On the History and Solution of the Four-Color Map Problem

John Mitchem

John Mitchem is Professor of Mathematics and Chairman of the Mathematics Department at San Jose State University where he has been since receiving his Ph.D. in 1970 from Western Michigan University. His research has been in graph theory including various coloring problems, but he has not *worked on the Four-Color Problem.*

In the summer of 1976, the University of Illinois had the audacity to use their postage meter to print the words "Four Colors Suffice" on the outgoing mail. The words referred to the recently announced computer-aided proof of the very famous Four-Color Conjecture by Kenneth Appel and Wolfgang Haken (with a significant contribution by John Koch) at the Urbana campus. Their announcement was originally greeted with a rather large degree of skepticism by the mathematical community. This skepticism was not without cause. For about a century there have been numerous rumors and claims of proofs or counterexamples to the conjecture. In fact, in 1971 a different computer-aided "proof" by Shimomoto was announced but later withdrawn after it failed to survive close scrutiny. Furthermore Hakin's name had somehow been associated (perhaps unfairly) with this ill-fated work.

Despite this initial skepticism the Appel-Haken proof has withstood over three years of close examination. It has been published by the highly respected Illinois Journal of Mathematics [1] and has been accepted by a number of experts throughout the world.

Let us now review the statement and history of the Four-Color Theorem. The conjecture was solely the discovery of Francis Guthrie in the early 1850's. While a student at University College, London, Guthrie discovered that he could color a map of England's counties with only four colors in such a way that each county has exactly one color, and two counties which share a common boundary line will have different colors. Furthermore he tried to prove that the counties (i.e., regions) of any map on the plane could be so four-colored. Not entirely satisfied with his "proof," Francis discussed his four-color problem with his brother Frederick, who communicated it to his teacher Augustus De Morgan. On October 23, 1852, De Morgan discussed the problem in a letter to Sir William Hamilton. Hamilton, perhaps showing more wisdom than many other mathematicians through the decades, refused to become involved with the problem.

This is a revised version of a lecture given on April 22, 1978, at the second annual Saturday Morning Mathematics Conference sponsored by Dean L. H. Lange at San Jose State University.

Until recently various books and papers stated that the problem of four-coloring maps was known to ancient cartographers and was discussed by Möbius in the 1840's. K. O. May [12] put these misconceptions to rest. There is no evidence that mapmakers knew or were interested in the problem, and it is clear that Möbius was interested in a problem which was only superficially related to Guthrie's problem. In their excellent book, [6] Biggs, Lloyd, and Wilson explain how the erroneous Möbius connection probably occurred.

Apparently word of the problem traveled by the mathematical grapevine. By the 1860's the American logician and philosopher C. S. Pierce was working on the problem. Arthur Cayley, at a London Mathematical Society meeting in June, 1878, asked if anyone had proved the conjecture. In 1879, in the American Journal of Mathematics, A. B. Kempe, a London barrister, gave the first published "proof" of the Four-Color Conjecture. That is, Kempe "proved" that any map on the surface of a sphere can be four-colored. For a proper coloring of a map with four colors see Figure 1, and Figure 2 has an unacceptable coloring of the same map because Idaho and Nevada share a boundary edge but both have color 1. In addition to the

Figure 1.

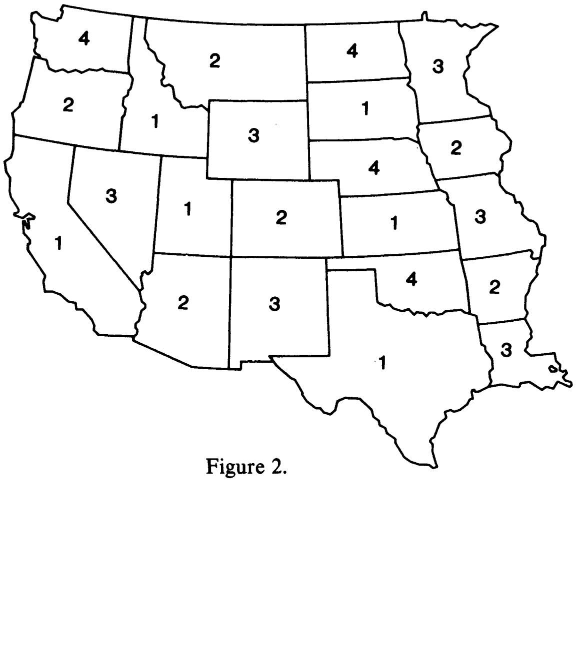

Figure 2.

