Spring 5-3-2020

The Four Color Problem: The Journey to a Proof and the Results of the Study

Rebecca M. Rogers
Eastern Kentucky University, rebecca_rogers44@mymail.eku.edu

EASTERN KENTUCKY UNIVERSITY

The Four Color Problem:
The Journey to a Proof and the Results of the Study

Honors Thesis

Submitted

in Partial Fulfillment

of the

Requirements of HON 420

Spring 2020

By

Rebecca M. Rogers

Mentor

Dr. Mathew Cropper

Department of Mathematics and Statistics

Abstract

The Four Color Problem:

The Journey to a Proof and the Results of the Study

Rebecca Rogers

Dr. Mathew Cropper Department of Mathematics and Statistics

The four color problem was one of the most difficult to prove problems for 150 years. It took several failed proofs and advancement in technology and techniques for the final proof to become possible. Some notable men include De Morgan first writing about the problem, Kempe giving the first proof, Heawood showing the flaws in Kempe's work as well as making advancements of his own. The first actual proof of the problem is then discussed, as well as it's shortcomings and the work done by other mathematicians to show improvements on them. The total of this work has lead to numerous great leaps in mathematics including the creation of the branch known as graph theory. This one problem also revolutionized proof writing, being the first to use a computer as an essential part of the proving process.

Keywords: Four Color Problem, Four Color Theorem, Graph Theory, Computer, Kempe, Appel and Haken

Contents

List of Figures

Acknowledgements

Thank you to Dr. Mathew Cropper for your time and effort in mentoring this project.

Thank you to the faculty of the Honors Department for helping along the way.

Lastly, thank you to my family for supporting me and being patient while I work on this.

Introduction

Most people who have taken a math class throughout the years, at almost any level has probably asked themselves how this certain problem, or even area of math was even created. Branches of mathematics are far reaching and often seem very daunting. Learning thought that an entire branch had stemmed out of one question may make it seem easier. The subject seems easier still when you learn that the question involves coloring a map. It is not until it is learned that this one simple question, which can be stated in one short sentence, took over a hundred years to find a proof for and a new branch of mathematics to solve.

History of the Problem

The first thing to be asked is what is this problem that keeps getting mentioned? The first time it can be found to be written is on October 23, 1852 in a letter from Augustus de Morgan to Sir William Rowan Hamilton [**11**]. The contents of the letter are as follows:

> A student of mine asked me to day to give him a reason for a fact which I did not know was a fact, and do not yet. He says, that if a figure be any how divided and the compartments differently coloured so that figures with any portion of common boundary line are differently coloured four colours may be wanted but not more. The following is his care in which four are wanted.
>
> A B C D are names of colours
>
> Query cannot a necessity for five or more be invented. As far as I see at this moment, if four ultimate compartments have each boundary line

> in common with one of the others, three of them inclose the fourth, and prevent any fifth from connexion with it. If this be true, four colours will coulour any possible map without any necessity for colour meeting colour except at a point. Now, it does seem that drawing three compartments with common boundary A B C two and two you cannot make a fourth take boundary from all, except inclosing one But it is tricky work and I am not sure of all convolutionns What do you say? And has it, if true been noticed? My pupil says he guessed it in colouring a map of England. The more I think of it, the more evident it seems. If you retort with some very simple case which makes me out a stupid animal, I think I must do as the Sphynx did. If this rule be true the following proposition of logic follows: If A B C D be four names of which any two might be confounded by breaking down some wall of definition, then some one of the names must be a species of some name which includes external to the other three[**11**].

Due to the nature of the problem it was named the Four Color Problem. Later it becomes known that De Morgan's student who mentioned this was Frederick Gutherie, but it was his brother Francis Gutherie who initially made the claim. This is why the four color problem is also commonly known as Gutherie's Problem. Simply put, the four color problem states that for any map only four colors are needed such that no areas which share a common boundary (more than a point) have the same color. This idea is a very simple one, so simple that even elementary school children can understand the idea behind it. What makes this truly a problem though is that no matter how simple to phrase, it is extremely difficult to prove.

Although the problem has now been mentioned and documented, it does not gain notoriety until after it is written of in the *Proceedings of the London Mathematical Society* by Cayley in 1878, asking if it had yet been proved [**22**]. When it had been found to not as of yet been proven, it was not long until there were many attempts being made toward a proof. In mathematics it is the thought that any statement which can be simply phrased can also be proven in a suscinct and efficient manner such that anyone with the necessary knowledge can easily follow the proof to see that it is true [**3**].

The first attempt at a proof came in the next year by Kempe. It was not until eleven years later, in 1890, that this proof was shown to be incorrect by Heawood. His findings lead to the proof of the five color theorem. Due to this finding, it was clear that the minimum number of colors needed to satisfy these conditions were at most five. From here, all that needed to be shown was if there were any cases in which five were necessary or if four colors were sufficient. Even though Kempe's proof was shown to be overall incorrect, it had many correct elements and the general idea used within the proof of reduction were eventual used in the first correct proof that was found.

During this time there were strides toward a proof and related graph ideas by many people all working toward the much desired proof of this one problem. Some of these honorable mentions include:

- 1880 Tait: found three-coloring the edges of a cubic map
- 1891 Peterson: Four Color Problem was equivalent to a problem on edge coloring
- 1898 Heawood: The Four Color Problem into algebraic form
- 1912 Veblen: The Four Color Problem to assertions in projective geometry
- 1912 Birkhoff: Chromatic Polynomials
- 1931 Whitney: Dual Graphs
- 1941 Brooks: Theorem that gives a bound on the chromatic number of any graph
- 1943 Hadwiger: A conjecture in which the four color problem is a special case

[**22**]Within these many years there were other strides toward finding the proof, such as many people showing that maps with a certain number of regions being able to be four colored. The first of these instances was in 1922 by Franklin. He showed that any graph that has 25 or fewer regions needed at most four colors. In 1926 this number was raised to 27 regions by Reynolds. Franklin raised the number to 31 in a paper published in 1938. It was raised two more times, first by Winn to 35 in 1940 and finally by Ore and Stemple to 39 in 1970[**22**]. Although this method was a good excercise and lead to many findings, this was not a way to prove the conjecture, as there is no stipulation to how many regions are possible.

It is not until 1977 that there is finally a proof answering the question. This proof is different than any other that came before it though. This proof utilized new technology in the form of a computer. Due to this new technology, the proof was not readily accepted. In fact, a book on the topic, entitled *The Four Color Problem: Assault and Conquest* written by Thomas Saaty and Paul Kainen, which was published just after the proof came out has many sentences that tell the reader to be cautious of putting too much stock into the proof. One such example reads, "Since the proof itself (if it is a proof) was discovered using the theory, and since the theory is heuristic, there is an added tendancy on the part of many mathematicians to mistrust the whole thing [**22**] ." This is just one example of the doubt surrounding the Appel-Henkel proof. Even though this was a controversial, it took twenty years before another result, confirming Appel-Henkel's was found. This is because the computing technology for such large quantities of data were still being improved upon.

Now, the Four Color Problem is acknowledged as being proven in 1977 by Appel and Henkel and the use of computer technology is becoming more accepted in mathematics. Although the problem has been solved does not mean that it's legacy is over. This problem that was started as an "innocent little coloring problem [**12**] " has expanded mathematics and banded together mathematicians for over 150 years.

The entire branch of math known as graph theory began to be explored in order to prove a question easy enough that even those new to math can understand it.

Background Math

Although the Four Color Problem is easy to understand, in the way that it has thus far been worded may lead some to wonder how it can be proven using mathematics. To do such, it is necessary to phrase the problem in mathematical terms. For reference, the original problem can be worded as, "can the regions of every map on a globe be colored with four colors so that regions sharing a nontrivial boundary have different colors [**29**]." Here terms such as map, region, and boundary are used. Though these terms are easy to understand, they are not mathematically the best way to look at the problem. In 1931 Whitney did work on dual graphs. This is the idea that each region can be represented by a vertice (or dot) and where two regions are connected by a boundary the corresponding vertex can be connected with a line known as an edge [**30**]. This new image is a graph, while the original is the map (see Figure 1). This new graph is known as the map's dual graph. One of the important features of a dual graph is that it allows a region coloration corresponds directly to a vertex coloration of its dual graph [**19**]. Using graphs instead of the corresponding maps is useful as the shape of the regions is no longer of concern. Now, maps that appear to be different, due to the shapes of the regions, can actually be seen to be the same through the lens of this particular problem (see Figure 2).

