Taking a stab at the four color theorem

Van den Driessche Willy.

With the invaluable aid of Van den Driessche Wouter.

In this document we present an alternative proof of the classical 4 color theorem.   
This theorem states that 4 colors suffice to color any planar graph   
in such a way that adjacent vertices always receive different colors.

30 December 2015.

Contents

[1 The four color theorem 2](#_Toc439260775)

[2 Maximal planar graphs 2](#_Toc439260776)

[3 Three rules 3](#_Toc439260777)

[3.1 Triangle rule 3](#_Toc439260778)

[3.2 Quadrilateral rule 3](#_Toc439260779)

[3.3 Pentagon rule 4](#_Toc439260780)

[4 Inductive approach 4](#_Toc439260781)

[5 Chromatic polynomial 4](#_Toc439260782)

[5.1 Shadow edges 5](#_Toc439260783)

[6 Triangle rule leaves the colofour unchanged 6](#_Toc439260784)

[7 Quadrilateral yields a positive colofour. 6](#_Toc439260785)

[7.1 Proof 1 9](#_Toc439260786)

[7.2 Proof 2 9](#_Toc439260787)

[8 Pentagon yields a positive colofour. 10](#_Toc439260788)

[8.1 Some symmetry 10](#_Toc439260789)

[8.2 3 empty pentagons 11](#_Toc439260790)

[8.3 2 empty pentagons 11](#_Toc439260791)

[8.4 Less than 2 empty pentagons : wrong way 13](#_Toc439260792)

[8.5 Less than 2 empty pentagons 14](#_Toc439260793)

[9 Conclusion 15](#_Toc439260794)

[10 References : 15](#_Toc439260795)

# The four color theorem

Before we attack the problem, we will first add some context to it [1]. [2] [3]

The original problem was stated in 1852 when Francis Guthrie was trying to color a map of England. He noticed that only 4 colors where required and wondered whether that was always true.

The problem is this : given any map on a plane or a sphere. If we want to color the countries in such a way that no two adjacent countries are colored using the same color, what is the maximum number of colors required so that we can achieve such a coloring. For this problem two countries are considered adjacent if they shared a border that is longer than a single point.

A first (ingenious but wrong) proof was given by Alfred Kempe in 1879. This proof was shown to be incorrect by Percy Heawood in 1891. From the remains of the proof the latter was able to show that 5 colors are sufficient. He was not able to show that 5 are required. Bizarrely, he was also able to give a formula that works on all surfaces of arbitrary genus. Unfortunately the formula did not work for a plane or a spherical surface (having genus 0).

Around 1960, Heinrich Heesch developed a method fo “discharging” that was amenable to automation by computer. Then on June 21, 1976 Kenneth Appel and Wolfgang Haken finally proved the conjecture and turned it into a theorem. Their proof is based on unavoidable sets and reducible configurations and requires a computer to verify [4] [5].

In 1996, the proof was “re-done” by Neil Robertson, Daniel P. Sanders, Paul Seymour, and Robin Thomas, using essentially the same approach. They use better algorithms and less reducible configurations and are therefore able to give an alternative computer proof.

Although the computer proof is now widely accepted, it is no secret that many mathematicians are a little bit frustrated that nobody has found a simpler proof that can be understood without the use of a computer. We are aware of the delicate nature of this problem and it is therefore in the humblest possible way that we present our attempt at a proof of the problem.

# Maximal planar graphs

We approached the problem in a classical way. It is easy to see that maps can be replaced with planar graphs.

First of all, we reduced the search to *maximal* planar graphs. A planar graphs is graph that allow for an embedding in the plane without having two edges intersect each other. A famous result by Kuratowski [5] [6] proves that a graph is planar if it does not contain a subgraph that is homeomorphic to K5 or K3,3.

|  |  |
| --- | --- |
| K5 | K3,3. |

A *maximal* planar graph is a planar graph in which no edge can be added without rendering the graph non-planar.

If we can prove that maximal planar graphs can be colored using 4 colors, then any planar graph can be colored using at most 4 colors. This stems from the simple observation that we can get to any planar graph from a maximal planar graph by deleting the appropriate edges.