"proof" in the American Journal, Kempe announced his result without proof in the prestigous British journal *Nature*. Kempe's work was apparently accepted and admired widely. In fact because of this work, and with the support of Cayley and others, Kempe was made a Fellow of the Royal Society. He brought out improved versions of the "proof." P. G. Tait at the University of Edinburgh soon outlined yet another "proof." Lewis Carroll (C. L. Dodgson) made a game in which one player made a map which his opponent needed to four-color. In 1886 the headmaster of an English boys' school posed the problem as a challenge to his school boys, but with the proviso that "no solution may exceed one page, 30 lines of manuscript, and one page of diagrams." (Keep in mind that the only current widely-accepted proof by Haken and Appel is well over 100 pages and used hundreds of hours of computer time.) In 1889 the Bishop of London, who later became the Archbishop of Canterbury, published his "proof" of the Four-Color Theorem in the Journal of Education.

In 1890 Heawood [11] pointed out a serious flaw in Kempe's work. He also showed how Kempe's method could be used to prove that any map on the plane can be five-colored. Kempe admitted that there was a flaw, and, furthermore, that he was unable to repair it. As perhaps often happens to bearers of bad news, Heawood received no public acclaim and certainly was not made a Fellow of the Royal Society.

We now will consider Kempe's method and its important successful variation almost 100 years later by Haken and Appel. For convenience we represent, as did many mathematicians who considered the problem, any map on the plane as follows: Inside each region we place a tiny circle called a *vertex*. If two regions share a common boundary line, we join their vertices with a curve through that boundary. These curves are called *edges* and the resulting collection of vertices and edges is a *graph*. Since no two edges of the graph intersect, we say that it is a *plane graph*. Thus any map on the plane can be represented as a plane graph with the vertices representing regions and the edges showing which regions are adjacent. Figure 3 illustrates the derivation of the plane graph from the map in Figure 1. That which remains after the removal of the boundaries of the original map is the plane graph. Two vertices of a graph are called *adjacent* if there is an edge joining them. Thus instead of coloring regions of a map we can color vertices of a plane graph. The Four-Color Conjecture becomes:

In any plane graph each vertex can be assigned exactly one of four colors so that adjacent vertices have different colors.

Observe that in Figure 3 certain nonadjacent vertices of the graph can be joined by an edge without forming any intersecting edges. For example, the vertices representing Arizona and Colorado could be so joined. Join as many vertices as possible. The result, given in Figure 4, is an example of a *maximal plane graph*. Thus any plane graph is contained in some maximal plane graph. Also if a maximal plane graph can be four-colored, then any plane graph contained in it can be four-colored as well. Thus in proving the Four-Color Conjecture it suffices to prove that any maximal plane graph is four-colorable.

Before using the following result of Euler [7] we define the *degree of any vertex* w, denoted deg w, as the number of vertices adjacent with w.

Theorem. *Any maximal plane graph* G *with at least five vertices has a vertex* v *of degree* 3, 4, *or* 5.

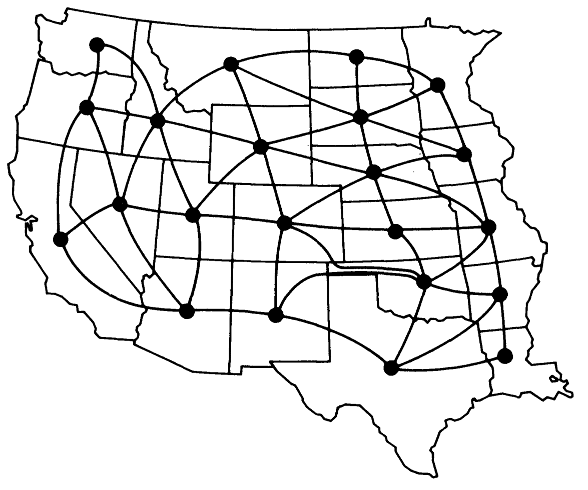

Figure 3.