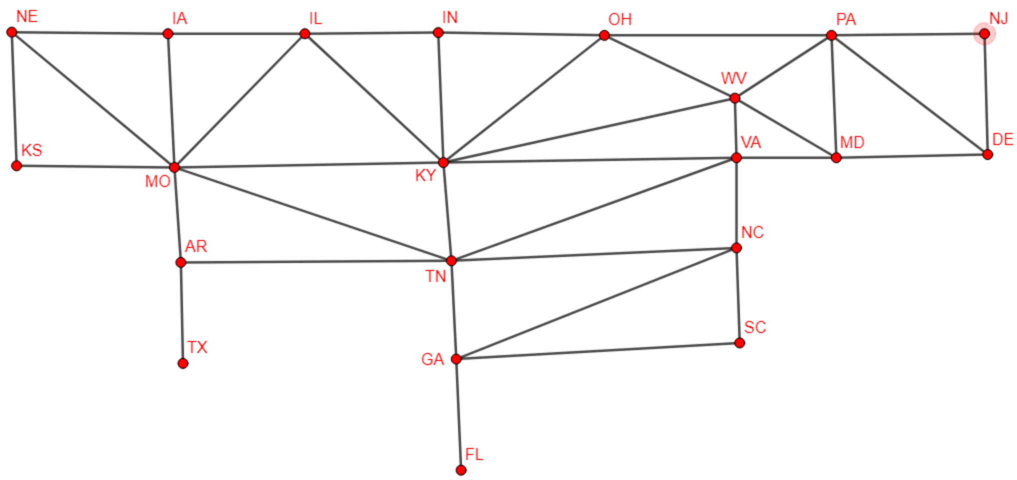
There are a few aspects of the dual graphs we will be using that are important to note. The first of these is that due to the duality with maps, all maps that we need to consider are planar maps. In the simplest terms this means that the graph can be drawn on a two dimensional surface or the surface of a sphere in such a way that the edges do not intersect when there is not a vertice [**16**]. Figure 3 shows an example of a non-planar graph. There are two parts of graphs that are not necessary to consider for this problem. The first is a loop. A loop is when one edge both starts

Figure 1. This shows how a map can be correlated with a dual graph. The map shows the continential Unites States, although only a few of the states were chosen to be included in the graph. Note that all of these states are somehow connected to one another. These chosen states also do not create any of the obstacles described by Appel and Haken as is described in the section "Appel and Haken."

and ends on the same vertice. If this were to occur in a dual graph it would mean that the map has a region that shares a boundary with itself. For maps there would not be a boundary there, so a loop does not make sense. The other are parallel edges. This is when two vertices have two seperate edges connecting the two vertices. In terms of maps this would be two regions sharing two boundary edges. Although this is possible, it does not change the coloring no matter how many boundary lines there are, so the graph can be simplified to only one edge. For an example of a loop or parallel edges see Figure 4.

Another definition that will prove useful to know is the degree of a vertex. Since there will be no loops in the graphs following, it can be said that it is the number of edges leading to (or from, based on perspective) a vertex. For maps this is equivalent to saying how many other regions share a boundary edge with the region in question. For example, the state of Kentucky has seven states that share a boundary. In the dual graph for the United States, the vertex representing Kentucky would be of degree seven as there are seven edges extending from that vetex (see Figure 1).

From here we can begin to look into the work leading up to the proof found by Appel and Henkel. This includes the attempted proof by Kempe and the disproof by Heawood. Although overall the proof by Kempe was found not true, most of the work done was correct and essential to understanding the proof that was eventually found. Heawood, during his work showing Kempe's to be incorrect, came up with a proof showing that five is the maximum minimal number needed for planar coloring problems.



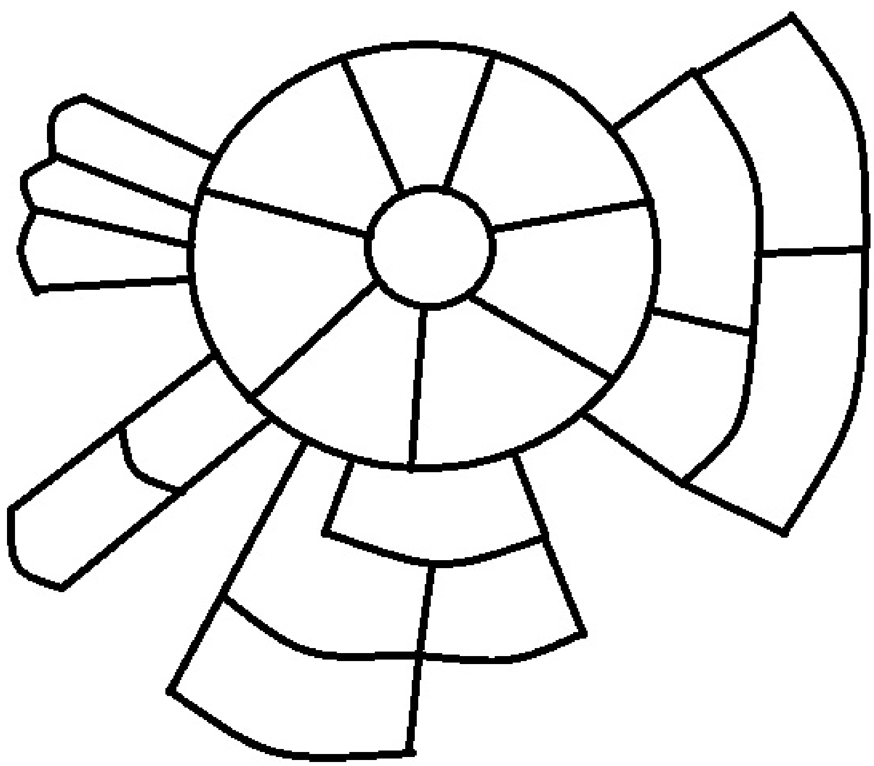


FIGURE 2. The top image shows the dual graph that corresponds to Figure 1. This graph has been laid out in a different manner, and although may look different is the same graph. Things such as length of edges and the angle between edges does not matter in graph theory as it does in subjects such as geometry. It is for this reason that maps which look totally different can have the same graph. The bottom image is a map which has the same dual graph as that shown in the top image, yet it looks very different from the map of the United States.

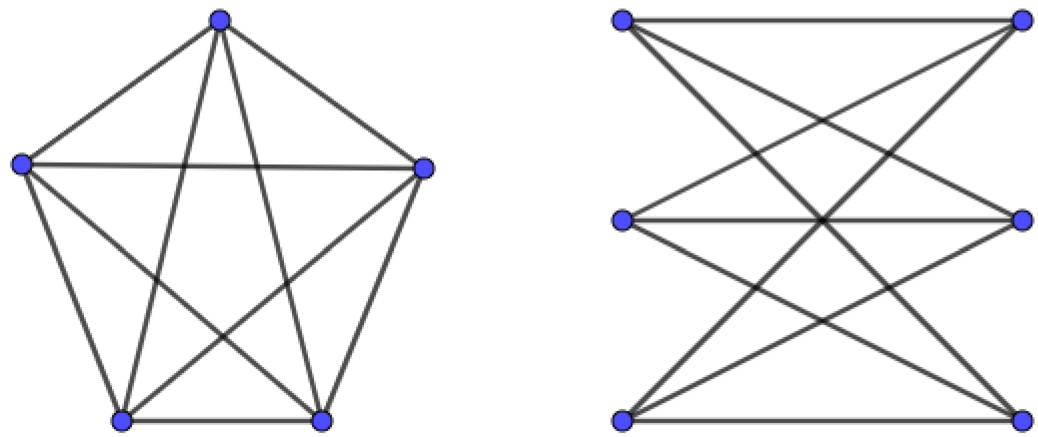


FIGURE 3. This figure shows two different examples of non-planar graphs. The first image is that of K_5 the complete graph with 5 vertices. The second is an example of a bipartite graph. Both of these images have edges that cross where there is no vertex. This is one of the main signs that a graph is not planar.

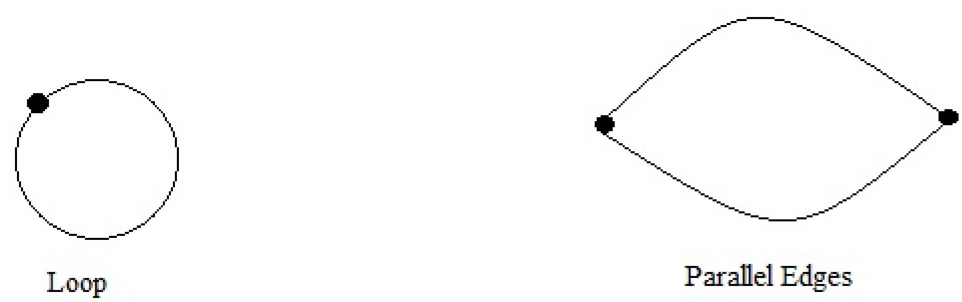


FIGURE 4. The first image shows a loop. Note that it is one vertex and one edge. The second image shows parallel edges. Here there are two vertices and two edges. Both of these are situations that will not have to be looked at or thought of as special cases when dealing with dual graphs of maps.

