## Colourability

Vassos Hadzilacos

*Colourability* is the following decision problem, abbreviated COL:

**Instance:** (G, k), where G = (V, E) is an undirected graph and  $k \in \mathbb{Z}^+$ .

**Question:** Is there a function  $f: V \to [0..k-1]$  such that if  $\{u, v\} \in E$ ,  $f(u) \neq f(v)$ . Such a function is called a *colouring* of *G*, and if it exists we say that *G* is *k*-colourable and that it has a *k*-colouring.

As the name suggests, we think of the numbers assigned to the nodes by the function f as "colours", and the colouring requirement is that adjacent nodes be assigned different colours.

Theorem 10.1 COL is NP-complete.

PROOF. It is straightforward to show that  $\text{COL} \in \mathbf{NP}$ : A nondeterministic Turing machine can, in polynomial time, "guess" a sequence of pairs (u, i), where  $u \in V$  and  $i \in [0..k - 1]$  and then check that (a) there is such a pair for each node u of G and (b) adjacent nodes of G have different "colours".

We prove that COL is **NP**-hard by showing that  $3SAT \leq_m^p COL$ . Given a 3-CNF formula F we show how to construct, in polynomial time, a graph G and a positive integer k such that

F is satisfiable if and only if G has a k-colouring. (\*)

Let  $F = C_1 \wedge C_2 \wedge \cdots \wedge C_m$  be a 3-CNF formula with variables  $x_1, x_2, \ldots, x_n$ . For each  $j \in [1..m]$ ,  $C_j = \ell_j^1 \vee \ell_j^2 \vee \ell_j^3$ , where each  $\ell_j^t$  is either  $x_i$  or  $\overline{x}_i$  for some  $i \in [1..n]$ . Without loss of generality we assume that  $n \geq 4$ .<sup>1</sup>

The instance of COL that we construct from F consists of the following graph G = (V, E) and k = n+1. (Recall that n is the number of variables and m is the number if clauses in F.) We group the nodes and edges according to their purpose, which is to simulate certain aspects of the formula.

There is a useful abuse of notation here, as we use  $C_j$  to denote both a clause of F and a node of G, and  $x_i$  and  $\overline{x}_i$  to denote both literals of F and nodes of G.

Let us first examine the time needed to construct G and k from F. |V| = 3n + m and  $|E| = \Theta(n^2) + \Theta(n) + \Theta(n^2) + \Theta(mn) = \Theta(n^2 + mn)$ . So the sizes of G and k are polynomial in the size of F, and they can be constructed from it in polynomial time. It remains to prove (\*).

[ONLY IF] Suppose F is satisfiable, and let  $\tau$  be a truth assignment that satisfies it. Then for each  $j \in [1..m]$  there is some  $t_j \in [1..3]$  such that  $\tau(\ell_j^{t_j}) = 1$ . Define the function  $f: V \to [0..n]$  as follows:

<sup>&</sup>lt;sup>1</sup>We can justify this assumption in various ways: We can add new clauses with new variables, say  $(x \lor x \lor x)$ , until we have enough variables; the resulting formula is obviously satisfiable if and only if F is (the new clauses are trivially satisfiable without affecting the original ones). Alternatively we can observe that if  $n \le 3$  there are at most eight truth assignments, so we can determine in polynomial time if F is satisfiable and accordingly map it to a graph that is or is not k-colourable.

- (1)  $f(v_i) = i.$
- (2) If there is some  $j \in [1..m]$  and  $i \in [1..n]$  such that  $\ell_j^{t_j} = x_i$  then  $f(x_i) = i$ ,  $f(\overline{x}_i) = 0$ , and  $f(C_j) = i$ .
- (3) If there is some  $j \in [1..m]$  and  $i \in [1..n]$  such that  $\ell_j^{t_j} = \overline{x}_i$  then  $f(x_i) = 0$ ,  $f(\overline{x}_i) = i$ , and  $f(C_j) = i$ .
- (4) If, for every  $j \in [1..m]$  and  $i \in [1..n]$ ,  $\ell_j^{t_j} \notin \{x_i, \overline{x}_i\}$  then arbitrarily set one of  $f(x_i)$  and  $f(\overline{x}_i)$  to i and the other to 0.