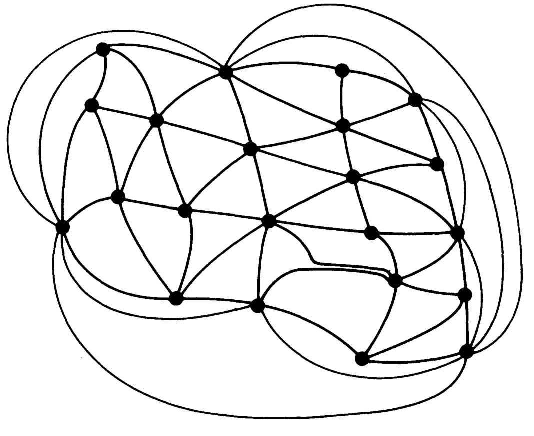

Figure 4.

We give now an outline of Kempe's work. Assume that the Four-Color Conjecture is false and let G be a maximal plane graph with the fewest vertices which is not four-colorable. Clearly G has at least five vertices. Let v be the vertex given in Euler's Theorem. Since G is a smallest non-four-colorable example, the removal of v and all of its edges from G results in a four-colorable graph. Four color it and consider the following three cases.

Case i. $\deg v = 3$. Now the vertices of G adjacent to v use only three of the four colors used on $G - v$. Thus the fourth color is available for v, a contradiction.

Case ii. $\deg v = 4$. If only three colors are used on the adjacencies of v, then as in case i we use the fourth color on v. Assume that the neighbors of v are clockwise v_1, v_2, v_3, v_4 and are colored respectively 1, 2, 3, 4.

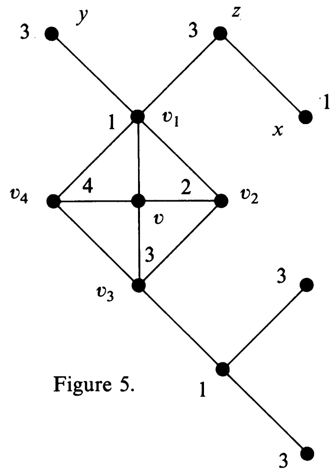

Figure 5.

Consider the points v_1 and v_3. If, as in Figure 5, there does not exist a chain of adjacent points colored 1 and 3 from v_1 to v_3, we interchange colors on *all* points colored 1 or 3 which can be reached by such a chain from v_1. In Figure 5, v_1 and x would then be colored 3 while y and z would now be colored 1. The result is indeed a coloring of $G - v$ such that v_1 and v_3 are colored 3. Vertex v can now be colored 1. This four-coloring of G is a contradiction.

If 1-3 chain from v_1 to v_3 as in Figure 6 exists, then there cannot exist a 2-4 chain from v_2 to v_4. For if such a chain did exist, then we must have two intersecting edges, which do not exist in a plane graph. Thus, similarly to the previous paragraph, we interchange colors on all points colored 2 or 4 which can be reached by a 2-4 chain from v_2. Now color 2 can be used on v, a contradiction.

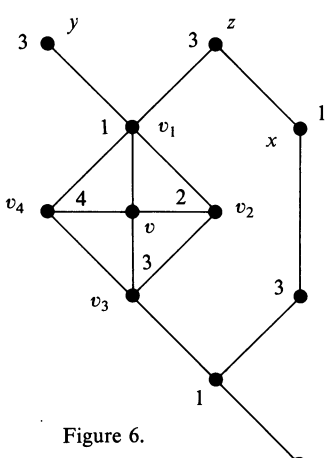

Figure 6.