Attempt of a Proof

After the problem had been initially stated in 1852 it took over twenty years for the problem to pick up noteriety. Once it did, however, it was not long before there were several mathematicians working toward the answer. A proof was quickly found, but over a decade later one map was shown to contradict part of the proof, leading it to be a failed attempt. Although the proof was not successful, it's revolutionary methods are still applied. For nearly a century the proof loomed just out of reach of mathematicians.

Kempe's Proof

In 1879, only 1 year after the question was posed by Cayley, Alfred Bray Kempe had an article entitled "On the Geographical Problem of the Four Colors" in the American journal of Mathematics declaring a proof to the seemingly easy problem. After explaining the necessity and usefulness of an answer to the problem, Kempe begins explaining his process to a proof. The first thing noted, which is significant to the proof, is "that four colours may be necessary will be at once obvious on consideration of the case of one distict surrounded by three others (see Figure 5), but that four colours will suffice in all cases is a fact which is by no means obvious [**15**]." Here, Kempe correctly points out that it is clear with one simple example that there are cases in which four colors are needed. Now, what needs to be shown is that four colors are always enough. An equivalent statement to this is that there are no instances in which five colors are needed. The rest of the article attempts to show that these statements are indeed true when applied to planar surfaces.

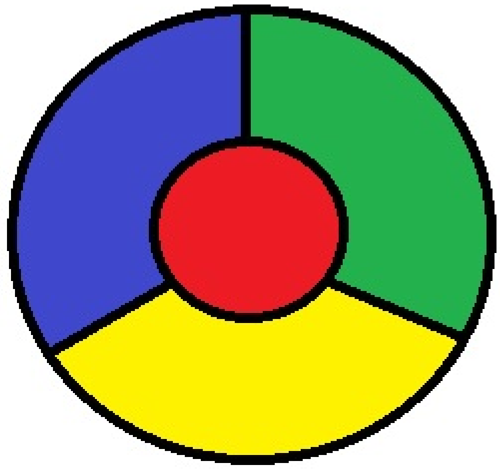


FIGURE 5. A simple example of how four colors are necessary. Note that the red section touches the blue, green and yellow. Similiarly, the blue region touches each of the other three, as well as the yellow touching the other three and the green touching the other three. Thus, none of the colors can be switched for another.

Unlike many mathematical papers that are written today, Kempe's "paper is virtually all prose which, though well written, makes it difficult to verify his work [**23**]." Rather than written as a clear mathematical proof, the paper holds a structure that can at times be rather difficult to follow. Timothy Sipka rewrote Kempe's proof in a different structure which makes it clear that it is actually a proof by mathematical induction which covers several cases [**23**]. Proofs by induction all hold a common principle. First, the first (or first several) statements need to be shown as true. Then, assume that there is a step farther along that is true. Finally, show that this assumption leads to the next step also being true. This method concludes that all steps are then true [**5**]. The general structure of induction is used loosely by Kempe, but explained plainly using Kempe's terminology through Sipka. The base case, as this first step is often called, is clear - when there is a map of four or fewer areas, then four or fewer colors suffice to color them in an acceptable manner. Now, assume that a map with n areas, or regions, can be colored appropriately with only four colors.

Kempe then shows that a map with $n + 1$ regions must have at least one region with five or fewer boundary lines with other regions. This means that there is always at least one region that is adjacent to five or fewer other regions. This result is found due to a method involving Cauchy's formula.

In order to get the map from having $n + 1$ regions to only n regions, Kempe came up with a process which he called patching. For this he says to literally cut out a patch the same shape as the region you wish to get rid of just slightly bigger. Then he says to cover this region up and extend the now unended boundary lines to meet at a vertex. If the region being patched over touches only two other regions, then rather bring the boundaries to a point connect them with a boundary line (see Figure 6). Patching in this manner will eliminate what Kempe refers to as islands and peninsulas. Islands are a region or a group of regions that do not connect with the rest of the map. Peninsulas are a region or group of regions that connect to the rest of the map through only a single vertex (see Figure 7). This process of patching being repeated several times will always eventually result in only one region remaining.

Before going any farther, it is necessary to look into why only regions with four or five suurrounding regions need to be gone through this patching process. In order to show this result Kempe works through some formulas, eventually leading to an equation credited to Cauchy, although it was derived from wok done by Euler. To begin it is important to define some variables. Let R be the number of regions at any given step with the patch still on. B is the number of boundaries and P the number of points of concourse -the number of times two boundary lines meet- at the same step with the patch on. Then, R' is the number of regions after the patch is moved at that particular step, B' the number of boundaries and P' the number of times boundary lines meet after the patch is removed. Now, let us look at the situation when the patch is covering an island. Then,

$$P' = P$$

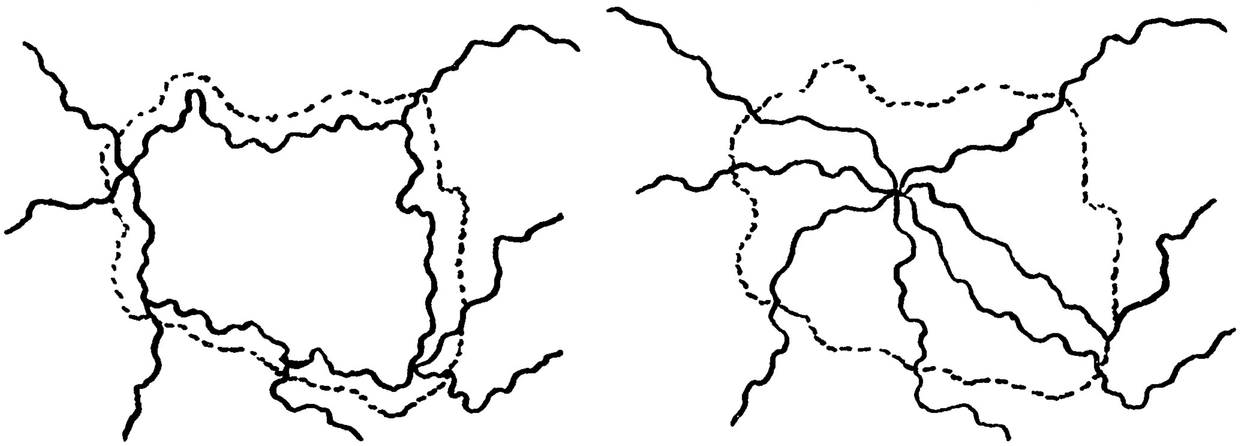


FIGURE 6. Here is the patching process as shown by Kempe in his original paper. He had these labeled as two seperate images, the first showing where the patch was going and the second showing the result after placing the patch and connecting all the boundary lines to a single point. Note that all the outer boundary lines do not show an ending, meaning that this is just a part of a bigger map.[**15**]

