To Four Colours and Beyond

Johanna Schwartzentruber

A thesis presented for the degree of:
Bachelor of Science

Supervised By: Dr. Hans U. Boden
Department: Mathematics
University: McMaster University
April 24, 2020

0 A Very Brief History

Map colouring has been of interest to map makers for as long as maps have been around. It involves assigning each region of a map a colour such that no two regions which share a border are coloured the same. In 1852, a student of Augustus De Morgan, Frederick Guthrie, asked him a very important question that did not get answered until over a hundred years later. The student's brother, later identified as Francis Guthrie, had found that any map he drew need not more than four colours to colour it. He brought it to his brother's attention who then passed the inquiry along to his professor De Morgan, who on that same day wrote to fellow mathematician William Hamilton inquiring about the problem. Hamilton was too preoccupied at the time to investigate further, and the problem stood untouched for almost twenty-five years [7]. This problem is famously known as **Guthrie's conjecture**.

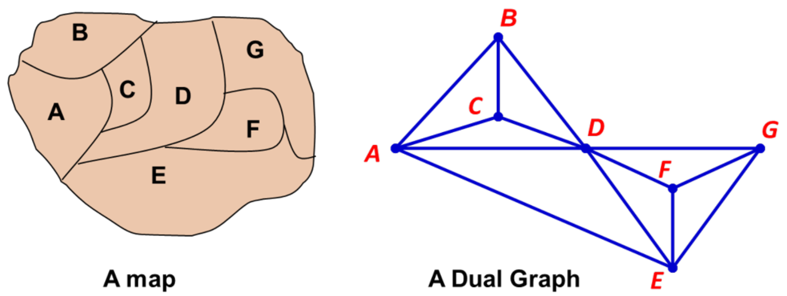
Conjecture 0.1. (Guthrie) Four colours are sufficient to colour any map drawn on the plane.

Then in 1878, mathematician Arthur Cayley sparked interest when he asked both the London Mathematical Society and the Royal Geographical Society if anyone had solved it [7]. Thus began the hunt for a proof of the four-colour conjecture. The news broke in 1879 that a proof had been found by Alfred Kempe and the subject was momentarily put to rest until eleven years later when Percy Heawood found one example of a map in which the techniques used in Kempe's proof to colour it, failed. The search was back on, and as time progressed it became increasingly obvious that this easily comprehensible statement did not have a simple explanation. Nearly a century had passed and with it numerous failed proof attempts had been made. Finally it was announced in 1976 that Kenneth Appel and Wolfgang Haken had done it. Their original proof contained a few small errors they later corrected but they were ultimately responsible for the declaration of the four-colour theorem. The proof however was far from what mathematicians call 'beautiful', and understandably unaccepted by many. It contained over 450 pages of hand-written proof and required over 1200 hours of computer time which made about ten billion logical decisions in the process [7]. Since then other proofs have surfaced, and in 2005 Canadian computer scientist Georges Gonthier wrote a formal proof and mechanically verified it on the proof checker Coq, basically removing any left over doubt of the validity of the statement. Although the actual statement has little known applications, its proof will certainly go down in history as one of the first proofs requiring more than just the human brain, leaving mathematicians and philosophers everywhere to reevaluate what it means to be a proof.

1 Preliminaries

The definitions in this section are referenced from [5], [8], and [9].

Given a **map** M, the **dual graph** G of M can be made by replacing each region in M with a vertex and for every border between two regions in M, an edge is created between the two respective vertices in G.



Since every map has a dual graph and vice versa, every statement made about graphs can be translated into a statement about maps. Thus because of their easily formalized nature, for the remainder of this report, graphs will be used in place of maps. We dedicate the rest of this section to introducing some standard but necessary graph theoretic concepts.

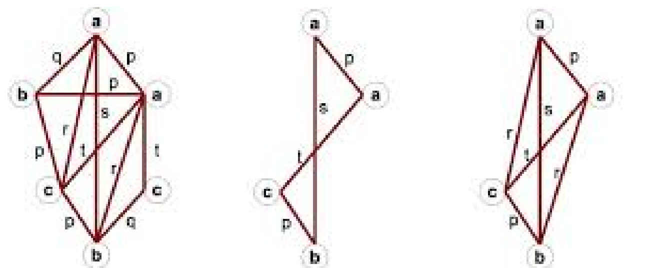
Definition 1.1. A ***graph*** G is a pair $G = (V(G), E(G))$, where $V(G)$ is a non-empty set and $E(G)$ is a set of unordered pairs $\{u, v\}$ for $u, v \in V(G)$. Elements of $V(G)$ are called *vertices* and elements in $E(G)$ are called *edges*. Graphs are typically depicted in the plane by drawing dots for vertices and line segments between vertices u and v to represent edge $\{u, v\}$.

A *loop* in a graph is an edge of the form $\{v, v\}$, and a graph is said to be *simple* if it does not contain any loops. A graph is *finite* if the set $V(G)$ of vertices is finite, and we will, for the remainder of the report, assume all graphs to be finite and simple unless otherwise stated.

An important and closely related concept is that of a *multigraph*, which allows loops and multiple edges. Thus, in a multigraph $G = (V(G), E(G))$, the edge set is generally a multiset of unordered pairs $\{u, v\}$, whereas in a graph, there is at most one edge between each pair of vertices.

Definition 1.2. A ***subgraph*** H of a graph G is a graph whose vertex and edge sets are subsets of the vertex and edge sets of G.

If G is a graph and $S \subseteq V(G)$ then the **induced graph** $G(S)$ is the graph with vertex set S and edge set containing the edges in $E(G)$ whose endpoints are both in S. A subgraph H of G is **induced** if $H = G(S)$ for some $S \subseteq V(G)$.



(a) Labeled Graph (b) Subgraph (c) Induced Subgraph

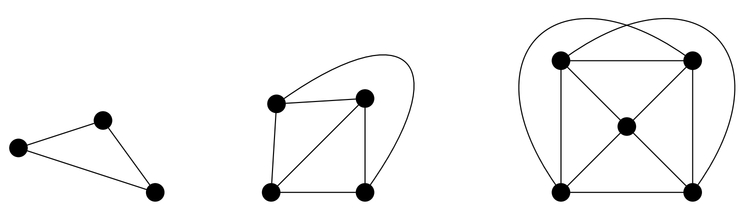
Two vertices, u, v of a graph G are **adjacent** if there is an edge between them, i.e., $\{u, v\} \in E(G)$. Two edges e and f are *adjacent* if they share a common endpoint, i.e., $e = \{u, v\}$ and $f = \{u, w\}$ for some $u, v, w \in V(G)$.

Two graphs G_1 and G_2 are **isomorphic** if there exists a bijection $f : V(G_1) \to V(G_2)$ such that vertices v and u of G_1 are adjacent if and only if vertices $f(v)$ and $f(u)$ of G_2 are adjacent.

A graph G is **bipartite** if its vertex set can be separated into two disjoint sets U and V such that each edge contains an endpoint in both U and V.

A **trail** of a graph G is a sequence of adjacent and non-repeating edges which joins adjacent vertices of G. A **circuit** is a trail beginning and ending at the same vertex. A **path** is a trail which has no repeating edges. A graph G is said to be **connected** if each pair of vertices is connected by a path. Given a connected graph G and integer $k \geq 1$ we say G is **k-connected** if G remains connected after the removal of any subset of $k - 1$ vertices. A **bridge** or **separating edge** is an edge of a graph G whose deletion would cause the number of connected components of G to increase. Graphs with bridges are thus 1-connected, and bridgeless graphs are at least 2-connected.

Given a set of vertices V, the *complete* graph on V is the graph G with vertex set V and edge set containing a unique edge connecting each pair of distinct vertices of V. A graph G is then said to be **complete** if it is isomorphic to the complete graph on $V(G)$.

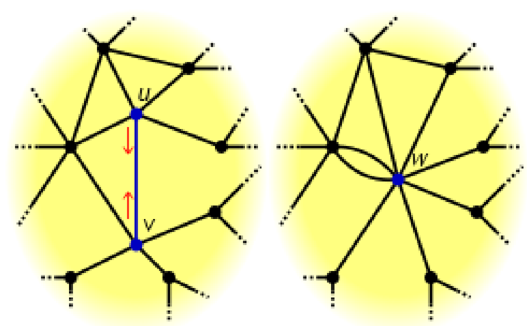


The complete graphs on 3, 4, and 5 vertices.

