On the Four Color Theorem

Zdeněk Dvořák

December 7, 2021

It will be convenient to use the following correspondence between 4-coloring and 3-edge-coloring in planar triangulations (we will see a proof of this claim later in the lecture on nowhere-zero flows).

Lemma 1. *A plane triangulation is 4-colorable if and only if its dual (a plane 2-edge-connected 3-regular graph) is 3-edge-colorable.*

Hence, the Four Color Theorem is equivalent to the following claim.

Theorem 2 (The Four Color Theorem, edge coloring version)**.** *Every planar 2-edge-connected 3-regular graph is 3-edge-colorable.*

In the reducibility part of the proof of the Four Color Theorem, we will work in the edge coloring setting. Let G be a plane 2-edge-connected 3-regular graph and let K be a cycle in the dual of G; then K corresponds to a simple closed curve that intersects G only in edges. Let Δ be the closed interior or exterior of K, and let $-\Delta$ be the closure of the complement of Δ (the closed exterior or interior of K). We define G_Δ to be the graph obtained from G by adding vertices on the intersections of edges of G with K and then deleting vertices and edges not contained in Δ; hence, G_Δ is drawn in Δ, its vertices drawn in the interior of Δ have degree three, and its vertices drawn on the boundary of Δ have degree 1. We fix a starting point and an orientation of the boundary of Δ arbitrarily (and set them for $-\Delta$ in the same way). Let e_1, ..., e_k be the edges of G_Δ incident with vertices of degree 1, listed in their order around the boundary of Δ (starting from the fixed starting point). A *k-tuple coloring* is a function $\psi : \{1, \ldots, k\} \to \{1, 2, 3\}$. It is *parity-compliant* if for each $c \in \{1, 2, 3\}$, we have $|\psi^{-1}(c)| \equiv k \pmod 2$. Let $\mathcal{C}_k$ denote the set of all parity-compliant k-tuple colorings. Let $S(G, \Delta)$ be the set of k-tuple colorings ψ such that G_Δ has a 3-edge-coloring in which the color of e_i is $\psi(i)$ for $i = 1, \ldots, k$. A *minimal counterexample* is a planar 2-edge-connected 3-regular graph that is not 3-edge-colorable and has the smallest number of vertices among the graphs with this property.

Observation 3. *Let G, Δ and k be as above.*

- *Each $\psi \in S(G, \Delta)$ is parity-compliant.*
- *G is 3-edge-colorable iff $S(G, \Delta) \cap S(G, -\Delta) \neq \emptyset$.*
- *If G is a minimal counterexample and $G_{-\Delta}$ is not a tree, then $S(G, \Delta) \neq \emptyset$.*

A matching M whose vertices are integers is *planar* if we can draw its vertices on the boundary of a disk Λ in increasing order and drawn the edges inside Λ without crossings. For a parity-compliant k-tuple coloring ψ and a pair of colors $\{c_1, c_2\} \in \binom{\{1,2,3\}}{2}$, a *$\{c_1, c_2\}$-matching* is any perfect matching M with vertex set $\psi^{-1}(c_1) \cup \psi^{-1}(c_2)$; note that this set has even size, since ψ is parity-compliant. We say that a k-tuple-coloring ψ' is *obtained from ψ by switching on M* if there exists a matching $M' \subseteq M$ such that for $i = 1, \ldots, k$,

$$\psi'(i) = \begin{cases} \psi(i) & \text{if } i \in \{1, \ldots, k\} \setminus V(M') \\ c_1 & \text{if } i \in V(M') \text{ and } \psi(i) = c_2 \\ c_2 & \text{if } i \in V(M') \text{ and } \psi(i) = c_1. \end{cases}$$

Consider a set S of parity-compliant k-tuple colorings. We say that $\psi \in S$ is *S-consistent* if for every $C \in \binom{\{1,2,3\}}{2}$, there exists a planar C-matching M such that all matchings obtained from ψ by switching on M belong to S. We say that S is *consistent* if all elements of S are S-consistent.