$$R' = R + 1$$

$$B' = B + 1.$$

Using these equations it can be seen that

$$P' + R' - B' - 1 = P + (R + 1) - (B + 1) - 1 \tag{1}$$

$$= P + R - B - 1 \tag{2}$$

For situations when the patch is over a peninsula region, the equations become,

$$P' = P + 1$$

$$R' = R + 1$$

$$B' = B + 2$$

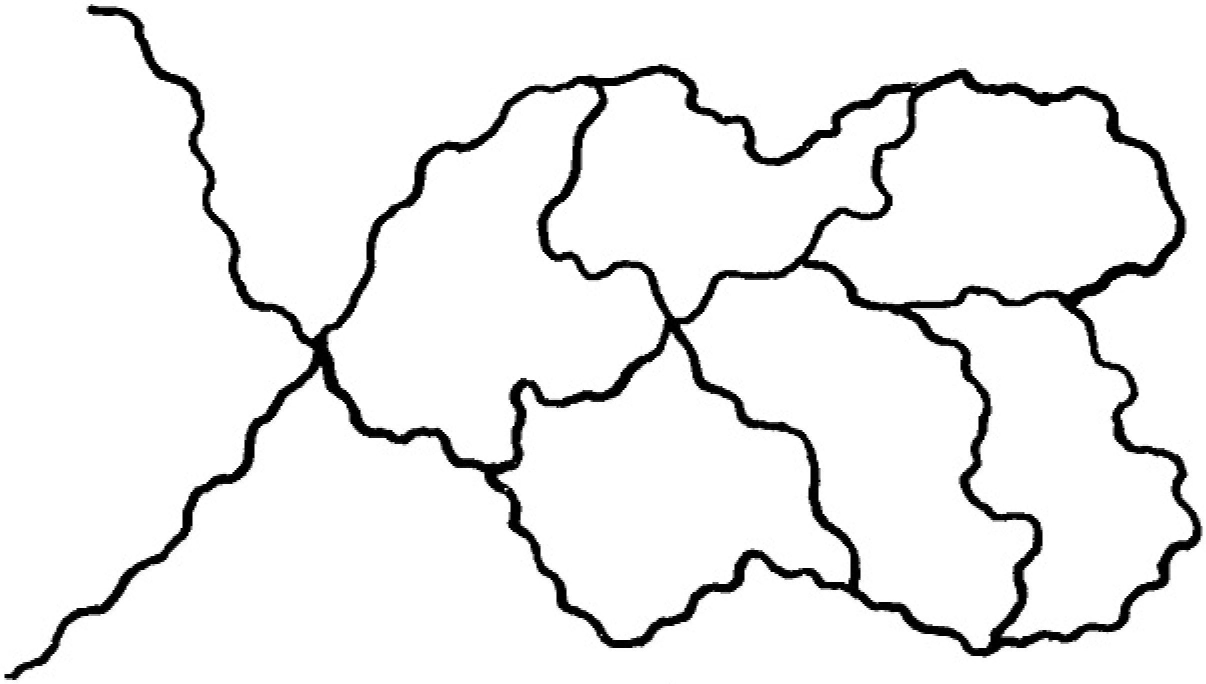


FIGURE 7. This is the example given by Kempe for a peninsula. This section looks very similiar to what is seen in Figure 6, yet there is a difference. This can be seen as a peninsula due to the outer boundary lines, on the peninusla area, not disappearing without an end. These boundary line can be seen to make a clear complete outer boundary, showing that it is indeed the edge of the map. [**15**]

These result in,

$$P' + R' - B' - 1 = (P+1) + (R+1) - (B+2) - 1 \tag{3}$$

$$= P + R - B - 1 \tag{4}$$

The final situation is when the patch is covering any other region, not including an island or peninsula region. The equations for this are,

$$P' = P + x - 1$$

$$R' = R + 1$$

$$B' = B + x$$

where x is the number of regions that are adjacent or connected to the one under the patch. Using these we get that,

$$P' + R' - B' - 1 = (P + x - 1) + (R + 1) - (B + x) - 1 \quad (5)$$

$$= P + R - B - 1 \quad (6)$$

From (2),(4), and (6), it can be seen that

$$P' + R' - B' - 1 = P + D - B - 1$$

in all scenarios. This is saying that at every step $P + R - B - 1$ has the same value, even after the patch is removed. Yet, it is known that when all patches are added , there is only one region left. Thus,

$$P = 0$$

$$R = 1$$

$$B = 0.$$

From here it can be seen,

$$P + R - B - 1 = 0 + 1 - 0 - 1 \quad (7)$$

$$= 0 \quad (8)$$

for every step. This is the part that can be attributed to Cauchy and Euler. Kempe takes this work a bit farther. For the rest of this section, let subscripts refer to the number of boundaries associated with each bit of information. So, r_1 refers to the number of regions with one boundary, while r_2 refers to the number of regions with two boundaries. Similiarly, p_3 refers to the number of points of concourse where three boundaries meet. In general, r_n is the number of regions that have n boundaries and p_n is the number of points of concourse where n boundaries meet. These lead to the equations

$$R = r_1 + r_2 + r_3 + ...$$

$$P = p_3 + p_4 + ...$$

Since two districts share one boundary,

$$2B = r_1 + 2r_2 + 3r_3 + ...$$

A similiar equation can be made regarding points of concourse, yet a couple of situations must be taken into account first. To begin, when there are no boundaries involved, for these situations b_0 will be used. When it is a peninsula and a boundary is made and not a point of concourse, it will be refered to as b_1. Now, it can be said,

$$2B = 2b_0 + b_1 + 3p_3 + 4p_4 + ...$$

Using multiplying (7) and (8) by 6 and rearranging some terms results in,

$$(6R - 2B) + (6P - 4B) - 6 = 0.$$

Substituting in what is known,

(9) $$(6R - 2B) + (6P - 4B) - 6 = 6(r_1 + r_2 + r_3 + ...) - 6(r_1 + 2r_2 + 3r_3 + ...)$$

(10)

$$+(6(p_3 + p_4 + ...) - 2(2b_0 + b_1 + 3p_3 + 4p_4 + ...)) - 6$$

(11) $$= 5r_1 + 4r_2 + 3r_3 + 2r_4 + r_5 - ...$$

(12) $$= 0$$

Where all terms not listed in _____ are subtracted from the equation. In order for this equation to equal zero, which it must, then there must be at least one of the following: r_1, r_2, r_3, r_4, r_5. So, there must be at least one region that has less than six boundaries in every map. From here Kempe goes on to say,"Consequently, if we develop a map so patched out (with patches only going over districts with less tham six boundaries), when taken off, discloses a [region] with less than six boundaries, not more than five boundaries meet at the point of concourse on the patch[**15**]." This is how it was shown that at most five regions will be surrounding the patched off region.

Now, the map with $n+1$ regions, which must have a region, X, with five or fewer adjacent regions, has this patching procedure done on region$X$. This results in the map with n regions, which we have already assumed to be four colorable. All that Kempe had left to show was that in all scenarios, where five or fewer regions were adjacent to this centralized region which just had its patch removed, could be four colored. This would mean that the up to five regions in a ring like pattern around the center could be colored with only three colors, leaving the fourth for the uncovered middle region. As Sipka describes it, this was shown by Kempe in different cases. Since it is clear that if the ring had at most three regions, then it could be colored with three or fewer colors. With this in mind, the first case is that which has 4 regions in this ring. This case is then broken into subcases based on a concept known as Kempe chains. To describe a Kempe chain, "He first asked that we consider all the [regions] (he called them districts) in the map which are colored red and green; then he observed that these [regions] form one or more red-green [sections]. Kempe's notion of a red-green [section] was simply a continuous 'chain' of [regions] colored red or green. He then made the important observation that one could interchange the colors in any red-green [section], and the map would still remain properly colored [**23**]."

The first subcase of case one describes a map that has four regions in the ring surrounding region X, where region A and region C belong to different red-green chains (see Figure 8). Since the two regions are part of different chains, then the colors of one chain, say the one containing region A, can be inverted - all red regions in this chain become green and all green become red, so that now A and C are the same color. This means the four regions in the ring are colored with three colors leaving the fourth color for region X.

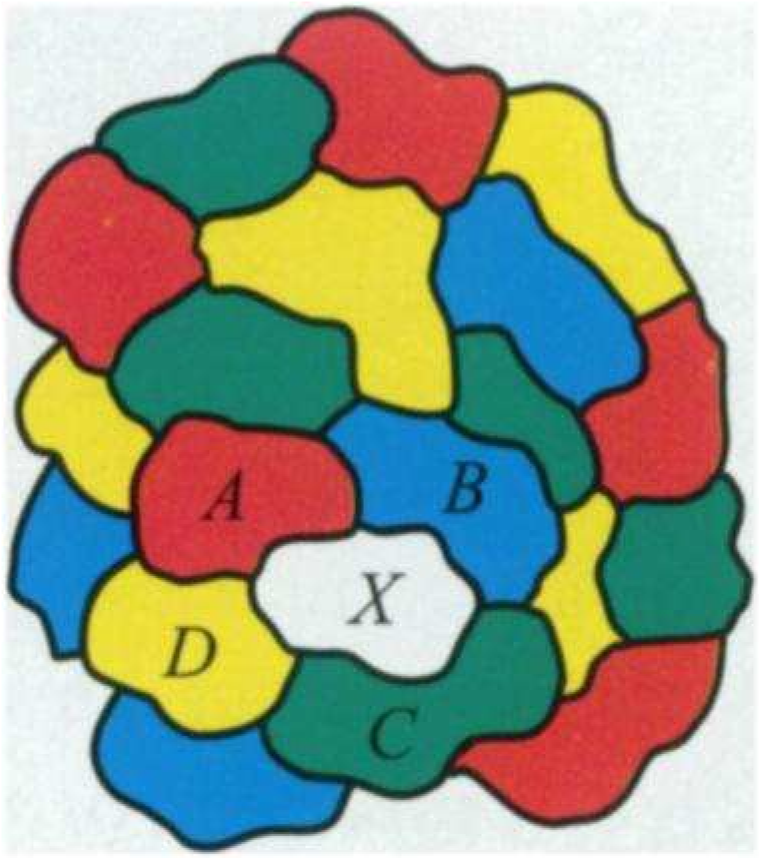


FIGURE 8. This image is provided by Sipka as the scenario in which Kempe uses as his first situation. In the modern terms that Sipka uses this is subcase 1.1. [**23**]

The other subcase of case one is where regions A and C are part of the same red-green chain. In order for this to happen, then regions B and D must not be a part of the same blue-yellow chain (see Figure 9). Now the same principle from subcase one can be applied to this case of the blue-yellow chains. Invert the colors of one of the chains, then there will only be three colors used around the region X leaving the fourth color for X.