A graph can be **n-coloured** if there is a labelling of its vertices with n colours in such a way that if two vertices are adjacent they are labelled differently. The **chromatic number** of a graph is the minimum number of colours needed to colour it. Notice that the chromatic number of the complete graph on n vertices (K_n) is n.

The **degree** of a vertex v is the number of edges that are incident to it and is denoted $deg(v)$. Define C_n to be the graph with n vertices, $v_1, \ldots, v_n$, and n edges connecting v_i to v_{i+1}, with i taken modulo n. A **cycle** of a graph G is a subgraph of G isomorphic to C_n for some n. A connected graph is a **tree** if it contains no cycles. A graph is said to be **triangulated** if for every cycle of length 4 or more there exists an edge which is not part of the cycle but joins 2 non-adjacent vertices in the cycle.

If G is a graph, the graph obtained by **contracting** edge e of G with endpoints v and u, or equivalently by **identifying vertices** v and u is the graph of G with the vertices v and u and edge e replaced by a single vertex, with edges connecting to each vertex that v and u were originally connected to. This is called an **edge contraction**. A graph H is a **contraction** of a graph G if H can be obtained from G by a sequence of edge contractions. A graph H is a **subcontraction** of a graph G if it is isomorphic to a contraction of a subgraph of G.



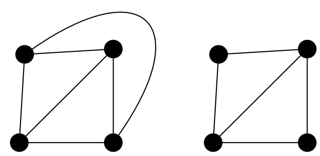
Identifying vertices v and u.

As you may have noticed the definition of a graph is quite abstract (though maybe not as abstract as the topological definition), but all of the examples of graphs given so far have been drawings on a flat surface. It is standard practice to draw graphs in this way which makes it easy to visualize the relation between vertices. It can be pointed out that these are not the only way to draw these graphs.

If you have any topology background you will be familiar with the idea of a surface, a two-dimensional manifold. If not, you can think of a surface as a deformation of the plane. A graph is **embedded** in a surface S if it can be drawn on S in such a way that no two edges intersect, except at vertices. Embedded graphs have an added feature- they have **faces** or **regions**, which are the connected components of S left over when the graph is removed from it.

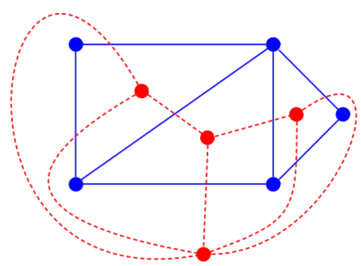
What made the four colour theorem so famous in the first place was that it was understandable by most people. It didn't deal with graphs embedded on Mobius strips or on double tori, it only concerned graphs that could be drawn on paper. A graph is said to be **planar** if it can be embedded in the plane. These graphs are probably the ones you're most familiar with and will be the subject for a majority of this report.

Definition 1.3. A ***near triangulation*** is a planar, 2-connected graph in which every (finite) region is a 'triangle', or bounded by three edges, except for one designated region, called the ***infinite region***.



A triangulation (left) and a near triangulation (right).

It is often convenient to refer to a graph's dual instead- the **dual graph** D of a **planar graph** G is the graph in which every face in G is represented by a vertex in D and for every border separating regions in G, an edge is created in D. It should be pointed out that the dual of a graph may be a multigraph (observe below).



A graph (blue) along with its dual graph (red).

We finish off with some properties of planar graphs. The **degree** of a face f of a planar graph is the number of edges that bound it, and is denoted $deg(f)$. Two faces are *adjacent* if they share a common bordering edge. A planar graph G is **maximal** if the addition of an edge to G makes it non planar.

2 Basic Results

The results in sections 2 and 3 can be found in most graph theory textbooks, see for instance the book [5] by Harary or [9] by Ore.

There are a few basic facts about planar graphs that can be pointed out before we dive in. We will use V, E, and F to denote the number of vertices, edges, and faces respectively of the indicated graph. To begin, a proof of one of, if not the most, famous results in graph theory:

Theorem 2.1. (Euler's Formula for Planar Graphs) If G is a connected planar graph then $V - E + F = 2$.

Before proving this theorem we require a lemma:

Lemma 2.2. If a tree is planar then $V = E + 1$.

Proof. We proceed by induction on E.

Base Case: $E = 0$
Then $V = 1$, and the formula holds.

Induction Step: Assume the formula holds for all trees with E edges, and let G be a graph with $E + 1$ edges. There must exist a vertex v of degree 1. Remove the vertex and its incident edge. By our induction hypothesis this new graph has the property that $V = E + 1$. After adding the edge and vertex just removed, the formula still holds. □

We are now ready to prove the theorem.

Proof. We proceed by induction on E.

Base Case: $E = 0$
In this case $V = 1$ and $F = 1$ so the formula holds.

Induction Step: Assume the formula holds for every graph with E edges and let G be a graph with $E + 1$ edges. If E does not contain any cycles, then $F = 1$ and $V = E + 1$ by lemma 3.2, and so the formula holds. Otherwise consider any cycle in G, and any edge on this cycle. This edge borders 2 faces, so removing that edge creates a graph with one less face. By our induction hypothesis, $V - E + F = 2$. After the addition of the edge we have $V - (E + 1) + (F + 1) = 2$ as well. Thus in either case the formula holds. □

We can now prove some basic relationships between the vertices, edges, and faces of planar graphs.

Lemma 2.3. If G is a connected embedded graph then $3F \leq 2E$.

Proof. Let e_i be the number of edges bounding the face i, such that if an edge appears on both sides of a face it is counted twice. Since every face is bounded by at least 3 edges, we have that $\sum_{i=1}^{F} e_i \geq \sum_{i=1}^{F} 3 = 3F$. Further, since every edge is counted twice in the sum, we have that $\sum_{i=1}^{F} e_i = 2E$, giving us our conclusion. □

Lemma 2.4. If G is a connected planar graph then $E \leq 3V - 6$.

Proof. From Euler's formula we have that $3F = 3E - 3V + 6$, and from lemma 3.3 we have that $3F \leq 2E$. Thus $3E - 3V + 6 \leq 2E$ or equivalently, $E \leq 3V - 6$. □

A classification of bipartite graphs:

Lemma 2.5. A graph G is bipartite if and only if each cycle in G is of even length.

Proof. ($\Rightarrow$) Since G is bipartite there exists two disjoint sets U and V such that $U \cup V = V(G)$ and each edge contains an endpoint in both U and V. If there were an odd cycle C_n, with n odd, it would pass through vertices $\{v_1, ..., v_n\}$. Without loss of generality, assume $v_1 \in U$. Then $v_k \in U$ for k odd and $v_k \in V$ for k even. This would imply an edge between v_1 and v_n which are both in U, contradicting the fact that U and V are the desired sets. Thus every cycle must be of even length.

($\Leftarrow$) Suppose every cycle of G is of even length, and let U and V be two empty sets. Assign the vertices of G to a set as follows: choose any vertex v_0 and assign it to set U. Then consider the subset $S_1 \subseteq V(G)$ which is the set of neighbors of v_0, i.e., all the vertices of G that are 1 edge away from v_0. Assign these to set V. Then consider the subset $S_2 \subseteq V(G) \setminus S_1$ which are all the unassigned neighbors of all the vertices in S_1, i.e., the set of all the vertices 2 edges away from v_0 which have not already been assigned to a set. Assign these to set U. Repeat in this manner. Let $S_n \subseteq V(G) \setminus (\bigcup_{i=1}^{n-1} S_i)$ be the set of all vertices n edges away from v_0 not already assigned. If $n = 2k$ for some integer k, assign all the vertices in S_n to set U, otherwise assign them to set V. We claim each edge in $E(G)$ contains an endpoint in both U and V. First note this algorithm will end because our graph is defined to be finite. Now suppose there was an edge with both endpoints x and y in the same set. Then x and y must be in the same subset S_n for some n since x is at most 1 edge further or closer from v_0 than y is and the vertices in S_K and S_{k+1} are in different sets for every k. So both vertices are n edges away from v_0. Consider the paths of alternating set assignments from x and y back to v_0. If they meet at a common vertex, v', each path from x and y to v' will be of the same length k, $k \leq n$. Since x and y are adjacent, these two paths along with the adjoining edge between x and y form a cycle of length $2k + 1$, an odd number! Thus there are no two adjacent edges assigned to the same set, so U and V are the desired sets and thus G is bipartite. □

We now show the equivalence of terms, allowing us to use them interchangeably.

