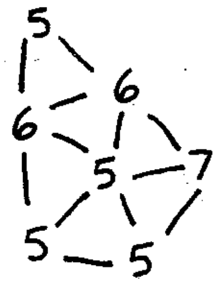
We will now extend the result to maps of fewer than 52 vertices. We will use a different definition of the value of a vertex, and the survey of neighborhoods will take longer, but otherwise the pattern of the proof is identical to the short proof in the last section.

In general we will make a great many adjustments to the definition of value, but we will still preserve our formula for the total value in our map:

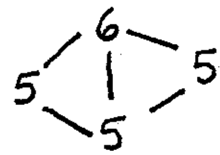
$$V = -5n_5 \quad + 7n_7 + 8n_8 + 9n_9 + \dots \tag{1}$$



5:66755



5:6(5)6755



5:565
6:555

Figure 1

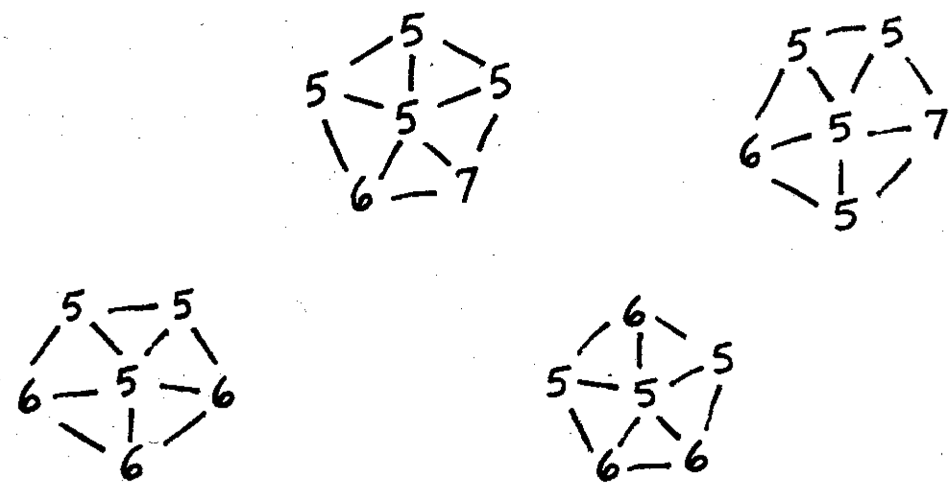


Figure 2. 5-vertices with value -2

Our use of Euler's Formula will be

$$78 = 6\ 1/2n_5 \qquad - 6\ 1/2n_7 - 13n_8 - 19\ 1/2n_9 - \dots \tag{5}$$

and we will consider maps of fewer than 52 vertices; that is, maps which satisfy

$$-78 < -1\ 1/2n_5 - 1\ 1/2n_6 - 1\ 1/2n_7 - 1\ 1/2n_8 - \\ - 1\ 1/2n_9 - \dots . \tag{6}$$

We will add formulae (1), (5), and (6) to obtain

$$V < \qquad -1\ 1/2n_6 - 1n_7 - 6\ 1/2n_8 - 12n_9 - \dots . \tag{7}$$

It follows that somewhere in our map is a k-vertex with value less than the coefficient of n_k in formula (7). We will finally show (by an exhaustive survey of possible first neighborhoods) that any such vertex has in its vicinity a reducible configuration.

The result will be the following theorem.

Theorem: Any map with fewer than 52 vertices contains a reducible configuration. Therefore, the four-color conjecture holds for all such maps.

§3. The Value of a Vertex

In this section we will define the value of a vertex in our map, illustrate it by some examples, and show that the total value of all the vertices in the map is given by formula (1). Our definition is highly unnatural, many complicated provisions having been contrived to help with the difficult cases in the sebsequent analysis.