Case iii. $\deg v = 5$. In this case let the vertices adjacent to v be clockwise v_1, v_2, v_3, v_4, v_5. Again we assume that four colors are used on the various v_i, for otherwise vertex v is easily colored with the fourth color. In fact by symmetry we can assume that v_1, v_2, v_3, v_4, v_5 are colored respectively 1, 2, 1, 3, 4 as in Figure 7. If there exists no 2-4 chain from v_2 to v_5, then as before we can interchange colors on all vertices which can be reached by a 2-4 chain from v_2. Then v can be colored with 2, a contradiction. Thus we assume that a 2-4 chain from v_2 to v_5 exists. Similarly we may assume a 2-3 chain from v_2 to v_5 exists. It now follows from the fact that we have a plane graph, that a 1-3 chain from v_1 to v_4 cannot exist nor can a 1-4 chain from v_3 to v_5.

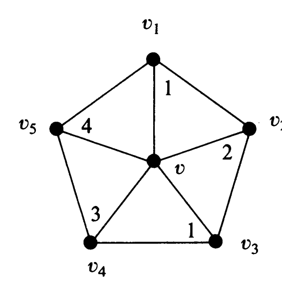

Figure 7.

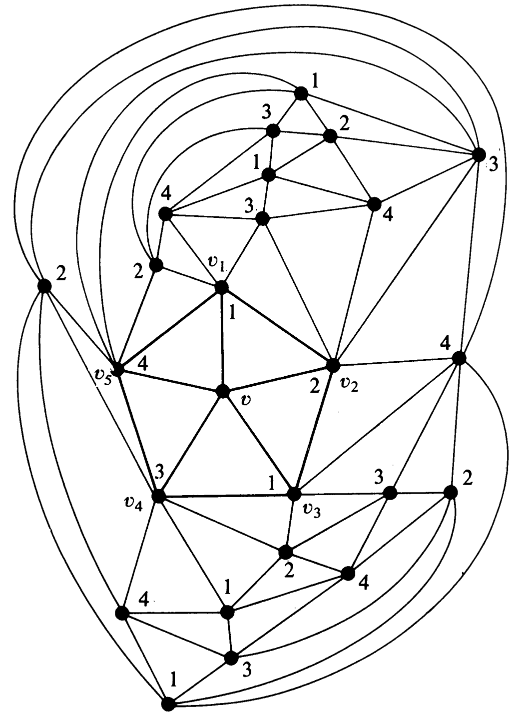

Figure 8.

At this point Kempe's argument runs into difficulty. According to Kempe we need only to interchange colors along the maximal 1-3 chain from v_1 and also interchange colors along the maximal 1-4 chain from v_3. Now no v_i, $i = 1, 2, 3, 4, 5$ is colored 1. Thus we use 1 on v, and according to Kempe the proof is complete.

However, Heawood showed by an example (see Figure 8) that it is possible that after interchanging colors on the 1-3 chain, the 1-4 chain connects v_3 with v_5. Thus interchanging colors on the 1-4 chain leaves v_5 colored 1, v_3 colored 4, and no color is available for v. A similar trouble occurs in Heawood's example if the first interchange occurs on the 1-4 chain. Thus Kempe's "proof" is not complete.

I hasten to add that the example which Heawood gave can easily be four-colored. It simply showed that Kempe's technique did not necessarily work.

Although Kempe's work was not complete, it contained most of the essential elements used almost a century later by Appel and Haken. We now introduce some of the terminology used in their work.

Let G be a smallest maximal plane graph which cannot be four-colored. A graph H is called *reducible* if H cannot be contained in G. As we discussed above Kempe showed that H_1 and H_2 are reducible, where H_1 consists of a vertex of degree 3, whose three neighbors are mutually adjacent and H_2 consists of vertices v, v_1, v_2, v_3, and v_4 as given in Figure 5. To complete the proof all that remains is to show that the graph in Figure 7, call it H_3, is reducible.

A set S of graphs is *unavoidable* if each maximal plane graph with at least 5 vertices must contain at least one element of S. According to Euler's theorem, we have that (H_1, H_2, H_3) is an unavoidable set.

A method of attack on the Four-Color Conjecture is to:

1. Find an unavoidable set S.
2. Show that each element of S is reducible.

A proof is complete when 1 and 2 are accomplished because: We assume that G is a smallest non-four-colorable maximal plane graph. By 1 we know G must contain at least one element of S. However, according to 2 all elements of S are reducible, i.e., G can contain no element of S. It follows that there is no smallest counterexample and hence no counterexample.