Lemma 2.6. A planar graph is maximal if and only if it is triangulated.

Proof. ($\Rightarrow$) Suppose a planar graph G is maximal. If G wasn't triangulated there would exist a cycle of length 4 or more with no edge connecting non-adjacent vertices, or equivalently, there would exist a face of degree 4 or more. Then clearly we could add another edge onto G, connecting those two non-adjacent vertices, without compromising its planarity. Thus a maximal graph is triangulated.

($\Leftarrow$) Now suppose G is a triangulated planar graph. If we could add an edge, it would need to connect two non-adjacent vertices (as graphs here are defined to be loop-less without multiple edges) and we would need to do so without intersecting any other edge. The only way to do this would be to connect a pair of vertices which are bounding the same face. But every pair of vertices bounding a face in a planar triangulation are already joined by an edge, since every face is a triangle. Thus G is already maximal. □

Our next result relates to our main topic, the four colour theorem, and thus deserves to be mentioned. First let us present the infamous theorem:

Theorem. *(Four Colour) Any planar graph can be 4-coloured.*

Lemma 2.7. To prove the four colour theorem, it suffices to prove that any maximal planar graph can be 4-coloured.

Proof. The main idea here is that if a graph can be n-coloured then so can any of its subgraphs. Note that any planar graph is a subgraph of a maximal planar graph (just remove vertices and edges). If all maximal planar graphs can be 4-coloured, then we can colour any planar graph G by the following: first triangulate G, thus making it maximal. Then 4-colour it, and remove any vertices (and corresponding edges) to re-obtain G. We claim this is a proper 4-colouring of G. If not, there would be a pair of adjacent vertices coloured the same, but that would mean they'd have to be coloured the same in the triangulated graph of G as well, which had a proper 4-colouring. Thus it suffices to show that any maximal planar graph can be 4-coloured □

3 Other Colour Theorems

As we have seen, some planar graphs require at least four colours to colour them (refer to K_4 above). A natural question arises- when is it possible to colour planar graphs with less than four colours? We will answer this question with the requirements necessary and sufficient for graphs to be 2 and 3-colourable. We conclude this section with Heawood's proof of the 5-colour theorem, which tells us that 5 colours is all that is needed to colour any planar graph. All graphs below are assumed to be connected, though all results can be extended to disconnected graphs by applying the arguments to the connected sub-components. It should be pointed out that we will use the obvious fact that cycles of even (respectively odd) length have chromatic number 2 (respectively 3).

The following result is a property of *all* graphs, planar or not.

Theorem 3.1. A graph is 2-colourable if and only if it is bipartite.

Proof. ($\Rightarrow$) If a graph G is 2-colourable, say in colours black and white, let U be the set all vertices coloured black and V be the set of vertices coloured white. Then clearly $U \cup V = V(G)$. As well, if an edge contained both endpoints in one set then both endpoints would be coloured the same, contradicting the fact that G is properly 2-coloured. Thus each edge contains an endpoint in each of U and V so G is bipartite.

($\Leftarrow$) Suppose G is bipartite and sets U and V are the disjoint sets with each edge in G containing an endpoint in both U and V. Then colour all vertices in U black and all vertices in V white. Claim this is a proper 2-colouring. Suppose not, then there exist adjacent vertices x and y coloured the same. But that would mean x and y are in the same set, contradicting the fact that U and V are the desired sets. Thus G is 2-colourable. □

Recall the notion of near triangulations, see Definition 2.3. A vertex of a near triangulation is said to be *internal* if it is not adjacent to the infinite region.

Theorem 3.2. A near triangulation G is 3-colourable if and only if every internal vertex of G is of even degree.

In 1873, Carl Hierholzer showed that a graph has vertices all of even degree if and only if it is **Eulerian**. A Eulerian graph is a graph with a **Eulerian path** which is a path that visits every edge exactly once (allowing for revisiting vertices). Further, it is known that this path can be made non-crossing. Then the proof requires a fact regarding the number of edges in a Eulerian path, namely [4]:

Lemma 3.3. In a non-crossing Eulerian path of an Eulerian near triangulation, the length of every subcircuit is divisible by 3.

We can then prove Theorem 4.2 [4]:

Proof. (⇒) If G contains an internal vertex v of odd degree then the circuit around v is 3-colourable, thus there is no left over colour for v.
(⇐) Suppose G is a planar, near triangulation whose internal vertices are all even. Then it contains a Eulerian path P so we can colour the vertices of G as follows: start at any vertex and follow along P colouring the vertices 1, 2, and 3 as you traverse the path. Then, since every subcircuits' length is divisible by 3, if we come back to a vertex while traversing P which has already been coloured, it will be assigned the same colour. Thus we get a proper 3-colouring of G. □

Then the necessary and sufficient requirements of 3-colourability for planar graphs follow easily:

Corollary 3.4. A planar graph G is 3-colourable if and only if it is the subgraph of a planar near triangulation whose internal vertices are all of even degree.

If you know any computability theory you are familiar with the concept of NP (nondeterministic polynomial time), and it may interest you to know that the decision problem of whether or not a planar graph is 3-colourable is NP-complete. A decision problem is NP if its solution (which is either yes or no) has proofs of polynomial length whose validity can be checked in polynomial time. For example, given a graph and a colouring of its vertices (which can be stored in a data structure of polynomial length with respect to the number of vertices) it can be checked in polynomial time if that colouring is indeed a valid 3-colouring. If in addition it has no known polynomial time algorithm to determine the solution, it is NP-hard. A problem is thus NP-complete if it is NP-hard and any NP problem can be reduced in polynomial time to it. Basically the problem of determining if a graph is 3-colourable is a very hard problem, which can be inferred by Corollary 4.4. Clearly those requirements are quite difficult to check, and if in fact P ≠ NP there will not be much improvement to the classification of 3-colourable graphs. On the other hand, if one can find a polynomial-time algorithm to determine if a planar graph is 3-colourable then this would solve the infamous P = NP question! This is because that decision problem is NP-complete, so basically that algorithm would provide a polynomial-time algorithm for *all* NP problems, leaving us to conclude P is in fact equal to NP.

Shortly after Kempe's failed attempt to prove the four-colour theorem, Heawood used his ideas to prove the five-colour theorem. This proof requires a fact about planar graphs:

Lemma 3.5. Every planar graph has a vertex v with $deg(v) \leq 5$.

Proof. Let G be a planar graph and let V, E, F denote the number of vertices, edges, and faces of G respectively. From lemma 3.4 we have that $E \leq 3V - 6$. Then if $deg(v_i) \geq 6$ for all i, $2E = \sum_{i=1}^{V} deg(v_i) \geq 6V$. But then since $3V - 6 \geq E$ or $6V - 12 \geq 2E$ we have that $6V \leq 2E \leq 6V - 12$, a contradiction. Therefore G must contain a vertex of degree 5 or less.

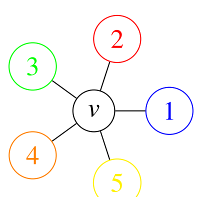
□

Theorem 3.6. Every planar graph can be 5-coloured.

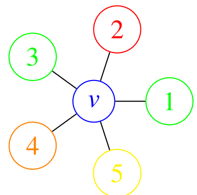
Proof. Let $|V(G)|$ be the number of vertices of the planar graph G. Then proceed by induction on $|V(G)|$. We will suppose each graph G has $|V(G)| \geq 3$.

Base Case: $|V(G)| \leq 5$. Trivial.

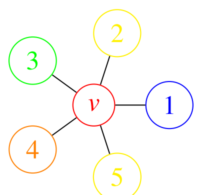
Induction Step: Suppose every graph G with $|V(G)| = n$ can be coloured in five colours. Let H be a graph with $|V(H)| = n+1$. Then by Lemma 1, there is a vertex v with $deg(v)$ ≤ 5. Consider the graph $H' = H \setminus v$. Then $|V(H')| = n$ so it can be coloured in five colours. If $deg(v) \leq 4$, add v back into H' to obtain H and assign it one of the remaining colours not assigned to its neighbors (note there will be at least one colour to choose from). Else $deg(v) = 5$. Add v back into the graph to obtain H. Suppose the neighboring vertices of v (1, 2, 3, 4, 5) are coloured with five distinct colours, else we can assign it a remaining colour.