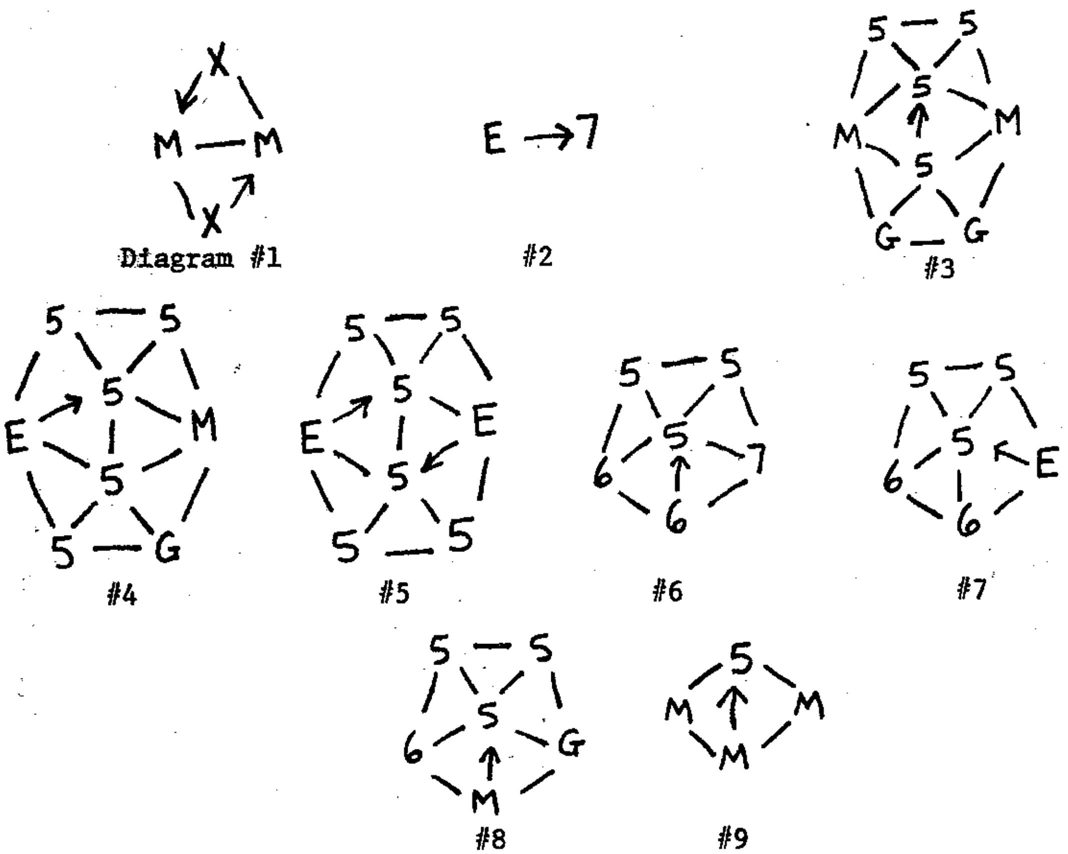


Figure 3a. Adjustments to value

(± 1/4 only)

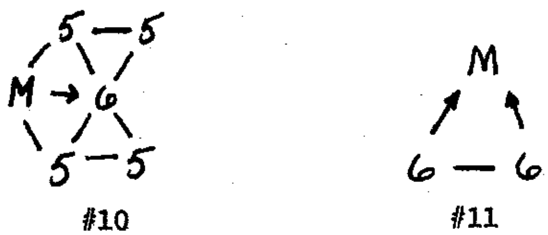


Figure 3b. Adjustments to value of ± 1/4

The value of a vertex begins with the basic definition from § 1, and is then increased or decreased according to the valences of other vertices in the vicinity.

Definition: The value of a vertex is the sum of the charges of its neighbors, except that neighbors with valence greater than 7 are counted as having charge +1, and

a) for each time a configuration illustrated in figure 3a appears in the map, the value of the vertex at the base of each arrow is decreased by 1, and that of the vertex at the point of each arrow is increased by 1; and

b) for each time a configuration illustrated in figure 3b appears in the map, the values of the vertices indicated by the arrows are increased or decreased by 1/4 . (If a major vertex has three consecutive 6-neighbors, diagram 12 is allowed to apply twice to the same 6-M edge.)