Since deleting an edge is essentially allowing for more possibilities, the number of colorings can never drop due to a deletion of an edge.

# Three rules

We approached the generation of maximal planar graphs in a classical way. Suppose we already have all maximal planar graphs with *n* vertices. We can get to the planar graphs with *n+1* vertices using 3 simple rules.

These rules have been used before to generate all maximal planar graphs [7]. These rules are easiest to visualize when we have an embedding of the graph in the plane.

## Triangle rule

The triangle rule is as follows [9] :

 becomes 

With this rule, we start with an empty triangular face. Then we add an extra vertex in the center of it and connect it to the existing vertices of the triangle. The triangle is part of a maximal planar graph and can therefore have many connections to other vertices *outside* of the triangle.

Clearly, this graph has 1 more vertex and it is again maximal planar.

## Quadrilateral rule

In the quadrilateral rule, we start with a quadrilateral on the left (again this is a part of a maximal planar graph). We delete the vertical edge and place a new vertex in the middle. This vertex is then connected to the vertices that formed the original quadrilateral.

becomes 

Notice again that if we start with a maximal planar graph, this rule will generate another one with one more vertex.

## Pentagon rule

The final rule is the pentagon rule. It can be pictured like this :

becomes 

For the pentagon rule, we first delete two edges from the original graph, we add a new vertex to the center and then join each existing edge in the pentagon to the center.

The nice thing is that these 3 rules suffice to generate all maximal planar graphs [7] [8].

We implemented these rules in a c# program [9] [10] [11] [12] [13] [14] [15] [16] and made heavy usage of Mathematica for testing our hypotheses [18].

# Inductive approach

Since we will often use the “number of ways a graph can be coloured with 4 colours”, we wish to define the colofour as the number of ways a graph can be colored with 4 colours, divided by 24.

Since our approach is based on the 3 rules, we wish to establish that none of the rules reduces the colofour to 0. It is also important to view this as a process instead of just a graph property.

So we start by noticing that K3 and K4 are maximal planar graphs and that they can be coloured with only 4 colours. In what follows we will assume that we can color maximal planar graphs with n vertices. We will use our 3 rules to prove that this implies that it holds for n+1 vertices.

# Chromatic polynomial

First we need to tell something more about the chromatic polynomial (chromial for short) [17] [18] [19] [20] [21] [24]. We don’t need to know a lot about this concept. We just need to know that it is a polynomial in x that can be computed from a graph. When we substitute 4 for x in that polynomial we get the number of ways to color that graph using 4 colours. (i.e. the colofour \* 24)

There is a brilliant insight that allows us to compute this chromial from simpler graphs. The insight is the following. If we want to know in how many ways we can color a graph then we can start by looking at 2 vertices *a* and *b* of that graph. The total number of colorings is the sum of

* colorings where both vertices have the *same* color
* colorings where both vertices have a *different* color

Now the nice thing is that if both vertices have a different color, we can just as well join them with an edge. The edge is the constraint that says that they have to be different.

On the other hand, if both vertices have the same color, then we can just as well “contract” them into a single vertex.

We can picture this concept as :

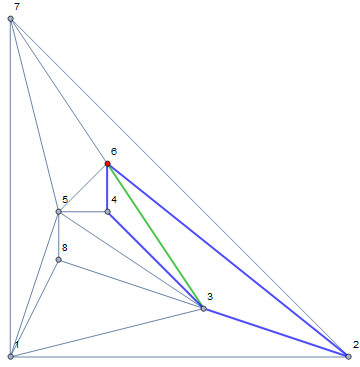
= + 

Or from most complex to less complex :

* =  - 

## Shadow edges

It is important that you view the diagrams that will follow as a snapshot on a complete maximal planar graph. So if we show , this is really a view on a graph like :



The blue portion is the part that has been shown in our snapshot. We make a clear distinction between what happens inside the blue quadrilateral and outside of it. Outside we do require everything to be maximal planar. But as far as we are concerned, an edge which is drawn *inside* the snapshot could also exist *outside* the blue quadrilateral.