Consider the subgraph H_{bg} of H, i.e., the subgraph with only the vertices coloured green and blue. If vertices 1 and 3 lie in different connected components, then for every vertex in the connected component of vertex 1, if it is coloured green, colour it blue and vice versa. Then vertex v may be coloured with the remaining colour (in this case blue).



Else vertices 1 and 3 lie in the same connected component, thus 1 and 3 are connected by alternating green and blue coloured vertices. Now consider the subgraph H_{ry} of H, i.e., the graph with only the vertices coloured red and yellow. Note 2 and 5 must not lie in the same connected component, since any path from vertices 2 to 5 (not passing though v) must pass though a vertex coloured either blue or green. Then repeat the same process as above; for every vertex in the connected component of vertex 2, if it is coloured red, colour it yellow and vice versa. Then vertex v may be coloured with the remaining colour (in this case red).



Thus in either case, H has a five colouring.

□

4 Four Colour Theorem

All of the definitions and most of the results in this section can be found in [10] and interested readers are invited to refer to that paper for a more thorough walk through of the proof.

Every graph in this section is connected unless stated otherwise.

Theorem 4.1. (Four Colour Theorem) Every planar graph can be 4-coloured.

Most of the graph theoretic proofs of the theorem have used the same structure; they involve proving that a minimal counterexample cannot exist.

Definition 4.2. A ***minimal counterexample*** is a graph G that cannot be four coloured, such that every subgraph G' with $|V(G')| + |E(G')| < |V(G)| + |E(G)|$ can be four coloured.

Note if there is a map that cannot be four coloured, there must exist a minimal such one. To prove this, one usually shows the existence of a finite set of unavoidable, reducible configurations (defined later). Reducibility refers to the fact that if any of the configurations were to appear in a graph G that would imply that the colouring of

G would depend on a graph G' which has less vertices than G. Each of the configurations are also unavoidable in the sense that at least one must appear in any minimal counterexample, thus leading to our conclusion. In more technical terms, this involves showing the following:

1. Every minimal counterexample is an internally 6-connected triangulation.
2. If T is a minimal counterexample then no good configuration appears in T.
3. For every internally 6-connected triangulation T, some good configuration appears in T.

A good configuration is taken to be any configuration in Robertson, Sanders, Seymour, and Thomas' (Robertson et all.) reducible and unavoidable set of configurations.

Despite Appel and Haken finding the first proof of the theorem, the steps that follow are based off of Robertson et all.'s proof which was more widely accepted for many reasons. Most notably the size of the unavoidable and reducible configuration set was cut down to less than half (633 from 1405), it had a fully computerised unavoidability proof, and contained a four-coloring algorithm which ran in quadratic time as opposed to Appel and Haken's quartic time one [7].

Consider the following definitions:

Definition 4.3. A ***configuration*** K is a pair $(G(K), \gamma_k)$ where $G(K)$ is a near triangulation and $\gamma_k : V(G(K)) \to \mathbb{Z}^+$ such that:

1. For every vertex v of $G(K)$, $G(K) \setminus v$ has at most two components and if there are two, $\gamma_k(v) = deg(v) + 2$.
2. For every vertex v of $G(K)$, if v isn't incident to the infinite face then $\gamma_k(v) = deg(v)$, otherwise $\gamma_k(v) = deg(v) + 2$ and in both cases, $\gamma_k(v) \geq 5$.
3. K has ring size ≥ 2 where the ring size of $K = \sum_v(\gamma_k - deg(v) - 1)$ summed over all vertices v incident to the infinite region such that $G(K) \setminus v$ is connected.

Definition 4.4. A ***free completion*** S of a configuration K with ring R is a near triangulation such that:

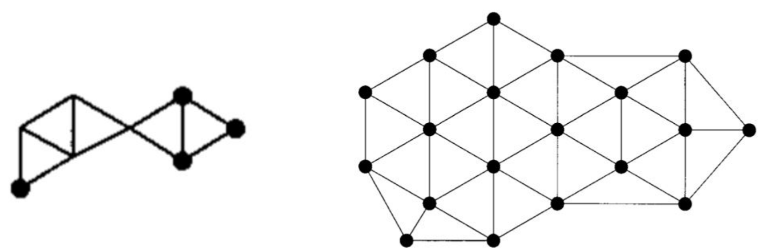
1. R is an induced circuit of S and bounds the infinite region of S.
2. $G(K)$ is an induced subgraph of S, $G(K) = S \setminus V(R)$, every finite region of $G(K)$ is a finite region of S and the infinite region of $G(K)$ includes R and the infinite region of S.
3. Every vertex v of S not in $V(R)$ has degree $\gamma_k(v)$ in S.

Some discussion is in order over these two definitions. First of all, the ring is not a part of the configuration, and its size is actually the size of the ring of the free completion. That being said, a configuration is meant to be embedded in a larger graph. The free completion of a configuration is not unique, but since a homeomorphism can be found between any two free completions of the same configuration, there is essentially one free completion for a configuration. Finally, the label $\gamma_k(v)$ in G(K) is meant to be the degree of v in the larger graph.

Configurations are also usually drawn with different vertex symbols representing $\gamma_k(v)$:

●	$\gamma_K(v) = 5$
·	$\gamma_K(v) = 6$
○	$\gamma_K(v) = 7$
□	$\gamma_K(v) = 8$
▽	$\gamma_K(v) = 9$
⬠	$\gamma_K(v) = 10$

Figure 1: Vertex naming conventions.



A configuration (left) along with its free completion (right).

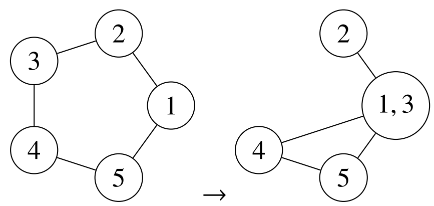
Part one of the proof involves reducing the problem to certain graphs. Consider the first discoveries made about minimal counterexamples [2]:

Lemma 4.5. If G is a minimal counterexample, then every vertex v of G must have $deg(v) \geq 5$.

Proof. First suppose G has a vertex v with degree ≤ 3. Then we can 4 colour the graph $G \setminus v$ (as it has less vertices than G) and then insert v back in and colour it (one of) the remaining colours. If G has a vertex v of degree 4, we can 4-colour $G \setminus v$. If the neighbors of v are not coloured in four different colours we may choose one of the remaining colours to colour v with. Otherwise we proceed with a Kempe chain argument as above. Suppose v has neighbors 1, 2, 3, 4, coloured a, b, c, and d respectively. We consider the graph G_{ac}. If vertices 1 and 3 are not on the same connected component of G_{ac} we can then swap the colours a and c on the connected component of 1 and colour v with the remaining colour, a. Otherwise, there is a Kempe chain connecting 1 and 3 in colours a and c, therefore there is not a Kempe chain connecting 2 and 4 with colours b and d. Thus we can use the same argument, swapping the colours b and d on the connected component of 2 on G_{bd}. This leaves one colour remaining for v, namely b. Thus in every case G is not a minimal counterexample. □

Lemma 4.6. If G is a minimal counterexample, then G is a triangulation.

Proof. Suppose G is a minimal counterexample which is not a triangulation. So G has a face F bordered by more than 3 vertices. Consider the graph obtained by identifying two non-adjacent vertices bordering F to a single point.



An example of a pair of non-adjacent vertices (1 and 3) getting identified on a face with more than 3 bordering vertices.

This graph has less vertices and edges than G and thus can be 4-coloured. This 4-colouring can then be extended back to G, as we can separate the previously identified vertices and colour them the same colour (since they do not share an edge). This leads to a colouring of G, contradicting the fact it is a minimal counterexample. □

Then in 1913, George Birkhoff showed that no minimal counterexample can contain a short circuit [2]:

Definition 4.7. A ***short circuit*** C of a triangulation T is a circuit such that $|E(C)| \leq 5$ and for each region M bounded by C, $M \cap V(T) \neq \emptyset$ and $|M \cap V(T)| \geq 2$ if $|E(C)| = 5$.

The case of a short circuit of size 3 had already been covered. If a minimal counterexample contained a ring with 3 vertices separating the graph into two disjoint, nonempty graphs, each of the distinct graphs along with the ring could be four coloured and by a simple permutation of the colours on the ring, the colourings could be combined to result in a colouring of the whole graph. In this case, a colouring of the whole graph only depends on whether or not the smaller graphs can be four coloured- which is not allowed if the original graph is a minimal counterexample. We then proceed to show the following:

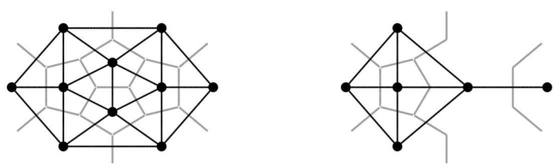
Theorem 4.8. No minimal counterexample can contain a short circuit of size 4.