Lemma 4. *If G, Δ and k are as above, then $S(G, \Delta)$ and $S(G, -\Delta)$ are consistent.*

Observe that if S_1 and S_2 are consistent, then so is $S_1 \cup S_2$. Hence, for every $X \subseteq \mathcal{C}_k$, there exists a unique maximal subset of X that is consistent; we denote it by $C(X)$.

Observation 5. *$C(X)$ is obtained from X by iteratively removing elements that are not X-consistent (in any order).*

By a string of k letters a, b, and c, we express all k-tuple colorings obtained by interpreting these letters as distinct colors. E.g., by abca, we mean the set of colorings $\{(1,2,3,1), (1,3,2,1), (2,1,3,2), (2,3,1,2), (3,1,2,3), (3,2,1,3)\}$. Note that if S is a consistent set of parity-compliant k-tuple colorings, it is closed under permutation of colors, i.e., if $\psi \in S$, then $\pi \circ \psi \in S$ for any permutation π of $\{1, 2, 3\}$.

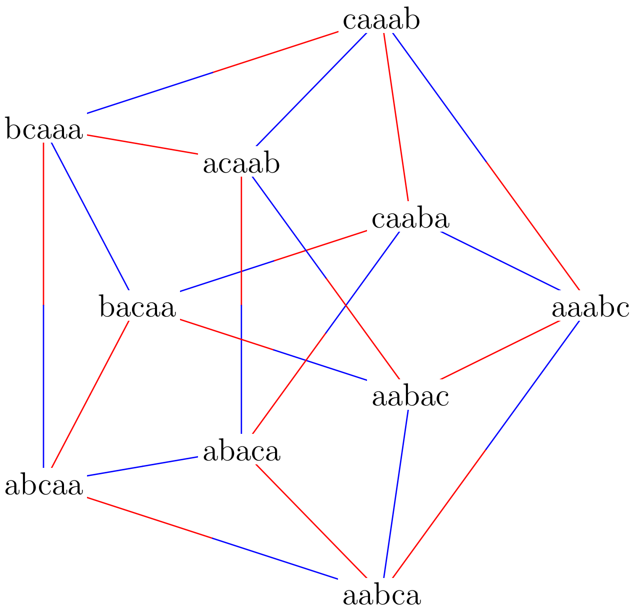
An *edge-cut* of a graph G is a partition (A, B) of its vertex set, the size of the edge-cut is the number of edges with one end in A and the other end in B. The edge-cut is *cyclic* if neither $G[A]$ nor $G[B]$ is a forest.

Lemma 6. *A minimal counterexample G has no cyclic edge-cut of size at most four. Moreover, if (A, B) is a cyclic edge-cut in G of size five, then $G[A]$ or $G[B]$ is a 5-cycle.*

Proof. Let (A, B) be a cyclic edge-cut of G of minimum size $c \leq 5$. Note that $c \geq 2$, since G is 2-edge-connected. We will give the argument for the cases $c = 4$ and $c = 5$, the arguments for the cases $c = 2$ and $c = 3$ are similar (and simpler). Note that in the dual of G, the edges between A and B form a cycle K of length c; let Δ be the closed disk bounded by K. Without loss of generality, we can assume $G[A]$ is drawn in Δ and $G[B]$ is drawn in $-\Delta$.