Of course, we restrict the existence of edges to the inside and the outside. We do not allow for arbitrary multiplicities of edges. If an edge from the inside is repeated on the outside, we call the one on the outside a “*shadow*” of the edge on the inside. Most of the time our diagrams will only be explicit on the existence or non-existence *inside* the snapshot. They give no information about the outside.

The existence or non-existence of shadows of certain edges will be crucial in our final arguments.

# Triangle rule leaves the colofour unchanged

We will start with rule1. Although it is trivial that this rule leaves the colofour untouched, we will show this using chromials. The key is to reduce the graph with (n+1) vertices to a graph with n vertices. The red edges show the edge on which we apply the chromial rule.

This becomes :

|  |  |  |
| --- | --- | --- |
|  | = - | *Deletion contraction rule* |
|  | =- 2 | *Deletion contraction rule* |
|  | = - 3 | *Deletion contraction rule* |
|  | =(x-3) | *Empty dot is worth an x since it is not connected to the rest* |

If we now calculate this for x= 4 then we have :

(4) = (4)

We now reverse this process to see that rule 1 does indeed leave the colofour unchanged. This implies that a graph with (n+1) vertices produced by this rule has a positive colofour because the one with n vertices has a positive colofour by our induction hypothesis.

# Quadrilateral yields a positive colofour.

The triangle rule has set the stage. We now look at the quadrilateral. This rule has already been proven by Kempe using Kempe chains.

Here we will use a different approach. Again we reverse the rule and start calculating the chromials. Again we try to use the red color to show where the “action” is. A red edge is used to show the edge we will collapse/delete. A red vertex shows two vertices (or the edge on the left side of the equation) that have been contracted.

|  |  |
| --- | --- |
|  | = - |
|  | =-- |
|  | = --- |
|  | =---- |
|  | =(x-2) -(+) |

If we now calculate this for x= 4 then we have :

(4) = 2(4)-((4)+(4))

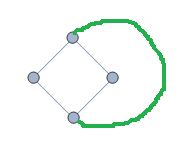
We want to show that the left side is positive. This can only be the case if the right side is positive. In other words, since all participants in the formula are >=0, we want to show that

2(4)> (4)+(4)

This can only happen if either or  is different from 

It is trivial to see that After all, we added a constraint to the right side so we cannot have more solutions than on the left side without the constraint.

One crucial insight is that the only way that  can be *equal* to is that the green constraint was already in the “empty quadrilateral”.

Ok, this doesn’t make a lot of sense does it ? To understand this statement it is important to remember the previous story about shadow edges. The graph parts we schematize are a part of a maximal planar graph. The “outside” of the graph is *there*. It is maximal planar except for the inside of the quadrilateral. Therefore, it is perfectly possible that  is in fact . In other words, the “top to bottom” edge in the example can exist as a *shadow* outside the quadrilateral. Therefore, if we add the new edge, it was “already there”. In that case we have equality of  and . But in all other case we have strict inequality.

If we add an edge to an “almost maximal planar graph” without shadow edges then we reduce the colofour.

We can show this with the quadrilateral.

=+

On the left side we see the case without an added edge. On the right side we the case with the added edge. We also see the (degenerate) contracted left-over of the quadrilateral. It is because all of these components are strictly positive (by our induction hypothesis) that we can see that  strictly < . Because  is strictly positive and it is between these two values.



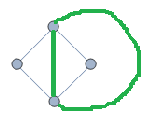
The only way that adding an edge to an “almost maximal planar” graph doesn’t reduce the colofour is because the edge was already in the original graph as a shadow edge.

The proof is completely analogous to the previous. In the previous we could only have equality if  is equal to 0. But since this is a planar graph with n-1 nodes, we should be able to use our induction hypothesis so that it is > 0 . The only reason this could then be = 0 is that the contraction introduced a loop : . But this loop can only have originated from a shadow edge in the source graph.