The next case to look at, case two, is that in which there are five regions in a ring around region X. Like case one, case two also has two subcases. The first of these subcases is when regions A and C are part of different red-yellow chains or if A and

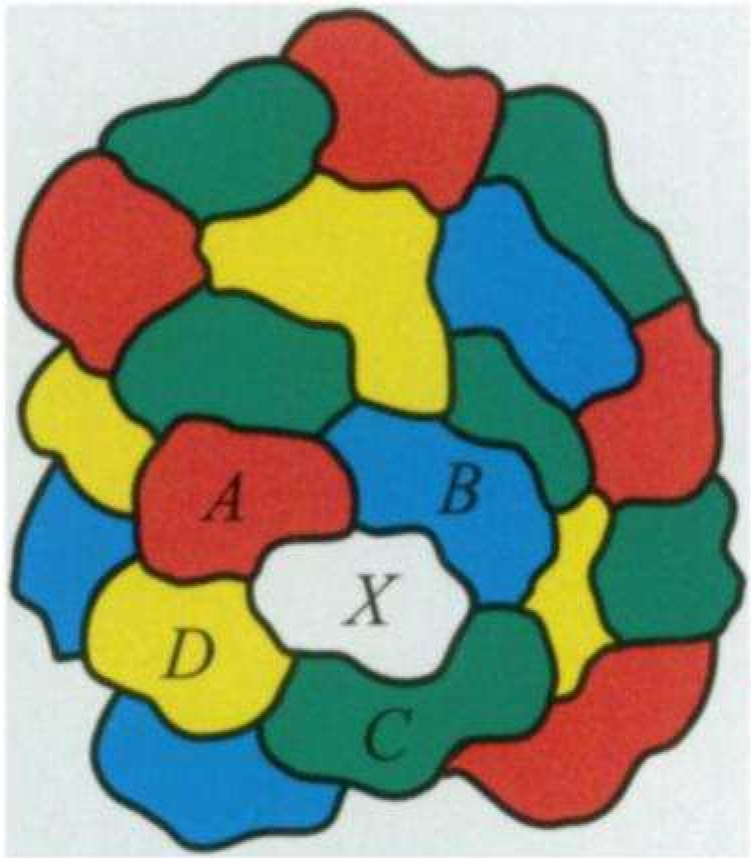


FIGURE 9. This is the second scenario discussed by Kempe, or in the terms of Sipka subcase 1.2. [**23**]

D are part of different red-green chains (see Figure 10). When one of these is true then those colors can be inverted leaving three colors in the ring surrounding region X. As with before the fourth color can then be used on region X.

The final subcase to look at is the second subcase of case two. For this case regions A and C are a part of the same red-yellow chain and regions A and D are a part of the same red-green chain (see Figure 11). Kempe then goes to state, "the two regions cut off B from E so that the blue-green region to which B belongs is different from that to which D and E belong, and the blue-yellow region to which E belongs is different from that to which B and C belong[**15**]." He continues to explain that the colors in the blue-green chain in which contain B need to be inverted as well as the colors in the blue-yellow chain that contains D. These two switches lead to the

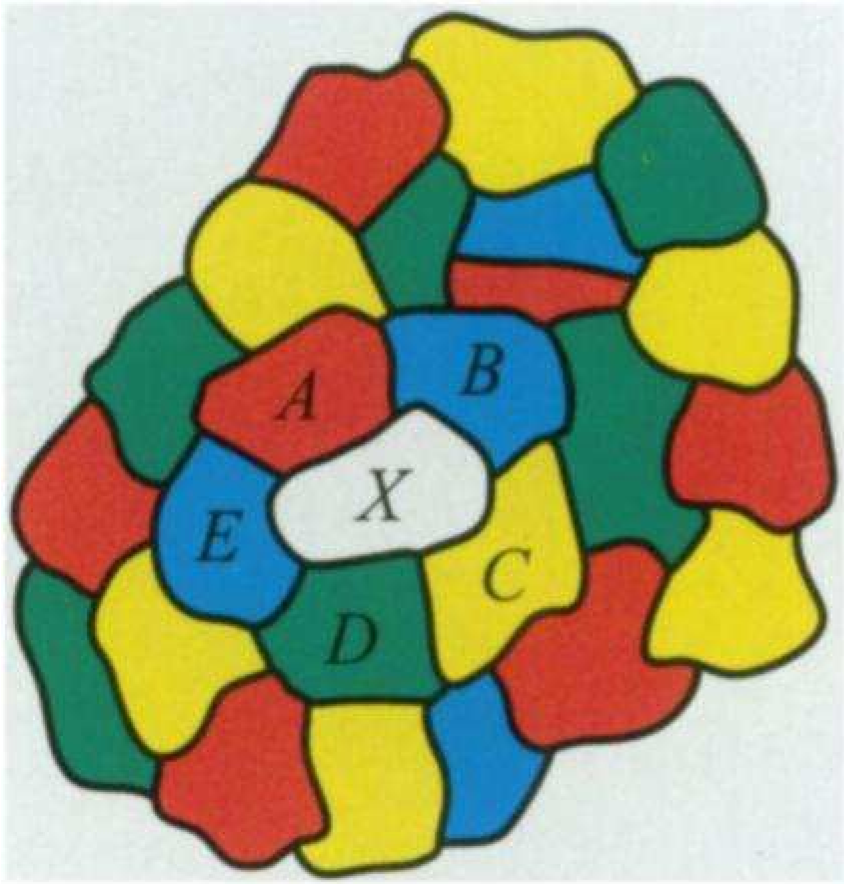


FIGURE 10. The third possibility talked about by Kempe, or Sipka's subcase 2.1. This is the first scenario that uses five regions surrounding X rather than four. [**23**]

regions in a ring around X only using three colors, leaving the fourth color for region X.

This concludes the proof given by Kempe for the four color problem. The methodology of Kempe chains is one that has endured throughout time, even after a flaw was found in Kempe's work. The flaw was not in the chains and this procedure has lasting impacts on mathematics. Although the proof was widely accepted, in 1890, eleven years after Kempe's paper, one map would be enough to show that it was not quite the answer needed to solve the problem.

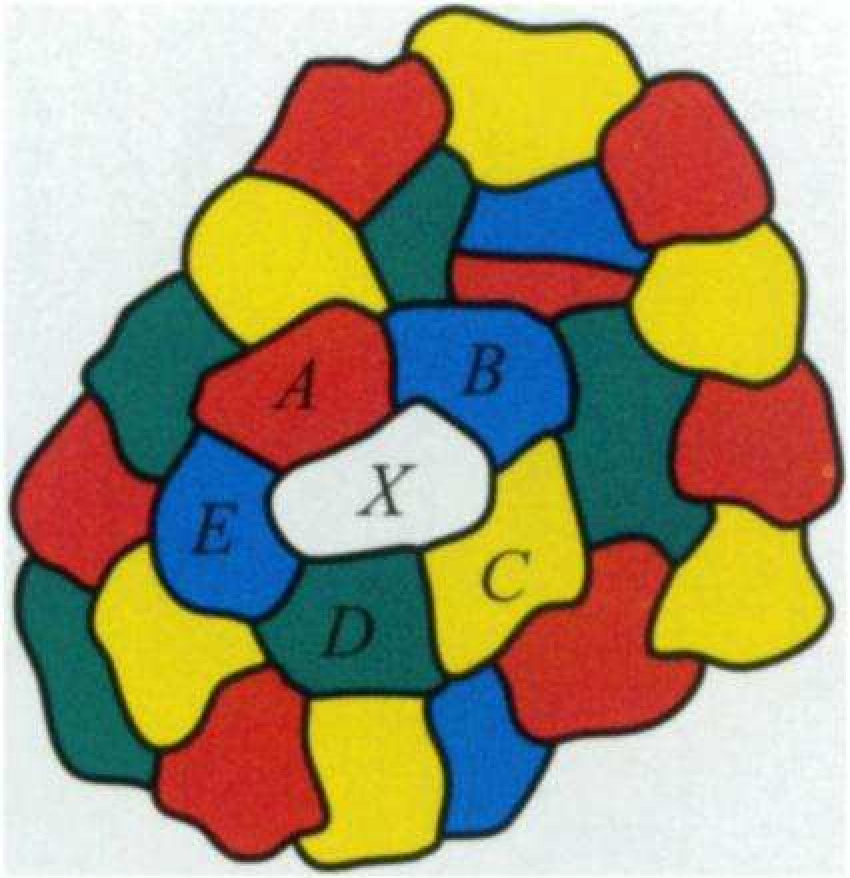


FIGURE 11. Kempe includes this as the final possible scenario of coloring a region getting a patch removed. This is called subcase 2.2 by Sipka.[**23**]

The Five Color Theorem

In 1890, Percy J. Heawood wrote a paper that included a discussion on the four color problem. Allthough the majority of this paper focused on the number of colors needed on surfaces other than those that are planar - for which Heawood did a great deal of work, the ending mentions the minimal colorings needed on any planar map. Heawood takes his time to describe Kempe's method, specifically what we refered to as the second subcase of case two. Heawood was complimentary of Kempe's work until one sentence, "it is conceivable that though either transposition would remove a red, both may not remove both reds[**13**]." This one sentence and the corresponding map, Figure 12, were enough to show that Kempe's proof was flawed and not enough to give a complete answer to the problem.