In 1913 G. D. Birkhoff [5] showed that H_4 is reducible where H_4 is exactly H_3 with the additional property that v_1, v_2, and v_3 have degree 5 and v_4, v_5 each have degree 6. Thus showing that $S_1 = (H_1, H_2, H_4)$ is an unavoidable set would complete a proof. However, it is known that S_1 is not unavoidable.

During the twentieth century various mathematicians, some famous, others not, worked on the conjecture. The story is often told of Minkowski's attempt at a proof. One variation of the story is that Minkowski boldly stated one day that the only reason the four-color problem was unsolved was that only third-rate mathematicians had worked on it. Minkowski, of course, knew that he was not a third-rate mathematician! Thus he set to work. After a few weeks Minkowski quietly went back to work on other problems apparently never to return to this one.

American mathematicians, in addition to G. D. Birkhoff, worked on the Four-Color Conjecture. In the 1940's a young American, A. Bernhart, wrote some papers on the conjecture. Shortly after he married, the new Mrs. Bernhart met the now middle-aged Mrs. Birkhoff at a mathematics meeting. Mrs. Birkhoff reportedly asked, "Did your husband make you color maps on your honeymoon, too?" Thirty years later the Bernhart's son, Dr. Frank Bernhart, worked on the problem. Haken and Appel continued this tradition of family involvement. Their teen-age children helped check some of the details of their work.

In 1922, Philip Franklin [8] at MIT, showed that every plane graph with at most 25 vertices can be four-colored. Four years later C. N. Reynolds [16] had increased that number to 27. By 1969, Ore and Stemple [15] increased it to 39. Shortly before Appel and Haken completed their work, Walter Stromquist [18], a graduate student at Harvard, showed that the Four-Color Conjecture holds for plane graphs with at most 52 vertices and J. Mayer [13] in Montpellier, France, increased that to 96. The essential idea in these works is that by limiting the number of vertices under consideration, a rather small unavoidable set could be obtained. Then each element in the unavoidable set was shown to be reducible.

In considering this technique for the possible complete proof of the Four-Color Conjecture, Appel and Haken knew from the work of H. Heesch of the University of Hanover that the required unavoidable set might have as many as 10,000 elements. Furthermore, such an avoidable set would contain some rather large configurations, each of which would have to be shown to be reducible. To show that one of these configurations, call it K, is reducible, one must in essence show that each of the numerous possible four-colorings of G-K can be extended to color K as well or be changed by a Kempe interchange of colors so that the result can be extended to K. Appel and Haken estimated that a single large configuration might take as much as 100 hours of computer time and require more storage than was available on any computer. However, by a method called a discharging procedure, Appel and Haken were able to find an unavoidable set with about 1500 members. Although many of the configurations in this set were not small, they were sufficiently lacking in size so that each was shown to be reducible in less than 30 minutes of computing time.

Of course, their work did not progress as smoothly and directly as outlined in the preceding paragraph. In fact by using a discharging procedure they obtained a set of unavoidable configurations. Then they would use their program to attempt to show all of the configurations were reducible. They would use the program until they encountered a configuration K for which reducibility was not verified after 30 minutes of computer time. They then would modify their discharging procedure to generate an unavoidable set which did not include K. By working back and forth in this manner, their proof was finally completed.

We now briefly discuss a discharging procedure. The exact procedure used by Appel and Haken is much too long and complicated to be given in detail here but we attempt to explain the idea. As previously we let G be a smallest maximal plane graph which cannot be four-colored. For each vertex v in G we assign it a number called its *charge*, 6-deg v. It follows from a theorem of Euler that the sum of all the charges of G is 12. Then rather involved rules (called the discharging procedure) are set up for moving positive charge from one vertex to another so that the decrease in charge at the first vertex is the amount of increase of charge at the second. Thus the new distribution still sums to twelve. The procedure eventually is used to show that if no element of a certain set S of configurations is contained in G, then each vertex eventually receives a nonpositive charge. This is impossible because the sum of charges is, of course, twelve. It then follows that S is an unavoidable set.