Figure 4 gives some examples of vertices and their values. It is clear that the sum of the values in the map is determined entirely by the first paragraph of the definition; that none of the diagrams which follow it can change the total. Therefore the total is still given by formula (1).

The purpose of the diagrams is to shift value from vertices with a "surplus" to those which are more in need. We are trying to show that neighborhoods with low values produce reductions; this task is eased if we shift value from neighborhoods which are likely to contain reductions, to neighborhoods which are not. For example, 5-5-5 triads tend to generate more low values than reductions, so we have introduced diagrams 3 through 8, and also diagram 9, to shift values toward these triads.

On the other hand, a vertex with consecutive major neighbors tends to have surplus value. Diagram 1 shifts this surplus to the major neighbors themselves. One effect is that a series of major neighbors to a given vertex contribute only +1 to the value of that vertex.

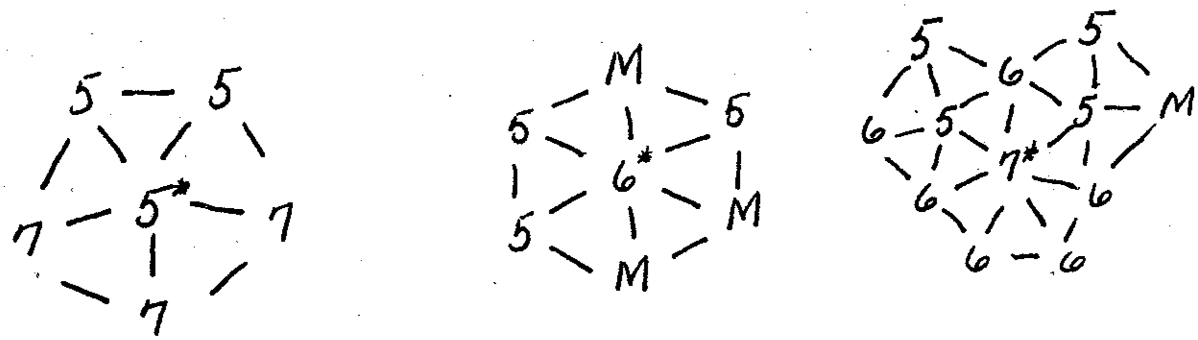
§4. Survey of Neighborhoods: 5- and 6-Vertices

We must now examine all possible neighborhoods, to show that any vertex which is not in (or near) a reducible configuration must have value at least equal to the corresponding coefficient in formula (7). This survey will complete the proof.

It will do no harm to underestimate the value of a vertex, so we will sometimes ignore diagrams which could increase the value of a vertex, or assume that "negative" diagrams apply when they might not.

We begin with 5-vertices. We must show that, in the absence of reductions, the value of any 5-vertex v is at least 0.

Because of the reductions 5:555, 5:565, 5:5665, and 5:mmmmm , we know that, except for possibly one pair of consecutive 5-neighbors, any two 5-neighbors of v must be separated by an M-vertex. Hence the charges of the neighbors must total at least -1, if v is in a 5-5-5 triad; or 0, if it is not. This remains true when we take into account diagram 1, and count only +1 toward the value of v for each series of consecutive M-neighbors.



*value = +1
-2 (diagram #1)
+1 (#9)
= 0

*value = 0
-1 (#1)
= -1

*value = -2
-1/4 (#10)
+1 1/2 (#11)
= -3/4

Figure 4. Examples of value

Now, if v is in a 5-5-5 triad, its value is increased to 0 by exactly one of diagrams 3-9 (unless v has neighbors 55M6M, in which case its value is already 0). Here we make implicit use of the reduction 7:5555.

It remains only to check for diagrams (other than 1) which can decrease the value of v. There is only one -- diagram 3, which is so specific that we can determine the value of any 5-vertex to which it applies. It must always be at least zero.

The case of a 6-vertex v is equally direct. We must show that the value of v is at least -1 1/2 . Now we have reductions 6:555, 6:565, 6:56665, 6:mmmmmm, and 6:5665 (the last being a recent addition by Bernhart) which again give us the result that, except for consecutive pairs of 5-neighbors, any two 5-neighbors must be separated by an M-neighbor. Hence the sum of the charges of the neighbors, even adjusting for diagram 1, is at least zero, minus 1 for each consecutive pair of 5-neighbors. The value may also be reduced by diagrams 6 and 11 and increased by diagram 10.