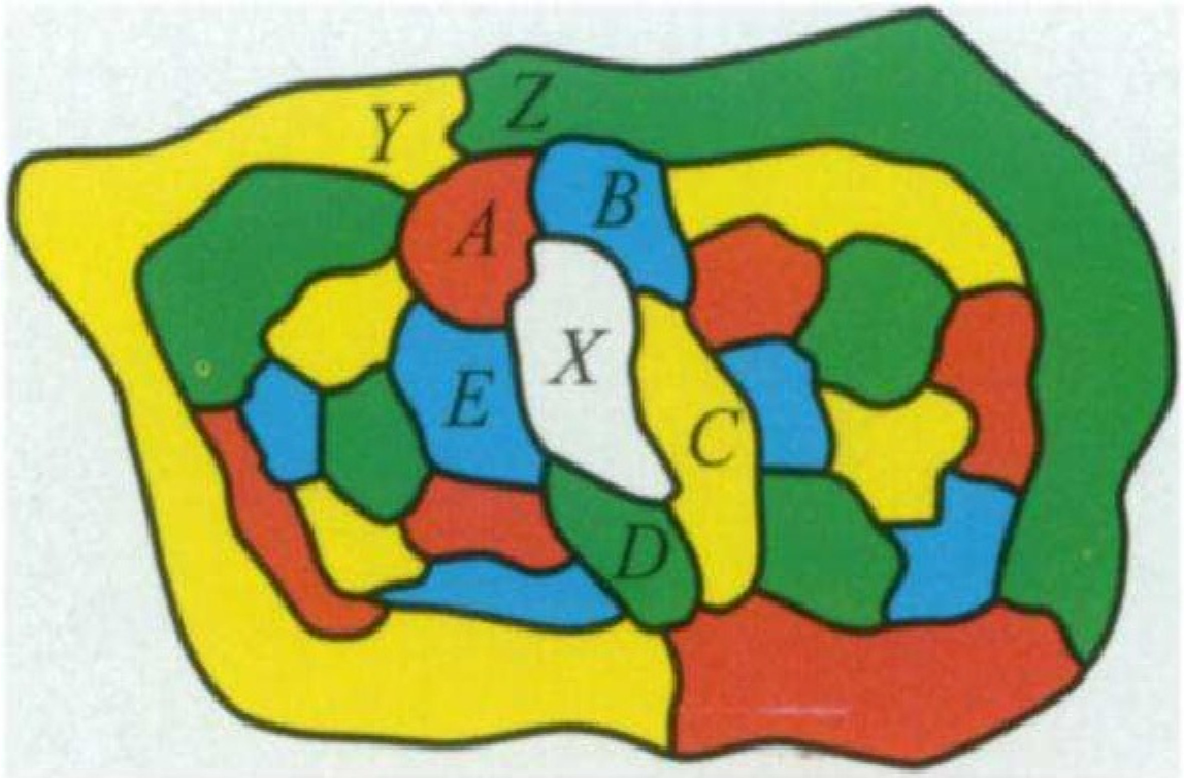


FIGURE 12. This is the image used by Heawood to show that Kempe's proof was not adequate for every possible map. This image is in contrast only to the fourth scenario Kempe mentions, leaving the other three as being correct. This is where both reds may not be removed.[**23**]

Although Heawood showed this flaw in the proof, he did not attempt to come up with a correct proof. Rather, he modified Kempe's work to show that it did prove another theorem, the five color theorem. This theorem, much as its name suggests, says that every planar graph can be five-colored. Heawood proved this theorem by slightly altering the work Kempe did to prove the four color theorem [**22**]. Using the ideas of Kempe, and knowing that his first case and the first subcase of case two are correct, it can be seen in the last subcase that at least one pair of regions must touch each other nowhere[**13**]. As with many theorems, proofs of different types have been tried through out the years. For the five color theorem, another proof that was used was by Paul C. Kainen. His proof was a proof by contradiction and using the idea that K_6 is not planar[**14**]. K_6 is the complete graph having 6 vertices (see Figure

13). This proof is mentioned here as the use of complete graphs play a part in future attempts at a proof.

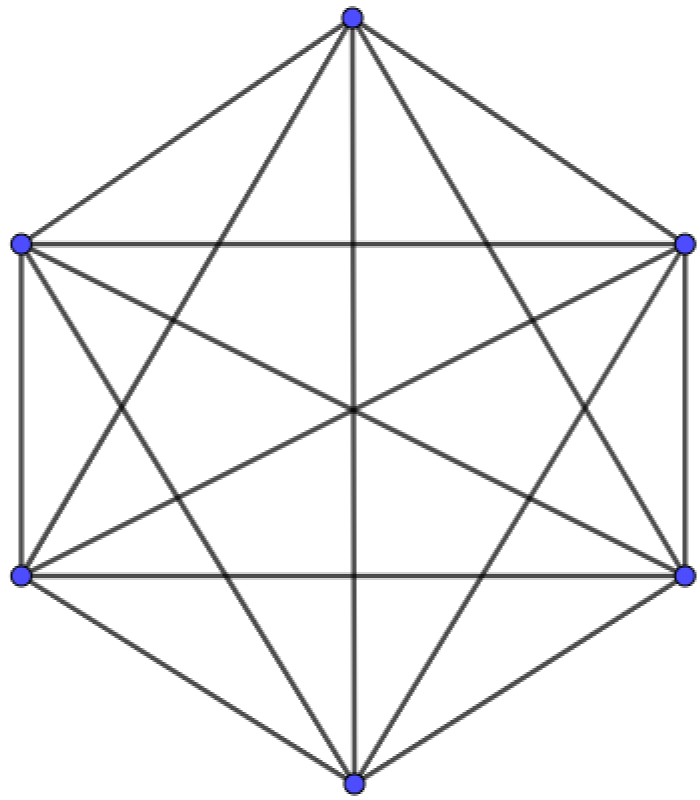


FIGURE 13. The K_6 graph. This was used by Kainen to prove Heawood's Five Color Theorem. K_6 is the graph that has six vertices and there is an edge connecting every vertice. This is known as a complete graph of degree six.

Heawoods' work with this problem did not stop at his proof of the five color theorem. He, as well as many other mathematicians would not stop working toward a definitive answer on the minimal number of colors needed. Since the proof was as of yet elusive, mathematicians worked along several different paths trying to find a revolution that would make it possible to solve. On way of doing this is to find ways to find equivalent forms of the problem in different mathematical principles. One example of this was by Heawood. He took the problem and found the pure algebraic equivalent [**8**].

Having the five color theorem proven, but not the four color theorem lead to a lot of doubt amongst mathematicians on whether the minimal number of colors needed is four or five. Although there was suspicion that four colors would not be sufficient, this did not stop mathematicians from developing the ideas emerging continually farther, trying to find certainty and clear ways of expressing the finding of mathematics.

More Advancements

Over the next 80 years, no proof for the four color problem was found. Although this sounds very discouraging, it does not mean that no advancements were made. In fact, it was said that all the techniques needed to prove the theorem were known before 1950, it was just that there was too much data for one person, or even a group of people, to compute and formulate by hand. Throughout this section several of the advancements are mentioned. This is not an exhaustive list of all advancements made in this time, as that would be so unweildy as to lose sight of this papers main focus.

The first of these advancements is done P.G. Tait. As Heawood wrote an equivalent problem to the four color problem in algrebraic form, Tait focused on writing an equivalent problem that stayed in the realm of graph theory. Tait accomplished this through the conjecture which can be stated as follows, "The edges of every cubic, bridgeless, planar graph are 3-colorable[**27**]." This conjecture is true if and only if the four color problem can be found to be true. Since the four color problem had yet to be proven, if this conjecture could be proven to be true, then the four color theorem would also be proven true. While Tait thought he had the proof to this conjecture, and thus the proof for the theorem [**24**], his work was disproven as Kempe's had been before him. Later, Tait's ideas were used in a more generalized manner by Tutte to show the ideas of nowhere-zero flows[**10**]. The work done by these men have expanded to cover much larger parts of flows and edge colorings.