Suppose $c = 4$. Note that $\mathcal{C}_4 = \mathrm{aaaa} \cup \mathrm{aabb} \cup \mathrm{abba} \cup \mathrm{abab}$. By Observation 3, $S(G, \Delta) \neq \emptyset$. Note that none of the sets aaaa, ..., abab is consistent, and that sets $\mathrm{aaaa} \cup \mathrm{abab}$ and $\mathrm{aabb} \cup \mathrm{abba}$ are not consistent. In particular, since $S(G, \Delta)$ is consistent by Observation 4, $S(G, \Delta)$ is a union of at least two of the sets. The same holds for $S(G, -\Delta)$. Moreover, $S(G, \Delta) \cap S(G, -\Delta) = \emptyset$ by Observation 3. By symmetry, we can assume $S(G, \Delta) = \mathrm{aaaa} \cup \mathrm{aabb}$ and $S(G, -\Delta) = \mathrm{abba} \cup \mathrm{abab}$. Let $v_1, \ldots, v_4$ be the vertices of G_Δ of degree one in order around the boundary of Δ. Let G' be the graph obtained from G_Δ by identifying v_1 with v_2 to a new vertex z_1, v_3 with v_4 to a new vertex z_3, and adding the edge z_1z_3. Then G' is planar cubic graph. Moreover, G' is 2-edge-connected. Indeed, if z_1z_2 formed an edge-cut, then G would contain a (necessarily cyclic) edge cut of size two consisting of the edges corresponding to v_1 and v_2, i.e., a cyclic edge-cut smaller than (A, B); and the other possible 1-edge-cuts are also easy to exclude. By the minimality of G, G' has a 3-edge-coloring. However, observe that no coloring in $S(G, \Delta)$ extends to a 3-edge-coloring of G', which is a contradiction.

Suppose $c = 5$ and for contradiction assume that neither $G[A]$ nor $G[B]$ is a 5-cycle. Note that $\mathcal{C}_5$ consists of aaabc, aabac, and their rotations. Consider a consistent subset S of $\mathcal{C}_5$. There are two planar $\{a, b\}$-matchings for a coloring aaabc, M_1 containing edges 12 and 34, and M_2 containing 14 and 23. Switching on M_1 transforms aaabc to aabac (and itself), switching on M_2 transforms aaabc to baaac (and itself). Hence, since S is consistent, we conclude that if S contains aaabc, then it also contains aabac or baaac. Similarly, we can consider planar $\{a, c\}$-matchings and conclude that S contains caaba or aacba. Repeating this idea for all other colorings in $\mathcal{C}_5$, we obtain the following picture, which should be read as follows: If a consistent set contains a coloring ψ, then it also necessarily contains at least one of the two colorings joined to it by an edge starting red and at least one joined by an edge starting blue.



In particular, $S(G, \Delta)$ and $S(G, -\Delta)$ induce in the graph depicted above disjoint subgraphs of minimum degree two.

Consider the graph G' obtained from G_Δ by deleting v_1 and v_2 and adding an edge between their neighbors, and by identifying v_3, v_4, and v_5 to a single vertex. Note that $|V(G')| < |V(G)|$, and moreover, observe that G' is 2-edge-connected. By the minimality of G, the graph G' is 3-edge-colorable, and we conclude that $(\text{aaabc} \cup \text{aabca} \cup \text{aabac}) \cap S(G, \Delta) \neq \emptyset$. By symmetry, this argument shows that both $S(G, \Delta)$ and $S(G, -\Delta)$ intersect all five triangles in the picture above.

Suppose that say aaabc $\subseteq S(G, \Delta)$ and caaab $\subseteq S(G, -\Delta)$. Then the neighbor of aaabc over the red edge belonging to $S(G, \Delta)$ is aabac, and the neighbor of caaab over the blue edge belonging to $S(G, -\Delta)$ is acaab. Since $S(G, \Delta)$ and $S(G, -\Delta)$ intersect all triangles, we have bcaaa $\subseteq S(G, \Delta)$ and aabca $\subseteq S(G, -\Delta)$. However, since aaabc and caaab have minimum degree two in the subgraphs induced by $S(G, \Delta)$ and $S(G, -\Delta)$, we conclude that caaba $\subseteq S(G, \Delta) \cap S(G, -\Delta)$, which is a contradiction. It follows that no edge of the outer cycle of the depicted graph joins a coloring from $S(G, \Delta)$ with a coloring of $S(G, -\Delta)$.