If there are two consecutive pairs of 5-neighbors, then diagram 10 applies twice, and diagrams 6 and 11 do not. The value of v is exactly -1 1/2 .

If there is one consecutive pair of 5-neighbors, then diagram 6 may not apply (becuase of the reduction 7:5655). Diagram 11 would have to apply three times to v to bring its value below -1 1/2 . There is only one way for this to happen, shown in figure 5; and in this case, the surfeit of major neighbors causes the value to be -3/4 .

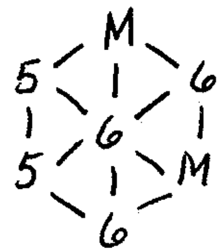


Figure 5

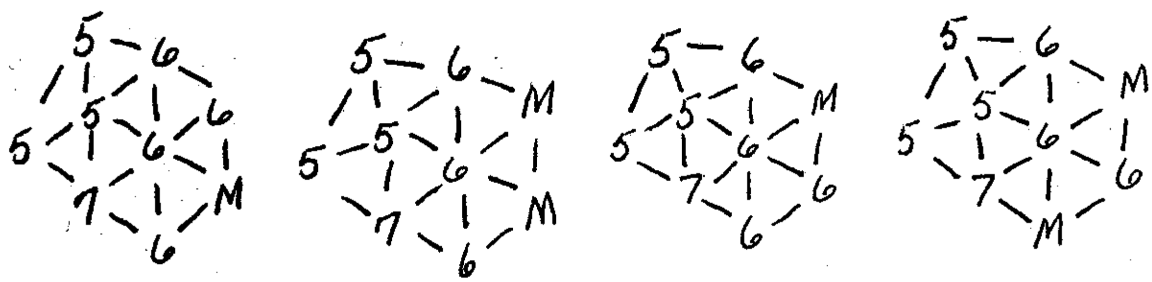


Figure 6

Suppose there are no consecutive pairs of 5-neighbors but diagram 6 applies to v. Again, we have no problem unless diagram 11 applies three times. There are four ways for this to occur (figure 6), and in each, the value of v turns out to be -3/4. Diagram 6 may not apply twice to the same 6-vertex (figure 7), because of the reduction 7:6(5)655 (Allaire-Swart).

Without diagram 6 or a consecutive pair of 5-neighbors, v must have value well above -1 1/2.

§5. 7-Vertices

The value of a 7-vertex v may be decreased by diagrams 1, 8, 9, and 10, and increased by diagrams 1, 2, and 11. We must show that in the absence of reductions, the value of v is at least -1.

Diagram 10 may apply only once to a single 7-vertex (reductions: 7:5655 and 5:556(5)76). In each case in which it appears, it turns out that the value is still much greater than -1, or diagram 11 applies at least once, in effect cancelling diagram 10. This fact requires checking, but in most cases we will leave this to the reader. Usually, we will ignore the effects of diagrams 10 and 11 on 7-vertices.

We will use the term "covered" to refer to any 5-neighbor which is part of diagram 8 or 9. A 5-neighbor which is part of a series of consecutive 5-neighbors cannot be covered.

We can summarize the effect of diagrams 1, 2, 8, and 9 on the value of v as follows: the value is the sum of +1 for each major neighbor or series of major neighbors, +1 (additional) for each 7-neighbor, +2 (additional) for each E-neighbor, -1 for each uncovered

5-neighbor, and -2 for each covered 5-neighbor.

We classify the possible 7-neighborhoods by the number of 5-neighbors. We will frequently use the reductions 7:5555 and 7:5655 without citation.

<u>Case 1 - Five 5-neighbors</u>. This can only occur if v has neighbors 555M55M, in which case its value is at least -1.