Another advancement that was done within these years was done by R. L. Brooks. Brooks has a theorem acredited to him that says for every graph with a maximum vertex degree of d has a d-coloring unless either the graph contains K_{d+1} or d = 2 and the graph contains an odd cycle[**9**]. Although this is a theorem that is more genral than the four color problem, it is still a useful advancement in graph theory. Another reason that this is useful to the four color problem is that Kempe chains can be used to prove Brooks theorem. Mel'nikov and Vizing used Kempe chains to write a fairly short proof of Brooks theorem[**17**]. This shows that the ideas that had previously been used were not just left, but expanded upon and used to grow into new areas. The four color problem was what caused the thinking behind the entirety of graph theory.

Although the previously mentioned advancements eventually veered away from directly working toward a solution for the four color problem does not mean that it was not still being worked on. As is mentioned early, throughout the early 1900s the problem was slowly whittled away on by several men who were working on raising the number of regions in a map that could always be four colorable. Franklin, the first to start this trend showed that any graph that has 25 or fewer regions was four colorable. In 1926 this number was raised to 27 regions by Reynolds. Franklin raised the number to 31 in a paper published in 1938. It was raised two more times, first by Winn to 35 in 1940 and finally by Ore and Stemple to 39 in 1970[**22**]. Each of these men had to show that no matter the configuration of the maps with this certain number of regions, it was always four colorable. Their work was based off the reducibility findings done by George Birkhoff in 1913[**6**]. Although these advancements helped us get a better understanding of the problem, this form of study could not be used to lead a proof.

Finally a Proof

Although many advancements were made and a great deal more was found about graph theory as we now know it, the proof to the four color problem was still elusive, even though it was over a hundred years since mentioned by De Morgan. Now that the 1970s have come around, technology has advanced enough so that all the work that was before seen as impossible to get done, may now have a chance.

Appel and Haken

In the early 1970s Kenneth Appel and Wolfgang Haken got together to work toward the solution for the four color problem[**1**][**2**]. The work that these two did together is described in their work "The Solution of the Four-Color-Map Problem. " The article begins going over the history of the problem up until that point. They continue on to then say that to begin their work, "We therefore decided to first study certain kinds of discharging procedure in order to determine the types of sets of obstacle-free configurations that might arise[**3**]." As can quickly be seen between this quote and ones given previously, the terminology has been greatly altered through the time span that the problem has been worked on. The idea of labeling each of the vertex with charges was brought upon by Heinrich Heesch. This idea follows from the work by Kempe done with Cauchy's formula. Since there are these charges now put upon the vertices there are procedures that can be done in which the charges are moved around the graph. The overall charge does not change, but the charge on any given vertice might. By picking a specific discharging procedure a finite list of configurations can be made. This process is explored, "Since the configurations signaled by this procedure must form an unaviodable set, if they are also reducible then the four-color conjecture is proved[**3**].” On the other hand, if they are not reducible, then there is not advancement toward a proof. After repeating this process on a computer several times, they had found “an un-avoidable set of good configurations.” It was not until after they found these that the proof was found that showed they were indeed what they claimed.

Now that Appel and Hanken have found and proven their unavoidable set, they begin to work on the reducibility section of the proof. This, they thought would be extremely difficult, even with the use of computers. Yet, they recalled that Heesch did work on reducibilty and he found some “phenomena that provide clues to the likelihood of successful reduction[**3**].” After working on reducibility for a few years, “by the fall of 1974, [they] had a lengthy proof that a finite unavoidable set of geographically good configurations does exist, and [they] had a procedure for constructing such a set with precise, although much larger than desirable, bounds on the size of the configurations of the set[**3**].” Once this was done, and a proof was found, they set out to change the parameters from good configurations to ones that are just obstacle-free. By obstacle free they are refering to three arrangements which lead to a graph not being able to be proven reducible. Here it is important to note that one of the arrangments occur on a graph when a map has exactly four regions meet at a single point (see Figure 14). Due to this fairly common complication many maps are not included in the proof by Appel and Hanken, such as the United States due to the Four Corners, and the state of Kentucky.

This work continued until the next summer when the duo finally started to believe they could find an unavoidable set of obstacle free configurations which were indeed redicible. As with all the steps before this, they spent much of their time writing and perfecting lines of code while the computers spent several hours performing calculations. Although it took the computer so long to do this, it would have taken people immensely more time. The work on programming for reduction continued for some time with the help of Koch. It was also slightly modified to fall more in line with

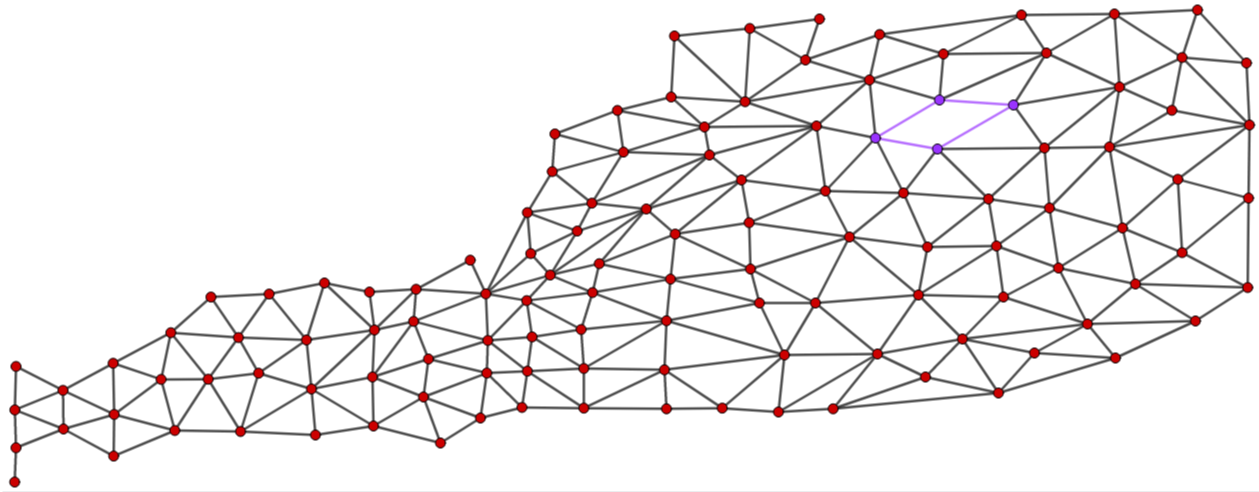


Figure 14. This is the dual graph of the Commonwealth of Kentucky. It is another example to show that all maps can be turned into graphs, as all maps are planar. This particular map is one of the ones that causes a problem for Appel and Haken, as the four vertices that are a seperate color do not create a triangulation like all the other areas do. This is one of the three types of obstacles that Appel and Haken did not know how to prove reducibility for, so they left them out of their set.

the work of Birkhoff. It was at this time that the work of the discharging procedure could go no farther without a complete rehaul of the computer code. Rather than do that, it was decided that it would be more worthwhile to continue this process by hand. Since humans were working on this process now, more "flexibility" could be put into the procedure allowing for the unavoidable set to be even more narrowed down. Although it was narrowed down, it would still take a great deal of computational ability to perform all the tasks needed. The computers were put to work and from here Appel and Hanken have proven the four color theorem.

Appel and Hanken conclude their paper saying that many mathematicians are not ready to accept the proof by computer, but that new ideas need to be embrased upon occasion. At this time, there were no other mathematical proofs that relied on computers. This particular proof was even mentioned in the article "Ugly Mathematics: Why Do Mathematicians Dislike Computer-Assisted Proofs?" Here, Montano shows that from the beginnings of computer assisted proofs there have been negative receptions for the mathematical community[**18**]. Montano has several compeling reasons, but for at least this one proof, some of the resistence may come from the lack of clarity and description given in the written and published proof. Even though computers were becoming more common, the proofs were still questionable.

Could this Be?

Several years after the work done by Appel and Hanken, there was still doubt into their computer proof, not only because it was done on a computer, but also because it was so inaccessible to be verified. A quote from their own paper states, "This leaves the reader to face 50 pages containing text and diagrams, 85 pages filled with almost 2500 additional diagrams, and 400 microfiche pages that contain further diagrams and thousands of individual verifications of claims made in the 24 lemmas in the main sections of text. In addition, the reader is told that certain facts have been verified with the use of about twleve hundred hours of computer time and would be extremely time-consuming to verify by hand. The papers are somewhat intimidating due to their style and length and few mathematicians have read them in any detail[**4**]." This quote is used by Robin Thomas in his article that gives an updated, and significantly more simple proof of the four color theorem[**26**]. Thomas uses this particular quote to show why it is beneficial for the mathematical world to spend time on a different proof than the one found by Appel and Hanken. Thomas was up for the challenge and did just that. He began by assessing the parts of the 1977 proof that were troublesome. Upon reflection he came up with two areas: the first is that a computer must be used, and the second is that even though part of it is said to be hand-checkable, it has not been due to length and lack of clarity. In order to settle at least the second of these issues, Thomas with his colleagues Neil Robertson, Daniel P. Sanders, and Paul Seymour first attempted to verify the part of the proof that was said could be done by hand. It did not take long before this attempt became clearly useless. Rather than leave both problems in place, they decided to create their own proof. Though they acknowledge that the first issue cannot be changed, the second can be, by making the hand done parts more accessible.