If both $S(G, \Delta)$ and $S(G, -\Delta)$ intersect the outer cycle, we can by symmetry assume that aaabc $\subseteq S(G, \Delta)$ and that caaab and aabca belong to neither $S(G, \Delta)$ nor $S(G, -\Delta)$. Since the subgraph induced by $S(G, \Delta)$ has minimum degree at least two, caaba and aabac belong to $S(G, \Delta)$. But then $S(G, -\Delta)$ intersects neither of the triangles containing aaabc, which is a contradiction.

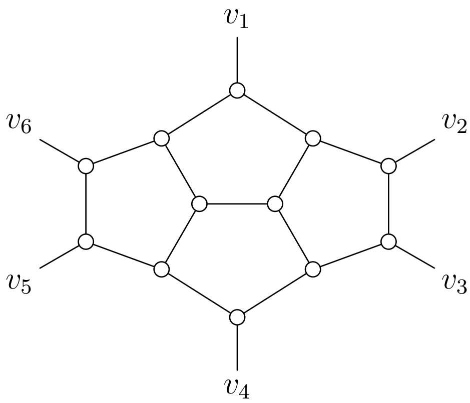
Hence, we can assume that $S(G, \Delta)$ is disjoint from the outer cycle. Let G'' be the graph obtained from G_Δ by adding the 5-cycle $v_1 \dots v_5$. Observe that G' is 2-edge-connected and has no 3-edge-coloring, contradicting the minimality of G (since $G[A]$ is not the 5-cycle). $\square$

The following is a key lemma that is used to prove reducibility in the proof of the Four Color Theorem. It is applied in the situation where G_Δ is a fixed configuration that we want to exclude, and G'_Δ is a fixed replacement graph.

Lemma 7. *Let G be a minimal counterexample and let Δ be the closed interior or exterior of a cycle K in the dual of G. Let G' be a planar 2-edge-connected 3-regular graph obtained from G by replacing G_Δ by a smaller graph; i.e., K is also a cycle in the dual of G', $G'_{-\Delta} = G_{-\Delta}$, and $|V(G')| < |V(G)|$. Then $S_{G',\Delta} \cap C(\mathcal{C}_{|K|} \setminus S_{G,\Delta}) \neq \emptyset$.*

Proof. By Observations 3 and 4, we have $S(G', -\Delta) = S(G, -\Delta) \subseteq C(\mathcal{C}_{|K|} \setminus S_{G,\Delta})$. By the minimality of G, the graph G' is 3-edge-colorable, and by Observation 3, $S_{G',\Delta} \cap S(G', -\Delta) \neq \emptyset$. $\square$

Let us now illustrate how this lemma is applied to deal with the smallest non-trivial reducible configuration, for which the graph G_Δ is depicted in the following picture.



An inspection of the 3-edge-colorings of this configuration shows that

$$\begin{aligned} S(G,\Delta) =\, & aaabab, aaabba, aabaab, aabbaa, aabbbb, aabccb, abaaba, ababaa,\\ & ababbb, ababcc, abbabb, abbacc, abbbab, abbbba, abbcac, abccba, \end{aligned}$$

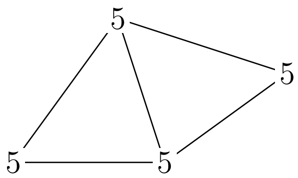
and thus

$$\begin{aligned} \mathcal{C}_6 \setminus S(G,\Delta) =\, & aaaaaa, aaaabb, aababa, aabbcc, aabcbc, abaaab, abacbc, abaccb\\ & abbaaa, abbcca, abcabc, abcacb, abcbac, abcbca, abccab \end{aligned}$$