<u>Case 2 - Four 5-neighbors</u>. Because of the reduction 7:56655 (Allaire-Swart), there must be at least two non-consecutive M-neighbors. Therefore if the value of v is below -1, two of the 5-neighbors must be covered. This can only occur in the configuration shown in figure 8. This contains a reduction, 6:5(5)5x5(5)5; and furthermore, each M-neighbor must actually be an E-neighbor (because of the reduction 7:5(5)755, by Heesch). So (even with diagram 10) the value of v in this configuration is at least -1/4 .

<u>Case 3 - Three 5-neighbors, including at least two consecutive</u>. There must be a major neighbor (by reductions of 7:5556666, Heesch, and 7:56655, Allaire-Swart), so for the value to be less than -1, one 5-neighbor must be covered and there must be no major neighbors except for one 7-neighbor. The three possibilities are shown in figure 9. The first contains the reduction 7:5(5)755 (Heesch); the other two contain the Allaire-Swart reduction 7:56655.

<u>Case 4 - Three 5-neighbors, no two of them consecutive</u>. This is the hardest case, and we will divide it into subcases according to the number and arrangement of major neighbors. First, if there are no major neighbors, we have 7:56565 which is reducible (Allaire-Swart). If there are three major neighbors, the value of v must be at least zero even

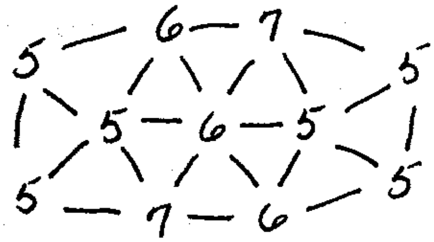


Figure 7

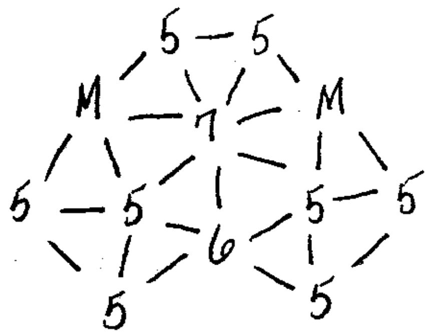


Figure 8

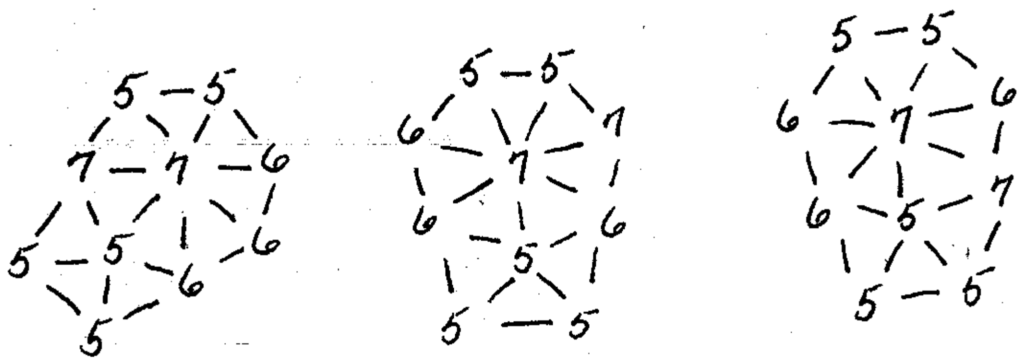


Figure 9

if all three 5-neighbors are covered. So there remain the cases of one or two major neighbors.

Case 4a - Neighbors 565M566. The first and third 5-neighbors cannot be covered, because of the reduction 7:5(5)665 (Allaire-Swart). If the second 5-neighbor is not covered, the value of v is -1. If the second 5-neighbor is covered but the M-neighbor is not a 7-neighbor, we still have value -1. We are left with figure 10, which contains the Allaire-Swart reduction 5:7(5)7655.

Case 4b - Neighbors 565656M or 56565MM. These include the configuration 7:56565, which is reducible (Allaire-Swart).

Case 4c - Neighbors 565M5M6. The first 5-neighbor may not be covered (reduction - 5:5567(5)6) and the second 5-neighbor may not be covered if the first M is to be a 7-neighbor (reduction - 5:7(5)7655, Allaire-Swart). Therefore the value may not be below zero.