After covering the basic history of the problem and several ideas that others before them have found that will be used, the paper gives a brief overview of what to expect from this proof. That overview is as follows:

> The main aspects of our proof are as follows. We confirm a conjecture of Heesch that in proving unavoidability a reducible configuration can be found in the second neighborhood of an "overcharged" vertex; this is how we avoid "immersion" problems that were a major source of complication for Appel and Haken. Our unavoidable set has size 633 as opposed to the 1,476-member set of Appel and Haken; our discharging method uses only 32 discharging rules instread of the 487 of Appel and Haken; and we obtain a quadratic algorithm to 4-color planar graohs, an improvement over the quartic algorithm of Appel and Haken. Our proof, including the computer part, has been independently verified, and the ideas have been and are being used to prove more general results. Finally, the main steps of our proof are easier to present, as I will attempt to demonstrate below[**26**].

Thomas makes it clear in his comparisons that this proof will follow a similiar idea to Appel and Hakens, but the actual proof has been made significantly more simplified. This simplification, as well as the better and more readily available computing options, makes this proof significantly more accessible by fellow mathematicians.

As with Appel and Haken's proof, one of the major components used by Thomas is that of reducibility. The ideas and practice of reducibilty is derived from work by Birkhoff, Bernhart, Heesch, Appel and Haken and others. The definitions of several types of reducibility can be found at [**21**]. They do make it clear in their 1998 paper that in order to do these types of reducibility, computers are needed, as one case can have up to 200,000 colorings to be checked. The rest of their proof follows the ideas used by Appel and Haken, just in a more systemized manner. This is in part due to there being so many less cases to check.

Continuing the Ideas

Now that the four color has been proven in a manner that it can feasibly be checked, without finding any errors, we can now say that the four color problem is indeed solved, after only 150 years. Although the proof was found, the work behind and for graph theory have not stopped. Graph theory was developed for this problem but has grown into so much more. With one search on the popular internet search engine, Google, hundreds of millions of results for graph theory appear. Although maybe the largest overarcing conseqence of the four color problem, graph theory is certainly not the only lasting result. Over the time it took to solve this problem there were many advancements made in hopes of making this problem solveable.

Now that the four color conjecture has been turned into a theorem, the work can now be put back into practical uses, just as it originally started out as. In the article "An Evolutionary Algorithm Based on the Four-Color Theorem for Location Area Planning" it describes how the four-color theorem is used to not to color maps, but actually to help plan where mobile network towers need to be placed to optimize the network reception for customers[**7**]. In this modern era, this plays a larger part than just coloring a map. It is also a significantly more difficult concept then coloring and more costly if done incorrectly. This shows that what was just a simple concept can have a great impact on society.

Another result of the four color theorem being proven is that it can now be used to prove other mathematical topics. Since the problem was found to have so many equivalent forms, the proof of our main theorem now can help prove all of these, as well as other concepts. One of these proofs using the four color theorem can be found in Alex Wendlands, "Coloring of Plane Graphs with Unique Maximal Colors on Faces" [**28**]. This takes the concept of the four color theorem and moves it further along into the world of coloring graphs.

Not only was the theorem used for practical purposes and to help prove even more things in mathematics, but even the proof of this revolutionary theorem has made a large impact on mathematics. The use of a computer was a new concept and one that was not accepted for many, many years. Although the use of this new technology was initially frowned upon the tides have changed in recent years. People such as Uwe V. Riss in his article "Objects and Processes in Mathematical Practice" take a strong stance in pushing to have computers accepted as a tool to help with proofs[**20**]. Although during the proof discussed in this paper it was ground breaking to use a computer, it is now becoming the solution to many of maths most difficult proofs [**25**]. Proofs using computers are expanding, all because of the four color problem.

Bibliography

[1] Kenneth Appel and Wolfgang Haken. Every planar map is four colorable. *Bulletin ofthe American Mathematical Society*, 82:711–713, 1976.

[2] Kenneth Appel and Wolfgang Haken. Every planar map is four colorable. *Illinois Journal of Mathematics*, 21:439–597, 1977.

[3] Kenneth Appel and Wolfgang Haken. The solution of the four-color-map problem. *Scientific American*, 237(4):108–121, 1977.

[4] Kenneth Appel and Wolfgang Haken. The four color proof suffices. *Mathematical Intellifencer*, 8:10–20, 1986.

[5] John A Bather. Mathematical induction. 1994.

[6] George D Birkhoff. The reducibility of maps. *American Journal of Mathematics*, 35(2):115–128, 1913.

[7] Lei Chen and Hai-Lin Liu. An evolutionary algorithm based on the four-color theorem for location area planning. *Mathematical Problems in Engineering*, 2013, 2013.

[8] HSM Coxeter. The four-color map problem, 1840-1890. *The Mathematics Teacher*, 52(4):283–289, 1959.

[9] Daniel W Cranston and Landon Rabern. Brooks' theorem and beyond. *Journal of Graph Theory*, 80(3):199–225, 2015.

[10] Miquel A Fiol, Giuseppe Mazzuoccolo, and Eckhard Steffen. On measures of edge-uncolorability of cubic graphs: A brief survey and some new results. *arXiv preprint arXiv:1702.07156*, 2017.

[11] R. Fritsch, J. Peschke, and G. Fritsch. *The Four-Color Theorem: History, Topological Foundations, and Idea of Proof.* Springer New York, 2012.

[12] Georges Gonthier. Formal proof–the four-color theorem. *Notices of the AMS*, 55(11):1382–1393, 2008.

[13] Percy J Heawood. On the four-colour map theorem. *Quart. J. Pure Appl. Math.*, 29:270–285, 1898.

[14] Paul C Kainen. A generalization of the 5-color theorem. *Proceedings of the American Mathematical Society*, 45(3):450–453, 1974.

[15] Alfred B Kempe. On the geographical problem of the four colours. *American journal of mathematics*, 2(3):193–200, 1879.

[16] B.G. Luisa, R.L.G.M.G. László Lovász, R.L. Graham, M. Grötschel, L. Lovasz, L. Lovász, J.M. Lemm, and G. Meurant. *Handbook of Combinatorics Volume 1*. Handbook of Combinatorics. Elsevier Science, 1995.

[17] L.S. Mel'nikov and V. G. Vizing. New proof of brooks' theorem. *J. Combin. Theory*, 1969.

[18] Ulianov Montano. Ugly mathematics: Why do mathematicians dislike computer-assisted proofs? *Math Intelligencer*, 34:21–28, 2012.

[19] Oystein Ore. *The four-color problem*. Pure and Applied Mathematics, Vol. 27. Academic Press, New York-London, 1967.

[20] Uwe V Riss. Objects and processes in mathematical practice. *Foundations of science*, 16(4):337–351, 2011.

[21] Neil Robertson, Daniel Sanders, Paul Seymour, and Robin Thomas. The four-colour theorem. *journal of combinatorial theory, Series B*, 70(1):2–44, 1997.

[22] Thomas L. Saaty and Paul C. Kainen. *The four-color problem*. McGraw-Hill International Book Co., New York-Bogotá-Auckland, 1977. Assaults and conquest.

[23] Timothy Sipka. Alfred bray kempe's "proof" of the four-color theorem. *Math Horizons*, 10(2):21–26, 2002.

[24] PG Tait. Remarks on the coloring of maps. In *Proc. R. Soc. Edinb*, volume 10, pages 501–503, 1880.

[25] Pingzhong Tang and Fangzhen Lin. Computer-aided proofs of arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

[26] Robin Thomas. An update on the four-color theorem. *Notices of the AMS*, 45(7):848–859, 1998.

[27] David B Weinberger. Network flows, minimum coverings, and the four-color conjectures. *Operations Research*, 24(2):272–290, 1976.

[28] Alex Wendland. Coloring of plane graphs with unique maximal colors on faces. *Journal of Graph Theory*, 83(4):359–371, 2016.

[29] Douglas B. West. *Introduction to graph theory*. Prentice Hall, Inc., Upper Saddle River, NJ, 1996.

[30] Robin J. Wilson and John J. Watkins. *Graphs*. John Wiley & Sons, Inc., New York, 1990. An introductory approach, A first course in discrete mathematics.