Now there are two ways to prove that 2(4)> (4)+(4)

## Proof 1

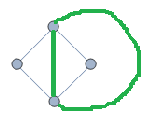
In the first way, we notice that it is possible for  or to be equal to , but not both at the same time. The reason is that equality requires (the example could just as well be done with the “horizontally connected” quadrilateral.) The green connection on the outside makes it impossible to have a green connection on the outside for the horizontal edge.  Therefore they cannot be both be equal to , proving that + < 2  and therefore that  is strictly positive.

## Proof 2

The second way is more subtle but it is a very important way of reasoning because we will need it for the pentagon rule later.

Remember that we want to see all of this as a *process*. We want you to look at the quadrilateral rule.

“becomes ”

We first start from . From this we delete the vertical edge to get  and then we continue with the rest of the process. However, because of this process, we could never have gotten the situation where the vertical edge is also present as a *shadow* outside of the quadrilateral. Because we started from a maximal planar graph and because our 3 rules guarantee that we don’t have duplicated edges. Because we have explicitly deleted the edge from the inside, we know it doesn’t exist as a shadow on the outside.

In other words, due to the way we construct the graphs, we know that < . (We don’t know this for sure for  because our construction works only on the vertical edge). Again, this proves that the sum + < 2 . And therefore again that  is strictly positive.

This constitutes the second proof for the quadrilateral.

# Pentagon yields a positive colofour.

We now look at the pentagon. This rule was the weak part in the Kempe chain approach. Here we will again use a different approach. Again we reverse the rule and start calculating the chromials.

|  |  |
| --- | --- |
|  | = - |
|  | = -- |
|  | = --- |
|  | = ---- |
|  | =----- |
|  | =(x-2) -(++) |

If we now calculate this for x= 4 then we have :

(4) = 2(4)- ((4)+ (4)+ (4))

We want to show that the left side is positive. This can only be the case if the right side is positive. In other words, we want to show that

2(4)> (4)+ (4)+ (4)

Let’s call the right portion of this inequality the *3 amigos*. We will encounter them a lot.

## Some symmetry

During the calculation of the chromial in the previous section, we picked an arbitrary edge to start with. Any other edge would have done too. This will possibly yield another formula that looks very similar to the one we found (2 double diagonals and 1 single diagonal – the union equal to K5 ). However, the equation we found comes in two flavors that are of interest to us :

++

And

++



(Another way to see this is to have a look at what happens to the rules in we flip the pentagon around the vertical symmetry axis that goes through the top vertex and cuts the bottom edge in half)

It will turn out to be very important that has a left and a right alternative. In what follows we will always work around  as the central amigo, although of course any corner would do (provided that is also the corner we use to apply tnhe pentagon transformation rule).

## 3 empty pentagons

We have to start *somewhere* with this problem. The first thing to notice is that each of the 3 amigos contains more “constraints” than an empty pentagon. So we have

>= and  >=and >= (the equal sign indicates we don’t make any assumption about shadow edges yet)



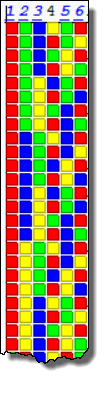
And so we have

3(4)>= (4)+ (4)+ (4)

This is clearly not enough. Since we proved that the colofour of the  is “not too negative”, while in fact it is always >=0.

## 2 empty pentagons

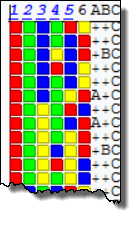
Suppose we list all possible colorings of the empty pentagon. So this will be a list of all vertices of the complete graph, inside or outside the pentagon. This list might look like this :



Each solution to an amigo must also figure in the empty pentagon.

Since an amigo adds a constraint to the empty pentagon, it will remove a solution form the solution set. On the other side, adding a constraint can never introduce new solutions. In this way we can see that the solution set of an amigo is a subset (possibly non-proper) of the solution set of the empty pentagon.

So we can count the number of times each solution of the empty pentagon is also a solution to any amigo. This counting list could look like this :



There is no real simple way to know how often and where which amigo will be a match. We do know one thing : if the solution of the empty pentagon would be a solution to all 3 of the amigos at the same time then in effect it would be a solution to , which is K5 and therefore impossible.