Case 4d - Neighbors 565M56M. Because of the reduction 6:5(5)5x5(5)5, it is impossible for all three 5-neighbors to be covered. The worst that can happen is that two 5-neighbors are covered and that the major neighbors are 7-neighbors (for example, figure 11). The value of v would still be -1. But even this case is impossible, since it contains the reduction 5:7(5)7655.

Case 4e - Neighbors 5M5M566. There is no problem here unless all three 5-neighbors are covered, but this is prevented by the reduction 7:5(5)665 (Allaire-Swart).

Case 5 - Two consecutive 5-neighbors. If there are no major neighbors, we have the reducible configuration 7:5566666. If there is a major neighbor, then v has value at least 0, since neither 5-neighbor may be covered.

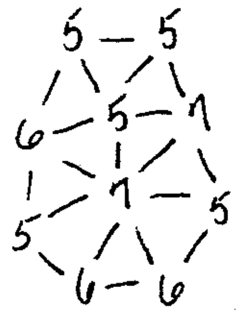


Figure 10

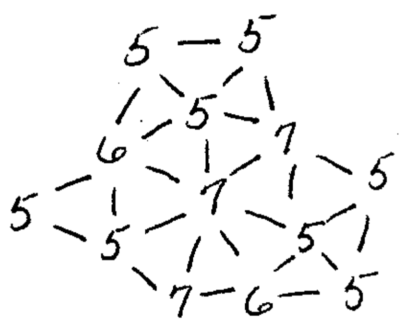


Figure 11

<u>Case 6 - Two non-consecutive 5-neighbors</u>. If there is a major neighbor, then both 5-neighbors must be covered to bring the value below -1. Using the reductions 7:5(5)665 (Allaire-Swart) and 6:5(5)5x5(5)5, we may limit the possibilities to the two shown in figure 12. In each case, diagram 11 applies at least four times and diagram 10 does not apply at all; so v finally has value at least -1.

If the neighbors are 5665666, then neither 5-neighbor may be covered because of the reduction 7:5(5)665 (Allaire-Swart). Diagram 11 applies six times, so v has value $-1/2$.

If the neighbors are 5656666, then again the 5-neighbors are not covered (reduction: 7:5(5)656666; Mayer). The same reduction prevents diagram 10 from applying. Since diagram 11 applies six times, v must have value $-1/2$.

<u>Case 7 - One or no 5-neighbors</u>. If there is an M-neighbor, or if there is not a covered 5-neighbor, v must have value at least -1. The only remaining case is shown in figure 13, but this configuration contains a reducible Errera Circuit.

§6. Vertices of Higher Valence

We leave the cases of vertices of higher valence to the reader. They are theoretically easy, because of the extreme negative coefficients attached to $n_8, n_9, \ldots$ in formula (7). However, the case of 8-vertices is still time-consuming because of the multiplicity of cases and the relative scarcity of known reductions involving high-order vertices. The following reductions are helpful for 8 vertices: 8:555x555, 8:55565, 8:55655 (Allaire-Swart), 8:555(5)5(6)6 (a new reduction by the author),

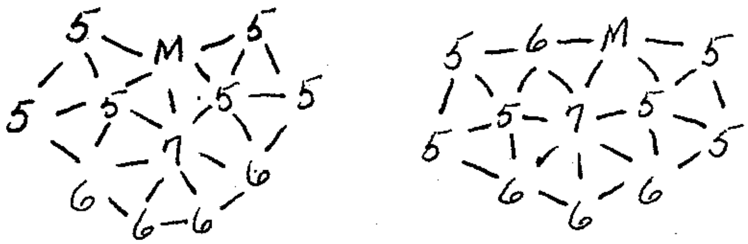


Figure 12

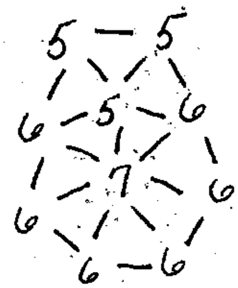


Figure 13

6:6(5)5x5(5)5 (Bernhart), and several of Mayer's reductions.