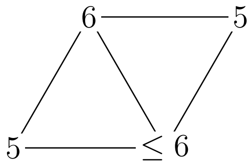
We now use Lemma 5. We can check that aaaaaa is not consistent with this set, by going over all planar $\{a, b\}$-matchings, and checking that for each of them, we can switch aaaaaa to a coloring not belonging to the set. Hence, we can delete aaaaaa, and check whether other colorings are consistent with the rest of the set. Iteratively deleting inconsistent colorings, we eventually conclude that $C(\mathcal{C}_6 \setminus S_{G,\Delta}) = \emptyset$. Let G' be obtained from $G_{-\Delta}$ by adding the 6-cycle $v_1 \dots v_6$. It is easy to see that G' is 2-edge-connected. However, $S(G', \Delta) \cap C(\mathcal{C}_6 \setminus S_{G,\Delta}) \subseteq C(\mathcal{C}_6 \setminus S_{G,\Delta}) = \emptyset$, contradicting Lemma 7.

In total, to prove the Four Color Theorem, one needs to exclude more than 600 such configurations (in some of them, $C(\mathcal{C}_{|K|} \setminus S_{G,\Delta})$ is non-empty, and in these cases, one needs to choose the reduction G' carefully so that $S(G', \Delta) \cap C(\mathcal{C}_6 \setminus S_{G,\Delta}) = \emptyset$ to obtain a contradiction; Lemma 6 is useful in showing G' is 2-edge-connected). For illustrating the discharging rules, we will use the following reducible configurations (we draw their duals, with the numbers indicating the degrees of the vertices; for example, the configuration depicted above corresponds to the first configuration in the list below):

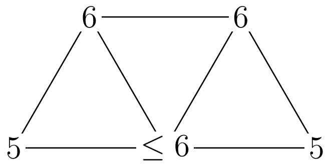
(C1)



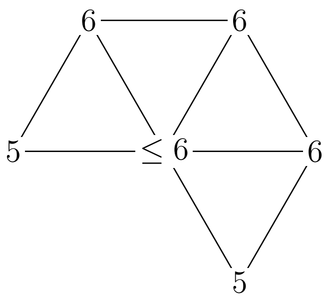
(C2)



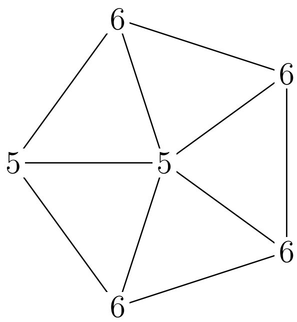
(C3)



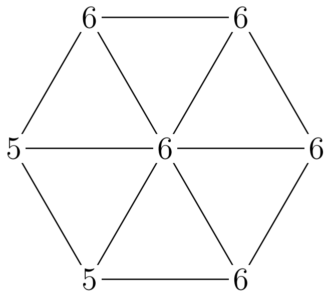
(C4)



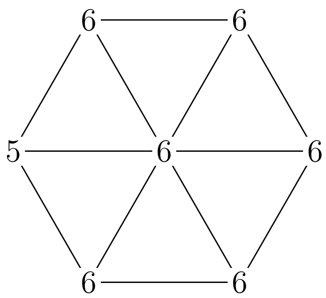
(C5)



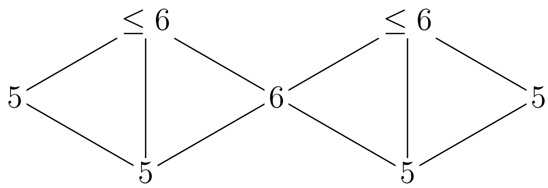
(C6)



(C7)



(C8)



We now switch to the primal setting, so now a minimum counterexample G is a non-4-colorable planar triangulation with the smallest number of vertices. As we have argued above, G contains none of the configurations (C1), …, (C8). Let us give each vertex v the initial charge $10\deg v - 60$, so that the sum of the charges is -120. Next, let us redistribute the charge according to the following rules (also defining over which edge the charge is sent, to help with accounting):

(R1) If a vertex v of degree at least seven has a neighbor of degree five, then v sends 2 to u over the edge vu.