Because of this, each solution can only be counted at most twice and we have

2(4)>= (4)+ (4)+ (4)

Again this is not enough. We have proved that >=0. But this also follows from the construction of chromials. Clearly we need more.

## Less than 2 empty pentagons : wrong way

The exercise in the previous section had a very nice side effect. We noticed that solutions that are only valid for 1 amigo are those that will make it into solutions for .

If you think about it that makes a lot of sense. It took us several months to get where Kempe started : trying to prove that there are solutions in which the empty pentagon is colored with only 3 colors. This leaves a free color for the additional vertex we want to place inside the pentagon. Because a solution to a single amigo is exactly when the empty pentagon is colored with 3 colors, these are the solutions to watch for.

An alternative way of saying this is that we are searching for a solution in which we have two “couples” of the same color in the pentagon.

What we had more than Kempe is the insight that the 3 amigo’s were the ones we needed to focus on. And that made all the difference in the world to us.

So now it is time to go back to the second proof of the quadrilateral. There we stressed the fact that the horizontal and vertical quadrilateral were created using the rules of the quadrilateral as a *process*. If we apply the same to the pentagon rule, then it is clear that the *first* amigo does not have a “shadow” outside of the pentagon. We explicitly delete it from the pentagon so the shadow could not have existed outside the pentagon. So this doesn’t exist :  (And not even the left or right green shadow edge alone can exist. They both do not exist).

If we look at the *second* amigo then we can’t say anything about it. Or can we ? As we pointed out, there are actually two second amigo’s: and . A shadow of such an amigo can come in different versions. After all, each edge of these amigos can occur as a shadow separately, or even in certain combinations. Here is what these would look like :



If you look at these 3 shadow cases it is clear that they cannot exist all 3 at the same time. At most two of them can coexist.

So we have a matrix in which we can picture all solutions to single and double amigos. (The images are somewhat complicated to understand but a red arrow means that the solution is not a solution for that arrow. This makes it possible to view every solution to a single only and a double amigo as something separate). If you don’t understand this picture, don’t worry; it doesn’t matter all that much.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Solutions to |  |  |  |  |  |  |
| … |  |  |  |  |  |  |
| … | X |  |  | X | X |  |
| … |  | X |  |  | X | X |

The crosses mean that at least one of this solution is not possible. So we have shown that certain solutions are excluded from the set of solutions to the empty set. Unfortunately, again this is not enough. If the crosses would have been all on the same line then we would have proved something. Now we have just proven that certain combinations are not possible for certain solutions of the empty pentagon. But since they can be relalized by the others for these combinations, we can still reach the full 2 empty pentagons.

## Less than 2 empty pentagons

The final solution comes when we look our previous findings. We have established that  doesn’t have a “shadow” outside the pentagon. At the same time we have established that at least one green edge inside or does not exist as a shadow.

We have tried to visualize the possibilities in the following table. It is indexed by the possible diagonal edges inside the pentagon, rows as well as columns. If there is a combination that exists without the diagonals having any overlap (forget the fact that they cross it doesn’t matter – in this picture the dash-dotted blue lines just mean that certain vertices will be combined; the edges themselves don’t exist) then we show it.

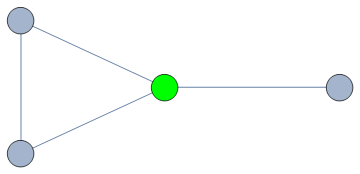
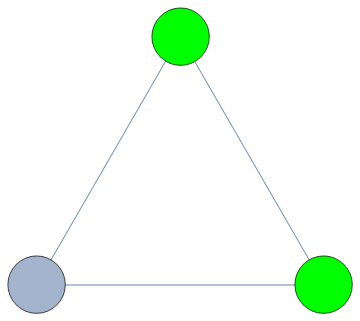
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

The green rows and columns indicate the edges that form the first amigo. As we have shown their shadows never exist so they are always available. The white rows and columns are not necessarily there are the same time but at least one of them is. As you can see there is always a green row for each white column.