Proof. Suppose G is a minimal counterexample which contains a short circuit, C, of size 4. Then C separates G into 2 distinct, non empty graphs, G_1 (interior) and G_2 (exterior). Label the vertices of the circuit a,b,c, and d. Consider the graph $A = G_1 + C$ with an edge joining a and c on the exterior. Then because A has less vertices than G it can be four coloured in the colours B, G, Y, W. There are two possible colourings of C (up to permutation of colours)- BGYG and BGYW. Now consider the graph $B = M_1 + C$ with an edge between b and d on the exterior. The possible colourings of C are then BGBY and BGYW. Thus either BGYW extends to M_1 or both BGYG and BGBY extend to M_1. We can then repeat this with M_2 and obtain the same possible colourings of C.

Thus there are three cases, if BGYW extends to both M_1 and M_2, then we can obtain a colouring for the whole graph. Similarly if both BGYG and BGBY extend to both M_1 and M_2 then we can obtain a colouring of the whole graph. Thus we only need to consider the case when, without loss of generality, BGYW extends to M_1 and both BGYG and BGBY extend to M_2. In this case we transform the ring colouring BGYW into one of the other two colourings by a Kempe chain argument.

If, in M_1, there is a B,Y chain connecting a and c then there is not a G,W chain connecting b and d, so we can swap the colours G and W on the connected G,W component of d. Thus we obtain the colouring BGYG. Otherwise we can swap the colours B and Y on the connected B,Y component of c, obtaining the colouring BGBW which is equivalent to BGBY by a mere permutation of colours. In either case we can extend the colouring of C to the whole graph. We can conclude that the colouring of G depends solely on if M_1 and M_2 are colourable, contradicting the fact that G is a minimal counterexample. □

We then define a graph to be **internally 6-connected** if it does not contain a short circuit. Along with the requirements of a minimal counterexample, Birkhoff is also credited for finding the first (and smallest) configuration found in many of the discovered sets of unavoidable and reducible configurations [12]. The configuration, coined the Birkhoff diamond, was found to have a 'reducer' which has less vertices and edges than the free completion of the diamond. It can also easily be shown that any colouring of the reducer extends to the free completion [12]. Thus if a minimal counterexample contained the Birkhoff diamond, the free completion could be replaced by its reducer, four coloured, then have the colouring extended back to the diamond. Therefore it is impossible for the Birkhoff diamond to appear in any minimal counterexample.



The free completion of the Birkhoff diamond (left) beside its reducer, both drawn on top of their dual maps.

This is an example of a B-reducible configuration (named after Birkhoff). There are 4 main types of reducible configurations namely A, B, C, and most famously, D. A, B, and C all include replacing the free completion with a reducer and their distinctions come from how many colouring permutations are needed to extend the colourings of the reducer to the whole configuration. In the case of A-reducibility, the colourings all directly extend, no permutations are needed. The problem with A, B, and C reducibility is that finding a 'safe' reducer (one without loops) is difficult and there is no single test to ensure a reducer is safe [12]. That's why a majority of the configurations are usually D-reducible. Checking for D-reducibility is 'easy' and doesn't require any reducers making it ideal. We now wish to define what it means for a configuration to be D-reducible. The definition that we prefer to use is one that references one of the equivalent formulations of the four colour theorem, discovered to be equivalent by Peter Tait in the late 1800's, which applies only to cubic graphs (graphs in which every vertex has degree 3). Consider the following definition:

Definition 4.9. A ***Tait colouring*** (or edge 3-coloring) of a graph G is a map $K : E(G) \to \{-1, 0, 1\}$ such that no edges which share an end point are labelled the same.

Now an equivalent statement to the four-colour theorem is as follows:

Theorem 4.10. Every bridgeless cubic planar graph has a Tait colouring.

We will now show the equivalence:

Proof. ($\Rightarrow$) Assume G is a bridgeless, cubic, planar graph. Create another graph G' as follows: create a vertex v for every face f in G. Create an edge in G' to represent adjacent faces in G. Note this is equivalent to treating G as a map and finding its dual graph. We can now 4 colour the vertices of G' (since it's planar, loop-less with no multiple edges) which leads to a face colouring of G in 4 colours. Assume G is face coloured with the labels of the Klein four-group: (0, 0), (0, 1), (1, 0), (1, 1). Colour the edges of G as follows: if an edge is between two faces labelled (a, b) and (c, d), colour it ((a + b) mod 2, (c + d) mod 2). Note since G is properly face coloured, the only possible edge colours are (0, 1), (1, 0), and (1, 1). Claim this is a proper edge 3-colouring. If not, there is a face with two incident edges coloured the same. Therefore the other faces which are bounded by each edge respectively must also be coloured the same. But since G is cubic, those faces must be touching and cannot be coloured the same. Thus G can be edge 3-coloured.

($\Leftarrow$) Assume G is a planar graph. Triangulate G and find its dual graph, G'. This graph is planar, cubic, and bridgeless (as G was triangulated). Thus we can find an edge 3-colouring of G' in the following labels, A, B, and C. Note if we can find a face colouring of G' (which includes the infinite face) this easily translates to a vertex colouring of the triangulation of G and thus G. Now observe that each face of G' is contained in the interior of either an even or odd number of A-B and A-C cycles. We will first colour the faces of G' by the following: first we will give each face two labels. If a face is is contained in an odd number of A-B cycles, label it α, else label it β. If it is also contained in an odd number of A-C cycles label it γ, else label it σ. Then give the faces a final colouring: if a face is coloured β and σ label it 0, α and σ, 1, β and γ, 2, and α and γ, 3. Claim this is a proper face colouring of G'. This is because if adjacent faces F_1 and F_2 share an edge e then e is labelled A, B, or C. If it is labelled B it is a part of one of the A-B cycles surrounding either F_1 or F_2 but not both. Thus the number of A-B cycles surrounding F_1 and F_2 differ by 1, so F_1 and F_2 cannot be labelled the same. Similarly, if e is labelled C it is a part of one of the A-C cycles surrounding either F_1 or F_2 but not both, so again F_1 and F_2 cannot be labelled the same. If e is labelled A it is a part of both an A-C cycle and a A-B cycle, so the above arguments apply. Therefore it is possible to properly face 4-colour G' and thus 4-colour G. □

We also need a few preliminary definitions. Let C be a circuit. A **match**, m, of C is an unordered pair of edges of C, and a **signed match** of C is a pair (m, u) where m is a match and $u = \pm 1$. A **signed matching** of C is a set M of signed matches such that if $(\{e, f\}, u)$, $(\{e', f'\}, u')$ in M are distinct then:

1. $\{e, f\} \cap \{e', f'\} = \emptyset$.
2. e, f belong to the same connected component on C if e' and f' are deleted.

Let $E(M) = \{e \in E(C) \mid e \in m \text{ for some } (m, u) \in M\}$.

For $\theta \in \{-1, 0, 1\}$ an edge colouring K of C θ-**fits** a matching M of C if:

1. $E(M) = \{e \in E(R) \mid K(e) \neq \theta\}$.
2. For every $(\{e, f\}, u)$ in M, $K(e) = K(f)$ iff $u = 1$.

A set E of edge colourings of C is ***consistent*** if for every $K \in E$ and every $\theta \in \{-1, 0, 1\}$ there is a signed matching M such that K θ-fits M and E contains every edge colouring that θ-fits M. We are now ready to define D-reducibility:

Definition 4.11. Let S be the free completion of a configuration K with ring R. Let E^* be the set of all edge colourings of R and let $E \subseteq E^*$ be the set of all restrictions to $E(R)$ of 3-edge colourings of S. Let E' be the maximum consistent subset of $E^* \setminus E$. K is ***D-reducible*** if $E' = \emptyset$.

You can think of D-reducibility as such- a configuration is D-reducible if any colouring of the ring of its free completion can be translated to a colouring of the configuration, with Kempe-chain colour swaps if needed. With this idea in mind it's easy to see why a D-reducible configuration cannot appear in a minimal counterexample, since it could be removed, the new graph 4-coloured, then inserted back in and coloured by extension.