(R2) If a vertex z of degree at most six has distinct neighbors v, x_1, …, x_k, u in order, where $0 \le k \le \deg z - 2$, v has degree at least 7, x_1, …, x_k have degree six, and u has degree five, then v sends 2 to u if $\deg z = 5$

and u and z have a common neighbor of degree five, and v sends 1 to u otherwise. The charge is sent over the edge vx_1 if $k \geq 1$ and over the edge vu if $k = 0$.

(R3) If uvz is a triangle with $\deg v \geq 7$, $\deg u = 5$ and $\deg v = 6$, then v sends 1 to u over the edge vz.

In applications of (R2) and (R3), we say that z is the *pivot*.

Lemma 8. *After performing these discharging rules, each vertex of degree* 5, 6, *or at least* 12 *has non-negative final charge.*

Proof. Note that the rules do not change the charge of vertices of degree six, and thus the final charge of these vertices is 0.

Consider now a vertex v of degree at least 7, and let us count the amount of charge sent over an edge vx incident with v. Note that the rule (R1) applies at most once, rule (R2) applies at most twice (with the pivots being the common neighbors of v and x), and rule (R3) also applies at most twice. If $\deg x = 5$, then only rules (R1) and (R2) can apply; if v sent charge to x by (R1) and the rule (R2) applied on vx twice, both times sending charge 2, then G would contain (C1). If $\deg x = 6$, then only rules (R2) and (R3) can apply; if both applied twice and both applications of (R2) sent charge 2, then G would contain (C8). Hence, v sends charge at most 5 over each incident edge, implying that its final charge is at least $10 \deg v - 60 - 5 \deg v = 5(\deg v - 12)$. Hence, if $\deg v \geq 12$, then the final charge of v is non-negative.

Let us now consider a vertex u of degree five, and let y_1, …, y_5 be its neighbors in order. Let s_i denote the amount of charge sent from y_i to u by (R1) plus the amount of charge sent to u by the applications of (R2) and (R3) with y_i being the pivot. By symmetry, we can assume that $s_2 \leq s_1, s_3, s_4, s_5$. If $s_1 + \ldots + s_5 \geq 10$, then the final charge of u is non-negative. Hence, suppose that $s_1 + \ldots + s_5 < 10$, and thus $s_2 \leq 1$.

In particular, y_2 does not send charge to u by (R1), and thus $\deg y_2 \leq 6$. Since G does not contain (C1) and (C2), we can by symmetry assume $\deg y_1 \geq 6$. Let u, y_1, x_{k-1}, …, x_1, v be neighbors of y_2 in order, where v is the first neighbor whose degree is not six (possibly $v = y_1$). Since $\deg y_1 \geq 6$ and G does not contain (C2), …, (C7), we have $\deg v \geq 7$, and the rule (R2) applies. If $\deg y_3 \geq 6$, then (R2) would symmetrically apply from the other side and we would have $s_2 \geq 2$; hence, $\deg y_3 = 5$. Since u only receives one unit of charge by (R2), we have $\deg y_2 = 6$. Since (R3) does not apply with pivot y_2, we have $\deg y_1 = 6$. Let $a \in \{3, 4, 5\}$ be maximum such that $\deg y_a \neq 6$. Since G does not contain (C3), (C4), and (C5), we have $\deg y_a \geq 7$. Note that (R3) applies twice and (R3) applies once with pivot

y_{a+1} (where $y_6 = y_1$), and thus $s_{a+1} = 3$. Observe that u does not have consecutive neighbors of degrees 5, 6, and 6 distinct from y_3, y_2, and y_1, and thus $s_i \geq 2$ for $i \neq 2$. Therefore, $s_1 + \ldots + s_5 \geq 1 + 3 + 3 \times 2 = 10$. $\square$

Corollary 9. *A minimun counterexample to the Four Color Theorem contains a vertex of degree 7, 8, 9, 10, or 11.*

One can finish the proof by introducing additional rules to deal with vertices of larger degrees (there are around 30 rules in total), making the final charge of all vertices non-negative (this involves case analysis around vertices of degree at most 11, performed by computer).