§7. Potential for Related Results

We have used the assumption that a map has fewer than 52 vertices to prove that it contained a reducible configuration. This assumption entered through equation (6), namely,

$$-78 < -1\ 1/2n_5 - 1\ 1/2n_6 - 1\ 1/2n_7 - 1\ 1/2n_8 - 1\ 1/2n_9 - \dots \quad (6)$$

Nothing in the structure of the proof required that the coefficients in (6) be equal. By starting with a different linear relation in place of (6), we would finish with a different formula in place of (7); but it might still be possible to define "value" in such a way that the proof would go through. In this way we could prove a variety of interesting results about the numbers $n_5, n_6, n_7, \dots$ in a minimal five-color map.

§8. Appendix: The Reducible Configurations

The following are the reductions used in this paper. The first group can be found listed in the paper by Ore and Stemple [2]; the bibliography of that paper leads to their proofs. The second group have been published by Heesch in [3] and [4]. The third group appear in Frank Bernhart's thesis [5], but at this writing, have not yet appeared in print. The next groups are by Jean Mayer [6], [7], and by F. Allaire and E. R. Swart [8]. When the last paper appears, it will become the best reference for all of the reductions.

One new reduction is also listed. Its "reduced configuration" is given in figure 14, which is drawn in the "face coloring" form, meaning that a face in figure 14 corresponds to a vertex anywhere else in the paper. The shaded area in figure 14 is to be collapsed into a single face in the reduced configuration.

List of Reductions

I. Cited by Ore and Stemple [2]

5:555	6:56665
5:565	6:5(5)5x5(5)5
5:5665	6:mmmmmm
5:mmmmm	7:5555
5:556(5)76	7:5566666
6:565	8:5555655

7:(three 6-neighbors and four 5-neighbors)
Errera Circuits, such as 7:6(5)5(5)66666
Winn Circuits, such as 6:65(5)765(5)7
and 7:5567(5)566

II. Reductions by Heesch [3], [4]

7:5(5)755
7:5(5)757(5)5

III. Reductions by Bernhart [5]

6:5665
7:5655
6:6(5)5x5(5)5

IV. Reductions by Mayer [6], [7]

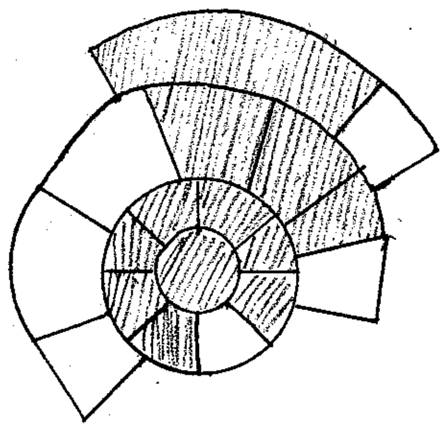
5:566(5)76
5:5667(5)6

V. Reductions by Allaire and Swart [8]

5:7(5)7655	7:5(5)665
7:56565	8:56555
7:6(5)655	8:55655
7:56655	8:555x555

VI. A new reduction (see figure 14)

8:555(5)5(6)6



8:555(5)5(6)6

Figure 14. The reduced configuration

REFERENCES

1. O. Ore, "The Four Color Problem," Academic Press, New York, 1967.

2. _______, and J. G. Stemple, Numerical Calculations on the Four Color Problem, J. Combinatorial Theory, 8(1970) 65-78.

3. H. Heesch, "Untersuchungen zum Vierfarbenproblem," Mannheim, 1969.

4. _______, Chromatic Reduction of the Triangulations T_e, $e = e_5 + e_7$, J. Combinatorial Theory (B), 13 (1972) 46-55.

5. F. Bernhart, thesis, Kansas State University.

6. J. Mayer, "Nouvelles reductions dans le probleme des quatre couleurs," Montpellier, Université des Sciences et Techniques due Languedoc, Cahiers mathematiques de Montpellier, No. 1.

7. _______, "Inegalites nouvelles dans le probleme des quatre couleurs," to appear.

8. F. Allaire and E. R. Swart, "A Systematic Approach to the Determination of Reducible Configurations in the Four-Colour Conjecture," to appear.