It is hopefully clear by now that if a minimal counterexample does exist, it cannot contain a reducible configuration. Thus comes the next and final part of the proof. The existence of a set of unavoidable configurations, *all* of which are reducible. Again, a set of configurations are unavoidable if every minimal counterexample must contain at least one configuration from the set.

The way in which Appel and Haken found unavoidable configurations is rather clever, and involved a technique known as discharging. Discharging involves assigning each vertex an initial 'charge' such that the total sum of the charges is positive. The charges are the redistributed according to a set of rules (keeping the total sum constant) and from the possible outcomes, there become configurations that must be present [1]. Consider this simple example showing the usefulness of discharging:

Lemma 4.12. Every maximal planar graph with vertices of degree at least 5 contains either two adjacent vertices of degree 5 or two adjacent vertices, one of degree 5 and the other of degree 6.

Proof. Let G be a maximal planar graph whose vertices all are of degree at least 5. Assign to each vertex v a charge of $60*(6 - deg(v))$. Denoting v_i to be the number of vertices of degree i and n to be the maximum vertex degree, we have that:

$$\begin{aligned}
TotalCharge(G) &= 60 * ((6-5)v_5 + (6-6)v_6 + (6-7)v_7 + ... + (6-n)v_n) \\
&= 60 * \sum_{i=5}^{n}(6-i) * v_i \\
&= 60 * (6 * \sum_{i=5}^{n} v_i - \sum_{i=5}^{n} i * v_i) \\
&= 60 * (6V - 2E) \\
&= 60 * 12 \\
&= 720.
\end{aligned}$$

Thus the total charge on G is 720, only vertices of degree 5 are positively charged, vertices of degree 6 have no charge, and the rest are negatively charged. Distribute the charges as such: all vertices of degree 5 give each of their neighboring vertices of degree 7 or more, 12 units of charge. After this redistribution, the total charge must still be positive. Observe that vertices of degree $d \geq 8$ have gained a charge of at most $12d$. That leaves their current charge at, at most:

$$60 * (6-d) + 12 * d = 360 - 48 * d \leq 360 - 384 = -24.$$

Thus there must be at least one positively charged vertex of degree 5 or 7. In the case of a positively charged vertex v of degree 5, this implies that v has a neighbor of degree either 5 or 6. In the case of a positively charged vertex v of degree 7 this implies that v has at least 6 neighbors of degree 5, and since G is a triangulation, at least 2 of those neighboring 5-vertices must be adjacent. □

It was with this technique that unavoidable sets were found. The tricky part was ensuring all the configurations in the set were reducible, for if it was found that a configuration was irreducible the rules had to be modified. The above is of course a very simple example of how discharging was used to restrict possible configurations, but it can be admired how much work would have gone into finding the over 400 rules Appel and Haken used in order to find their set. Robertson, Sanders, Seymour, and Thomas on the other hand were able to construct their set with a mere 32 rules (pictured below).

After verifying that each of their 633 configurations were reducible, Robertson et all. needed to prove that they were all in fact unavoidable. They did this in cases depending on the size of the cartwheel.

Definition 4.13. A ***cartwheel***, W, is a configuration such that there is a vertex w (called the ***hub*** of W) and two circuits C_1 and C_2 of $G(W)$ with the following properties:

1. $\{w\}$, $V(C_1)$, $V(C_2)$ are all pairwise disjoint with union being $V(G(W))$.
2. C_1 and C_2 are induced subgraphs of $G(W)$ and C_2 bounds the infinite region of $G(W)$.

3. w is adjacent to all vertices of C_1 and none of C_2.

In order to understand just how they managed to prove this we require a few more definitions. Recall the definition of a *path*, Q; whose length we will denote by $|E(Q)|$. It is called a **u-v path** if u,$v \in V(Q)$ and u and v are the vertices of Q with degree 1.

Definition 4.14. A ***pass*** is a quadruple, $P = (K, r, s, t)$ where:

1. K is a configuration.

2. $r \in \mathbb{Z}^+$.

3. s, t are distinct adjacent vertices of $G(K)$.

4. For every vertex $v \in V(G(K))$, there is an s-v path and t-v path in $G(K)$, both of length ≤ 2.

We call s the **source**, t the **sink**, and r the **value** of the pass. In all the 32 rules, r is at most 2.

Let $\mathscr{P}$ be a set of passes. A pass $P \sim \mathscr{P}$ if P is a pass isomorphic to a member of $\mathscr{P}$. A pass P **appears** in a cartwheel W if $K(P)$ appears in W, or in other words, $G(K(P))$ is an induced subgraph of $G(W)$, every finite region of $K(P)$ is a finite region of W (and so the infinite region of $K(P)$ includes the infinite region of W), and $\gamma_{K(P)}(v) = \gamma_W(v)$ for every $v \in (G(K))$.

Define:

$$N_{\mathscr{P}}(W) := 10(6 - \gamma_W(w)) + \sum(r(P) : P \sim \mathscr{P}, P \text{ appears in } W, t(P) = w) \\ - \sum(r(P) : P \sim \mathscr{P}, P \text{ appears in } W, s(P) = w),$$

where w is the hub of W.

Formally, a **rule** is a 6-tuple $(G, \beta, \delta, r, s, t)$ where:

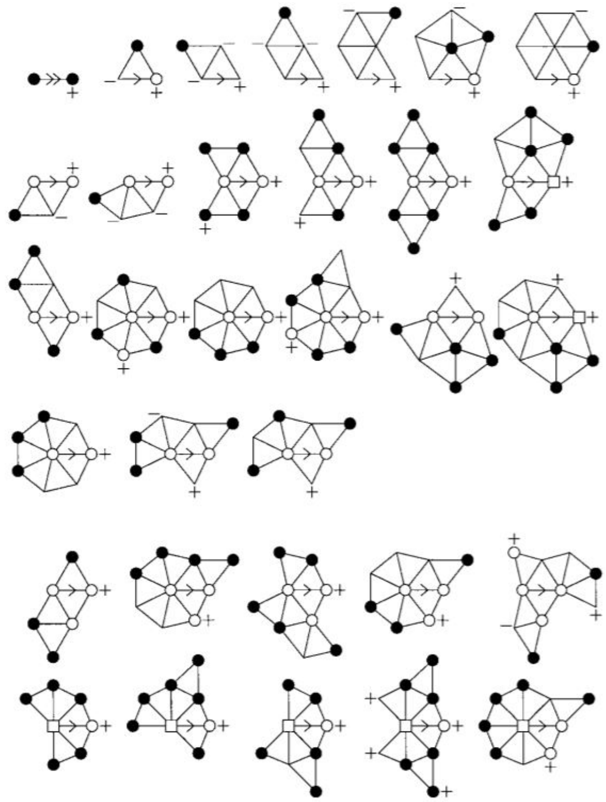
1. G is a near triangulation and $G \setminus v$ is connected for all $v \in G$.

2. $\beta : \mathrm{V}(G) \to \mathbb{Z}^+$ and $\delta : \mathrm{V}(G) \to \mathbb{Z}^+ \cup \{ \infty \}$ such that $\beta(v) \leq \delta(v)$ for all v.

3. $r \in \mathbb{Z}^+$.

4. s and t are distinct, adjacent vertices of G and for every $v \in V(G)$, there exists a v-s path and a v-t path of length ≤ 2 such that $\delta(w) \leq 8$ for the internal vertex w of the path if there is one.

Finally, a pass P **obeys** a rule $(G, \beta, \delta, r, s, t)$ if P is isomorphic to some (K, r, s, t) where $G(K) = G$ and $\beta(v) \leq \gamma_K(v) \leq \delta(v)$ for every $v \in \mathrm{V}(G)$.



The set of 32 discharging rules used by Robertson et all.

The symbols on the vertices adopt the same meaning as was presented above in figure 1. The arrows shown indicate s (tail of arrow), t (head of arrow), and r (number of arrow heads). Also in all cases, the options for $\beta(v)$ and $\delta(v)$ are as follows:

1. $5 \leq \beta(v) = \delta(v) \leq 8$, or
2. $\beta(v) = 5$ and $6 \leq \delta(v) \leq 8$, or
3. $5 \leq \beta(v) \leq 8$ and $\delta(v) = \infty$.