After this work of Appel and Haken a number of obvious questions come to mind. One such question is: What is the likelihood that a noncomputer-aided proof will be found? Of course, that is very difficult, perhaps impossible, to answer properly. However, Appel and Haken argue strongly that it is very unlikely that one could use their proof technique without the very important aid of a computer to show that a large number of large configurations are reducible. Of course, this does not rule out the possibility of some bright young person devising a completely new technique that would give a relatively short proof of the theorem.

Another question: What about the efficacy of other reported recent "proofs" of the Four-Color Theorem? A retired American mathematician named Thomas claims to have a noncomputer-aided proof or two of the theorem and an Englishman named Spencer-Brown received a fair bit of publicity in the U. K. two years ago with his "proof." As far as the author knows there exists no recognized expert in the field of graph theory and coloring problems who accepts the work of either Thomas or Spencer-Brown as a proof of the theorem.

Finally, what mathematical significance can be attached to Appel and Haken's work? This is probably the first proof of a major theorem which not only makes extensive use of the computer, but also would not be done without the computer. Some mathematicians believe that there may be many relatively easy statements which require tremendously long proofs. Currently the Four-Color Theorem seems to be an example of that. Furthermore this work of Appel and Haken will certainly herald a time when mathematicians attempt to make better use of the computer in obtaining new results. Of course the philosophical question occurs of what really constitutes a proof.

We close with a variety of references not previously mentioned for those who may want to read further. Many graph theory books are available for readers who may want to learn more of this recently popular branch of mathematics. Two of the better texts are Harary [10] and Behzad and Chartrand [3]. For the reader who wants detailed information on the Four-Color Problem, Ore [14] covers the topic very well up to about 1966, Saaty and Kainen [17] include a discussion of the work of Appel and Haken. Other discussions of Appel and Haken's work which are more technical than this article include Appel and Haken [2], Haken [9], and Bernhart [4].

REFERENCES

1. K. Appel, W. Haken, and J. Koch, Every planar map is four-colorable, Illinois J. Math, 21 (1977) 429–567.
2. K. Appel and W. Haken, The solution of the four-color map problem, Sci. Amer., 237 (Oct. 1977) 108–121.
3. M. Behzad and G. Chartrand, An Introduction to the Theory of Graphs, Allyn and Bacon, Boston, 1971.
4. F. Bernhart, A digest of the four-color theorem, J. Graph Theory, 1 (1977) 207–225.
5. G. D. Birkhoff, The reducibility of maps, Amer. J. Math., 35 (1913) 115–128.
6. N. Biggs, E. Lloyd, and R. Wilson, Graph Theory 1736–1936, Oxford, New York, 1976.
7. L. Euler, Demonstratio nonmullarum insignium proprietatum quibas solida hedris planis inclusa sunt praedita, Novi Comm. Acad. Sci. Imp. Petropol., 4 (1752–3) 140–160.
8. P. Franklin, The four-color problem, Amer. J. Math., 44 (1922) 225–236.
9. W. Haken, J. Graph Theory, 1 (1977), 193–206.
10. F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.
11. P. Heawood, Map colour theorems, Quart. J. Math. Oxford Ser., 24 (1890) 332–338.
12. K. May, The origin of the four-color conjecture, Isis, 56 (1965) 346–348.
13. J. Mayer, Inégalités Nouvelles dans le Problème des Quatre Couleurs, J. Combin. Theory Ser. B, 19 (1975) 119–149.
14. O. Ore, The Four Color Problem, Academic Press, New York, 1967.
15. O. Ore and G. Stemple, Numerical calculations on the four-color problem, J. Combin. Theory, 8 (1970) 65–68.
16. C. Reynolds, On the problem of coloring maps in four colors, Ann. of Math., 28 (1926–27) 477–492.
17. T. Saaty and P. Kainen, The Four-Color Problem, McGraw-Hill, New York, 1977.
18. W. Stromquist, The Four Color Theorem for small maps, J. Combin. Theory Ser. B, 19 (1975) 256–268.