First note that f is well-defined: Rules (2) and (3) do not result in contradictory definitions for  $f(x_i)$ and  $f(\overline{x}_i)$ , because for this to happen there would have to be  $j \neq j' \in [1..m]$  such that for some  $i \in [1..n]$ ,  $\ell_j^{t_j} = x_i$  and  $\ell_{j'}^{t_{j'}} = \overline{x}_i$ ; this is impossible because then  $\tau$  cannot satisfy both  $x_i$  and  $\overline{x}_i$ .

We can now verify that f is a valid (n+1)-colouring of G: By (1), all palette nodes get different colours, so f respects edges of type I. By (2), (3), and (4), each pair of literal nodes,  $x_i$  and  $\overline{x}_i$ , get different colours, so f respects edges of type II. Since the literal nodes  $x_i$  and  $\overline{x}_i$ , one of which is coloured i and the other 0, are only connected to palette nodes  $v_{i'}$ , for  $i' \neq i$ , and each  $v_{i'}$  is coloured i', f respects edges of type III. By (2) and (3), clause node  $C_j$  has different colour than every literal node except  $\ell_j^{t_j}$ , where  $\ell_j^{t_j}$  is a literal that satisfies clause  $C_j$  under  $\tau$ . By the definition of type IV edges, there is no edge between  $C_j$  and  $\ell_j^{t_j}$ , so f respects type IV edges.

[IF] Suppose G is (n+1)-colourable, and let f be an (n+1)-colouring of G. The type I edges form a clique among the n palette nodes. Thus we need n distinct colours for these nodes. Without loss of generality, let i be the colour of  $v_i$ .

Type III edges imply that for each  $i \in [1..n]$ , literal nodes  $x_i$  and  $\overline{x}_i$  cannot be coloured with any of the colours [1..n] except *i*. Furthermore, type II edges require that these nodes have different colours, since one of them is coloured *i* the other must be coloured 0. Therefore, for every  $i \in [1..n]$ ,

one of  $x_i$  and  $\overline{x}_i$  is coloured *i* and the other is coloured 0. (†)

We now define a truth assignment  $\tau$  that satisfies F:

$$\tau(x_i) = \begin{cases} 1, & \text{if } f(x_i) = i \\ 0, & \text{otherwise} - \text{i.e., by } (\dagger), f(x_i) = 0. \end{cases}$$
(\$\$

Consider any clause  $C_j$ ,  $j \in [1..m]$ . We claim that  $\tau$  satisfies at least one of the three literals of  $C_j$ . The clause node  $C_j$  is connected by type IV edges to every literal node except the three that correspond to the literals in clause  $C_j$ . Since  $n \geq 4$ , there is at least one variable  $x_{i'}$ ,  $i' \in [1..n]$ , that is not involved in the three literals of  $C_j$ , so, by ( $\dagger$ ), node  $C_j$  is not coloured 0 and it is not coloured i' for any variable  $x_{i'}$  that does not appear in its literals. So it must be coloured by one of the colours assigned to the three literal node and i be the colour assigned to it and to  $C_j$ , so  $\ell_j^t$  is either  $x_i$  or  $\overline{x}_i$ . We consider these two cases:

CASE 1. 
$$\ell_j^t = x_i$$
. Then  $f(x_i) = f(C_j) = i$  and so, by  $(\ddagger), \tau(x_i) = 1$ . Thus,  $\tau$  satisfies clause  $C_j$ .

CASE 2.  $\ell_j^t = \overline{x}_i$ . Then  $f(\overline{x}_i) = f(C_j) = i$ . By (†),  $f(x_i) = 0$ ; by (‡),  $\tau(x_i) = 0$  and so  $\tau(\overline{x}_i) = 1$ . Thus,  $\tau$  satisfies clause  $C_j$ .

Therefore  $\tau$  satisfies  $C_j$ , for every  $j \in [1..m]$ , and so it satisfies F.

It turns out that even the 3-colourability problem (fixing the parameter k to 3) is **NP**-complete. On the other hand, 2-colourability can be decided in polynomial (in fact linear) time using breadth- or depth-first search: A graph is 2-colourable if and only if it is bipartite.