To describe the first case, the same meaning as was presented in figure 1 is used. In the second case, figure 1 conventions along with a minus sign beside a vertex v to indicate $\gamma_G(v) = \delta(v)$. Similarly for the third case, figure 1 conventions are used along with a plus sign beside vertex v to indicate $\gamma_G(v) = \beta(v)$. The labels of $\beta(v)$ and $\gamma(v)$ indicate bounds for what $\gamma_{K(P)}(v)$ must be in order to apply the rule to a pass P. For example, the first rule requires all passes P which obey it to have a source s satisfy $\gamma_{K(P)}(s) = 5$ (as $\beta(s) = \delta(s) = \gamma_G(s)$).

They then showed the following:

Theorem 4.15. Let T be an internally 6-connected triangulation and $\mathscr{P}$ be a set of passes. Then there exists a cartwheel W appearing in T with $N_{\mathscr{P}}(W) > 0$.

Theorem 4.16. For every cartwheel W with $N_{\mathscr{P}}(W) > 0$, some good configuration appears in W.

We will omit the proof of the first theorem and prove the case for cartwheels with hub of degree 5 for the second theorem. The other cases are left as an exercise to the reader (note the attempt at this is not recommended as it took the original authors over 13,000 lines to show).

We will now show:

Theorem 4.17. Let W be a cartwheel with $N_{\mathscr{P}}(W) > 0$ and hub of degree 5. Then a good configuration appears in it.

Before we present the proof we require a lemma:

Lemma 4.18. Let W be a cartwheel with hub w of degree 5. For $k = 1,..., 32$ let p_k (respectively q_k) be the sum of $r(P)$ over all passes P obeying rule k and appearing in W with sink (respectively source) w. Suppose no good configuration appears in W. Then $p_1 = q_2 + q_3$.

Proof. Suppose W is a cartwheel with hub w of degree 5. Let X be the set of all triples (x, y, z) of neighbors of w in W such that x, y, z are all distinct and y is adjacent to x and z with $\gamma(x) = 5$. Then p_1, which is the sum of all passes P obeying rule 1 and appearing in W with sink w, is equal to $|X|$. This is because those passes must have a source s with $\gamma(s) = 5$, and the value of r for rule 1 is 2 which accounts for each neighboring vertex of w with $\gamma(x) = 5$ getting counted twice, as in X. q_2 is the number of triples (x, y, z) in X with $\gamma(y) \geq 7$ and q_3 is the number of triples (x, y, z) in X with $\gamma(y) = 5$ or 6 and $\gamma(z) \geq 6$. So $|X| = q_2 + q_3 + K$ where K is the number of triples (x, y, z) with $\gamma(y) = 5$ or 6 and $\gamma(z) = 5$.

First 2 configurations in the set of good configurations of Robertson et all.

But those possible configurations agreeing with K are good configurations, thus
$|X| = q_2 + q_3 = p_1$. □

We now prove the case for cartwheels with hubs of degree 5.

Proof. Suppose W is a cartwheel with hub w of degree 5 and suppose no good configuration appears in W. Then letting p_k and q_k be as above, we have that $p_1 = q_2 + q_3$.

Observe that $p_2,...,p_{32}$, $q_4,...,q_{32} = 0$ since all those rules require the sink v (respectively source v) to have $\gamma(v) \geq 6$. We have then that:

$$N_{\mathscr{P}}(W) = 10(6 - \gamma_W(w)) + p_1 - q_1 - q_2 - q_3 = 10 - q_1.$$

But $q_1 = 10$ which makes $N_{\mathscr{P}}(W) = 0$, a contradiction. Then a good configuration must appear in W. □

So to recap, Robertson et all. showed that every internally 6-connected triangulation has a cartwheel with $N_{\mathscr{P}}(W) > 0$, and therefore has a good configuration appearing in it. If you recall every minimal counterexample is an internally 6-connected triangulation and thus has a good configuration in it. But at the same time, each of the good configurations are reducible. So no minimal counterexample can exist, thus there is no planar graph that isn't four colourable! That wasn't so hard.

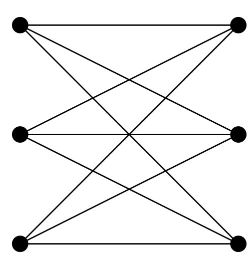
5 Colouring of graphs on higher genus surfaces

All graphs and surfaces in this section are connected unless otherwise stated.

In 1930, Kazimierz Kuratowski gave a necessary and sufficient requirement for a graph to be planar. Consider the following definition-

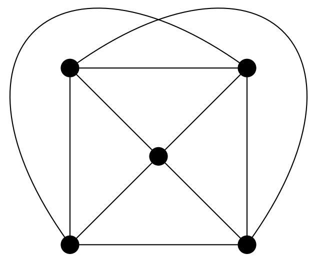
Definition 5.1. A ***subdivision*** H of a graph G is a graph obtained from G with the addition of vertices to the edges of G.

Theorem 5.2 (Kuratowski)**.** A graph is planar if and only if it does not contain a (Kuratowski) subgraph that is a subdivision of K_5 or $K_{3,3}$ where K_5 is the complete graph on 5 vertices and $K_{3,3}$ (pictured below) is the complete bipartite graph on 6 vertices.

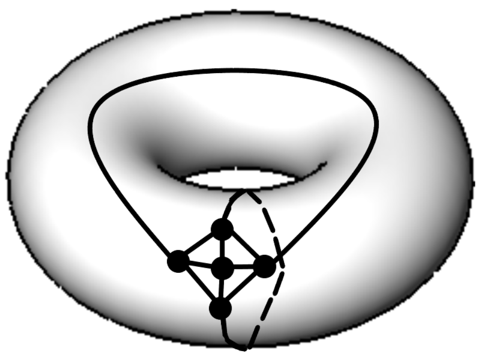


Another equivalent statement to planarity involves the crossing number of a graph. The *crossing number* of G is the smallest possible number of edge crossings of G when it's drawn on the plane. Clearly a graph is planar if and only if its crossing number is 0. Unfortunately, determining the crossing number of a graph is very difficult, and is consequently an NP-hard problem. That is why in practice Kuratowski's requirement is used as a Kuratowski subgraph can be found in polynomial time.

Until now, only planar graphs have been considered, but in fact any graph (recall a graph was defined to be finite) can be embedded. First, recall K_5 which is not planar by lemma 3.4 (since $E \leq 3V - 6$ for all planar graphs, and $E(K_5) = 10$ and $V(K_5) = 5$):



Clearly no matter how it's drawn, it can never be embedded on the plane, consequently it has crossing number 1. But consider K_5 drawn on a torus:



It can be drawn such that no edges cross. Now note, if another handle was added onto the torus, it would not affect the graph's embedding. Consider the following definition:

Definition 5.3. A **2-*cell embedding*** is an embedding in which every face of the graph is homeomorphic to an open disk.

This will be the definition of embedding that will be used from now on. Note this now restricts the genus of the surface that a graph can be embedded in. The **genus** of a graph G, denoted $g(G)$, is the smallest integer z such that G can be embedded on a surface of genus z. So $g(K_5) = 1$. The genus, p, of an orientable surface, S_p, can intuitively be thought of as the number of holes in the surface, or the number of tori glued together. The non-orientable genus, k, of a non-orientable surface N_k can be thought of as the number of projective planes glued together. A topological invariant of surfaces, known as the Euler characteristic, denoted $\chi(S)$, is defined to be:

$$\chi(S) = \begin{cases} 2-2p & \text{for} \quad S = S_p \\ 2-k & \text{for} \quad S = N_k. \end{cases}$$

Now Euler's formula for planar graphs presented in section 2 is actually a special case of the more general formula:

Theorem 5.4. (Euler's formula) Let G be an embedded graph with V, E, and F vertices, edges, and faces respectively, embedded in a surface S with Euler characteristic $\chi(S)$. Then $V - E + F = \chi(S)$.

Unfortunately, finding the genus of a graph may not be an easy task, but once it's found we can easily determine the chromatic number of the graph, thanks to Heawood. In 1890, Heawood found an upper bound for the chromatic number of any graph [5].

