Colourability

Vassos Hadzilacos

Colourability is the following decision problem, abbreviated COL:

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

Question: Does there exist a function $f : V \to [0..k-1]$ such that for every $\{u, v\} \in E$, $f(u) \neq f(v)$. 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 distinct colours.

Theorem 10.1 COL *is **NP**-complete.*

PROOF. It is straightforward to show that COL $\in$ **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 $\leq_m^p$ 3DM. Given a 3-CNF formula F we show how to construct, in polynomial time, a graph G and a positive integer k such that

$$F \text{ is satisfiable if and only if } G \text{ has a } k\text{-colouring.} \qquad (*)$$

Let $F = C_1 \wedge C_2 \wedge \cdots \wedge C_m$ be a 3-CNF formula with variables $x_1, x_2, \ldots, x_n$. Each clause C_j of F is a disjunction of three literals x_i or $\overline{x}_i$ for some $i \in [1..n]$. Without loss of generality we assume that $n \geq 4$.[1]

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.

$$V = \underbrace{\{v_1, \ldots, v_n\}}_{\text{“palette” nodes}} \cup \underbrace{\{x_1, \ldots, x_n\} \cup \{\overline{x}_1, \ldots, \overline{x}_n\}}_{\text{“literal” nodes}} \cup \underbrace{\{C_1, \ldots, C_m\}}_{\text{“clause” nodes}}$$

$$\begin{aligned}
E = &\big\{\{v_i, v_{i'}\} : i, i' \in [1..n] \text{ and } i \neq i'\big\} && \longleftarrow \text{ type I edges}\\
\cup &\big\{\{x_i, \overline{x}_i\} : i \in [1..n]\big\} && \longleftarrow \text{ type II edges}\\
\cup &\big\{\{x_i, v_{i'}\} : i, i' \in [1..n] \text{ and } i \neq i'\big\} \cup \big\{\{\overline{x}_i, v_{i'}\} : i, i' \in [1..n] \text{ and } i \neq i'\big\} && \longleftarrow \text{ type III edges}\\
\cup &\big\{\{C_j, x_i\} : x_i \text{ is \underline{not} a literal in } C_j\big\} \cup \big\{\{C_j, \overline{x}_i\} : \overline{x}_i \text{ is \underline{not} a literal in } C_j\big\} && \longleftarrow \text{ type IV edges}
\end{aligned}$$

Let us first examine the time needed to construct G and k from F. We have

$$\begin{aligned}
|V| &= 3n + m\\
|E| &= \Theta(n^2) + \Theta(n) + \Theta(n^2) + \Theta(mn) = \Theta(n^2 + mn)
\end{aligned}$$

[1] We can justify this assumption in various ways: We can add new clauses with new variables, say $(x \vee x \vee x)$, 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 $n \leq 3$ 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.

and so the size 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 $(*)$. First we show that $n+1$ colours are necessary and sufficient for a colouring of the palette and literal nodes, even ignoring (for now) the clause nodes C_j and their associated edges (type IV).

- 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. So we need a new colour, which we will call 0. Therefore, for every $i \in [1..n]$,

  $$\text{one of } x_i \text{ and } \overline{x}_i \text{ is coloured } i \text{ and the other is coloured } 0. \tag{†}$$

  The name of the new colour is suggestive of he intuition that, in a truth assignment that satisfies F, the literal that corresponds to a node coloured 0 is false.

Thus, the $n+1$ colours $[0..n]$ are necessary and sufficient for the palette and literal nodes. Given the colouring of the palette and literal nodes with these colours that was described above, we now turn to the colouring of the clause edges C_j, $j \in [1..m]$.

- Since the clause C_j has three literals, by the assumption that $n \geq 4$, there is some variable x_i so the C_j is adjacent to both x_i and $\overline{x}_i$ through type IV edges. Therefore, by (†),

  $$C_j \text{ cannot be coloured } 0. \tag{‡}$$

- Because of type IV edges, C_j is adjacent to every literal node except the three that correspond to the literals of clause C_j. So,

  C_j can be assigned a colour other than 0 if and only if at least one of the literal nodes that correspond to the three literals of clause C_j is assigned a colour other than 0. (§)

By (‡) and (§), G is $(n+1)$-colourable if and only if, for each $j \in [1..m]$, clause C_j has a literal whose associated node has a colour other than 0. Viewing colouring node x_i with i as setting the corresponding variable to true, and colouring it with 0 as setting the variable to false, this means that G is $(n+1)$-colourable if and only if some truth assignment satisfies every clause — i.e., if and only if F is satisfiable. □

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.