What we have established is that we always have 2 non-overlapping diagonals that are not connected through shadows outside of the empty pentagon. In other words : a set of 1 vertex, 2 vertices and another 2 vertices cannot be avoided.

This means that the vertices on these diagonals can receive the same color. And 2 diagonals that have the same colors on their edges implies that we only use 3 colours on the pentagon. And that is exactly what we needed.

Why can we colour these vertices on p.e. using the same color ? Is there not something else that prevents this ?

The answer is that we can count them and their number is strictly positive. We can “contract” both vertices to something like  (remember that the contraction is used to give two vertices the same color). That will again yield a planar graph which can be 4 colored (because it has n-1 vertices). This means that the number of assignments to where both ends of the green edge get the same color is positive. The same argument can be applied to the other diagonal.   
It can also be repeated for both diagonals at the same time giving the interesting contraction : 



# Conclusion

Given the fact that it has taken 163 years to come up with this proof and that all of our arguments are rather elementary, we have the uneasy feeling that we probably made a mistake. Nevertheless, it was a unique opportunity for a father and a son to work together on an interesting mathematical problem.

References :

|  |  |
| --- | --- |
| [1] | W. R., Four colors suffice, how the map problem was solved, Penguin books, 2002. |
| [2] | S. &. Kainen, The four color problem, Mc Graw Hill, 1977. |
| [3] | A. Soifer, The mathematical coloring book, Springer, 2009. |
| [4] | Haken en Appel, „Every planar graph is four-colorable - Part I : discharging,” vol. 21, nr. p429-490, 1977. |
| [5] | Haken en Appel, „Every planar graph is four-colorable Part II : reducibility,” nr. 21, 1977. |
| [6] | C. N. Nishizeki T., Planar graphs, theory and algorithms, Dover publications, 2008. |
| [7] | R. G. Hartsfield N., Pearls in graph theory, Dover publications, 2003. |
| [8] | Brinkmann en McKay, „Fast generation of planar graphs,” vol. Proceedings of the conference 'Computers in Scientific Discovery III', nr. Volume 58 - issue 2, 2007. |
| [9] | Polya, „On Picture Writing,” *American Math monthly,* vol. 63, pp. 689-697. |
| [10] | F. S. Bowen R, „Generation of triangulations of the sphere,” vol. Volume 21, nr. Number 98, 1965. |
| [11] | darrellp, „A C# Implementation of the VF Graph Isomorphism Algorithm,” 1 5 2008. [Online]. Available: http://www.codeproject.com/Articles/23144/A-C-Implementation-of-the-VF-Graph-Isomorphism-Alg. [Geopend 30 12 2015]. |
| [12] | Y. J. Gross J, Graph theory and its applications, Chapman & Hall / CRC, 2006. |
| [13] | T. Roberto, Handbook of graph drawing and visualization, CRC Press, 2014. |
| [14] | J. Dieter, Graphs, networks and algorithms, Springer-Verlag, 2013. |
| [15] | E. Shimon, Graph Algorithms, Cambridge University Press, 2012. |
| [16] | Siek, Lee en Lumsdaine, The boost graph library, Addison Wesley, 2002. |
| [17] | Lau, A java library of graph algorithms and optimization, Chapman & Hall / CRC, 2007. |
| [18] | Goldberg, Efficient algorithms for listing combinatorial structures, Cambridge university press, 1993. |
| [19] | Sloane, „Online encyclopedia of integer sequences,” [Online]. Available: http://oeis.org/A155944. |
| [20] | Birkhoff, „A determinant formula for the number of ways of coloring a map,” vol. 14, 1912. |
| [21] | B. Boerinc, „Computing Chromatic Polynomials for special families of graphs,” 1979. |
| [22] | Dong, Koh en Teo, Chromatic polynomials and chromaticity of graphs, World scientific publishing, 2005. |
| [23] | N. &. Wilf, Combinatorial algorithms, Academic press, 1975. |
| [24] | Grimaldi, Discrete and combinatorial mathematics, Addison wesley publishing, 1985. |
| [25] | Read, „An introduction to chromatic polynomials,” *J. Combin Theory,* vol. 4, pp. 52-71, 1968. |