Theorem 5.5 (Heawood)**.** Let S be an orientable surface of genus $p > 0$, and suppose G is a graph embeddable on S. Then the chromatic number of G, $\gamma(G)$, is at most $\gamma(G) \leq \gamma'(G)$ where:

$$\gamma'(G) \leq \left\lfloor \frac{7 + \sqrt{1 + 48p}}{2} \right\rfloor,$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Proof. Let G be a graph embedded on a surface of genus $p > 0$ with V, E, and F representing the usual. Let $c = \gamma'(G)$. We can assume $deg(v) \geq c - 1$ for every vertex v of G else we can consider the graph $G \setminus v$ which has the same chromatic number as G. Then we have that $2E \geq (c - 1)V$. From Euler's formula and lemma 3.3 we can also deduce: $E \leq 3V - 6 + 6p$. Therefore we can conclude that $6V - 12 + 12p \geq (c - 1)V$ or equivalently $0 \geq (c - 7)V + 12 - 12p$. Since the RHS of Theorem 6.5 is at least 7, we can assume $c \geq 7$. Then since $V \geq c$ we have that:

$$0 \geq (c - 7)V + 12 - 12p \geq (c - 7)c + 12 - 12p = c^2 - 7c + 12 - 12p.$$

Then solving this quadratic formula yields Theorem 6.5. □

Heawood believed his upper bound was actually tight and was under the false impression he had proven it. It took 78 years until Gerhard Ringel and Ted Youngs finally showed equality completing the last 3 cases of the 12 case proof. There is also a more general version written in terms of the Euler characteristic which also applies to graphs embedded on non-orientable surfaces [8], namely:

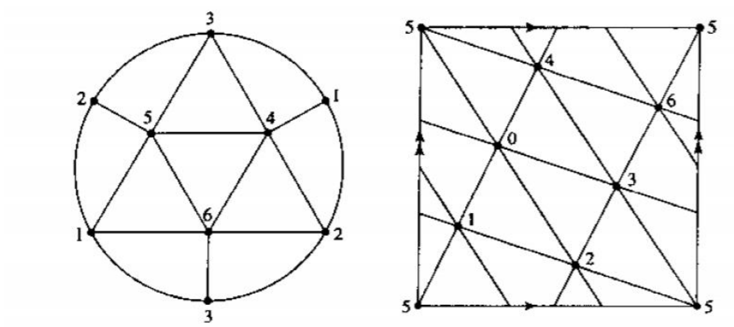
Theorem 5.6. Let S be a surface (other than the Klein bottle) with Euler characteristic χ, and suppose G is a graph embeddable on S. Then the chromatic number of G is at most $\gamma(G) \leq \gamma'(G)$ where:

$$\gamma'(G) = \left\lfloor \frac{7 + \sqrt{49 - 24\chi}}{2} \right\rfloor.$$

Graphs embeddable on the Klein bottle, N_2, are proven to have chromatic number at most 6.

We now have a way of determining a close bound to the chromatic number of any graph given its Euler characteristic. It's interesting that the colouring problem for surfaces of genus 0 was the most difficult to solve. This is due to the fact that to prove the four colour problem, one must show that *every* planar graph is four colourable.

Whereas to prove Theorem 6.6 and equality in Theorem 6.5 one needs to show that we can draw, on a surface with Euler characteristic χ, a graph with chromatic number n where $n = \left\lfloor \frac{7+\sqrt{49-24\chi}}{2} \right\rfloor$. Therefore the proof boils down to showing that K_n (which has chromatic number n) is embeddable on a surface of Euler characteristic χ if $n = \left\lfloor \frac{7+\sqrt{49-24\chi}}{2} \right\rfloor$ [3].



Triangulations of the projective plane, N_1, by K_6 and of the torus, S_1, by K_7 [3]

So in conclusion, the 4-colour problem and general surface colouring problem are really completely different problems, with one being immensely more difficult than the other.

6 Conclusion

The four colour problem is truly one of the greatest hurdles mathematicians have ever overcome. Mathematician William Tutte was quoted saying "It is dangerous to work close to The Problem", and he was right. When Appel and Haken first found their proof they hypothesized that over 10 million human hours had been dedicated to finding a solution [7]. Today, that number has been well surpassed as mathematicians in every field continue to hunt for a more intuitive and intelligible solution. There are still a number of open problems, that if solved would verify the four colour theorem and give insight as to why it should be true. Probably the most famous being Hadwiger's conjecture, made in 1943:

Conjecture 6.1. (Hadwiger) Every connected graph with chromatic number at least n is subcontractible to K_n.

The case with n = 5 is equivalent to the four colour theorem [11]. If there was a planar graph with chromatic number greater than 4, or at least 5, then it would be subcontractible to K_5, but K_5 is not planar and thus cannot be a subcontraction of any planar graph. The conjecture is proven for the cases $1 \leq n \leq 6$ and considered by many to be "one of the deepest unsolved problems in graph theory".

There are generalizations being done in homology theory regarding graphs embedded in 3-space which look to show cubic bridgeless graphs which are embeddable in the plane admit a Tait colouring. This is done by associating to the graph a non-zero finite-dimensional vector space J using homology. It is then conjectured that [6]:

Conjecture 6.2. If an embedded cubic graph K lies in the plane, so $K \subseteq \mathbb{R}^2 \subseteq \mathbb{R}^3$, then the dimension of J is equal to the number of Tait colorings of K, which is non-zero if and only if K is bridgeless.

There are also equivalent statements found that appear to have nothing to do with the four colour theorem. Consider the vector cross product in $\mathbb{R}^3$, which we all know to be non-associative. If an arrangement of $k-1$ brackets is given to k vectors $v_1 \times v_2 \times ... \times v_k$ such that the order of evaluation can be determined, we call this arrangement of brackets an *association*. Consider the following equivalent statement to the four colour theorem [13]:

Theorem 6.3. Let i, j, k be the usual unit vector basis of $\mathbb{R}^3$. If two associations of $v_1 \times v_2 \times ... \times v_k$ are given, there exists an assignment of i, j, k to $v_1, v_2, ..., v_k$ such that the evaluations of the two associations are equal and nonzero.

Hopefully it is clear that the four colour problem is more than just a problem for graph theorists. Like many statements in math, it can be reformulated in different theories to pose problems for topologists, algebraists, and geometers alike.

To this date, many mathematicians remain unsatisfied with the current solutions. Many of the proofs still require computer aid and none have quite provided the satisfying intuitive explanation that formal proofs are meant to deliver. It will be interesting to see in the years to come what improvements will be made, but in the meantime, the four colour problem will continue tallying up the hours its stolen from mankind.

References

[1] Kenneth Appel and Wolfgang Haken. "Every planar map is four colorable. Part I: Discharging". In: *Illinois Journal of Mathematics* 21 (1977), pp. 429–490. DOI: 10.1215/ijm/1256049011.

[2] George D. Birkhoff. "The Reducibility of Maps". In: *American Journal of Mathematics* 35 (1913), pp. 115–128. DOI: 10.2307/2370276.

[3] Bela Bollobas. *Modern Graph Theory*. Vol. 184. Graduate Texts in Mathematics. Springer-Verlag New York, 1998. ISBN: 9781461206194.

[4] L.I Golovina and I.M Yaglom. *Induction in Geometry*. Mir Publishers, 1979. ISBN: 9780486838564.

[5] Frank Harary. *Graph Theory*. Addison-Wesley Publishing Company, 1969. ISBN: 9780201410334.

[6] Peter Kronheimer and Tomasz S. Mrowka. "Tait colorings, and an instanton homology for webs and foams". In: *Journal of the European Mathematical Society* (2015).

[7] Donald MacKenzie. "Slaying the Kraken: The Sociohistory of a Mathematical Proof". In: *Social Studies of Science* 29 (1999), pp. 7–60.

[8] Bojan Mohar and Carsten Thomassen. *Graphs on Surfaces*. Johns Hopkins University Press, 2001. ISBN: 9780801866890.

[9] Oystein Ore. *The Four-Color Problem*. Vol. 27. Pure and Applied Mathematics. Academic Press, 1967. ISBN: 9780123745712.

[10] Neil Robertson et al. "The Four-Colour Theorem". In: *Journal of Combinatorial Theory, Series B* 70 (1997), pp. 2–44.

[11] Thomas L. Saaty. "Thirteen Colorful Variations on Guthrie's Four-Color Conjecture". In: *The American Mathematical Monthly* 79.1 (1972), pp. 2–43. DOI: 10.1080/00029890.1972.11992979.

[12] John P. Steinberger. "An Unavoidable Set of D-Reducible Configurations". In: *Transactions of the American Mathematical Society* 362 (2010), pp. 6633–6661.

[13] Robin Thomas. "An Update on the Four-Color Theorem". In: *Notices of the American Mathematical Society* 45.7 (1998), pp. 848–859.