Complexity of Coloring Graphs without Forbidden Induced Subgraphs

Daniel Král'[1], Jan Kratochvíl[1,⋆], Zsolt Tuza[2,⋆⋆], and Gerhard J. Woeginger[3]

[1] Department of Applied Mathematics and
Institute for Theoretical Computer Science[†],
Charles University,
Malostranské nám. 25, 118 00 Prague, Czech Republic
{kral,honza}@kam.ms.mff.cuni.cz

[2] Computer and Automation Institute,
Hungarian Academy of Sciences,
H–1111 Budapest, Kende u. 13–17, and
Department of Computer Science,
University of Veszprém, Hungary
tuza@sztaki.hu

[3] Technical university Graz, Austria
gwoegi@igi.tu-graz.ac.at

Abstract. We give a complete characterization of parameter graphs H for which the problem of coloring H–free graphs is polynomial and for which it is NP–complete. We further initiate a study of this problem for two forbidden subgraphs.

1 Preliminaries and Overview of Results

Graph coloring belongs to the most important and applied graph problems. It also belongs to the first identified NP–complete problems. Many classes of graphs were shown to allow polynomial-time solution (e.g., interval graphs, chordal graphs, etc.). In this paper we aim at classifying the computational complexity of this problem when restricted to graphs that do not contain certain forbidden induced subgraphs. Related results appear in [8], where 3-colorability is studied. We consider, on the other hand, the coloring problem with the number of colors being part of the input. For one forbidden subgraph we obtain a complete characterization of the complexity, which performs the polynomial-time/NP–complete dichotomy. First results in the direction of two forbidden subgraphs are gathered in the last section, but a complete characterization is not yet at hand.

We consider finite simple undirected graphs. We say that a graph G is H–*free*, where H is another graph, if G does not contain an induced subgraph

⋆ This author acknowledges further partial support of Czech research grant GAUK 158/1999.

⋆⋆ Research supported in part by the Hungarian Scientific Research Fund, grants OTKA T–026575 and T-032969, and Czech Grant GAČR 201/99/0242 (DIMATIA).

† Project LN00A056 supported by The Ministry of Education of Czech Republic

A. Brandstädt and V.B. Le (Eds.): WG 2001, LNCS 2204, pp. 254–262, 2001.

isomorphic to H. Typically H will be a fixed "small" graph while G will be part of the input of the problem. For a class of graphs $\mathcal{A}$, we say that G is $\mathcal{A}$*–free* if it is H–free for every $H \in \mathcal{A}$. The graphs of $\mathcal{A}$ are then also referred to as forbidden induced subgraphs in the sense that $\mathcal{A}$ is the set of forbidden induced subgraphs for the class of $\mathcal{A}$–free graphs.

As usual, K_n denotes the complete graph on n vertices, C_n denotes the cycle on n vertices and P_n denotes the path with n vertices (i.e., path of length $n-1$). By $G \oplus H$ we denote the disjoint union of G and H, and in this sense kG means the disjoint union of k copies of G. We use the notation $G \leq H$ when G is isomorphic to an induced subgraph of H. The graph C_k^+ is a cycle of length k with a pending edge, e.g. $C_3^+ = \overline{P_3 \oplus K_1}$. The complement of a graph G is denoted by $\overline{G}$. The chromatic number of G is denoted by $\chi(G)$. The clique covering number of G is denoted by $\sigma(G)$ ($\sigma(G) = \chi(\overline{G})$).

We study the computational complexity of the following problem:

H-Free Coloring
Input: An H–free graph G, and an integer k.
Question: Is $\chi(G) \leq k$?

Our main result is summarized in the next theorem.

Theorem 1. *The problem H-*Free Coloring *is polynomial-time solvable if H is an induced subgraph of P_4 or of $P_3 \oplus K_1$, and NP–complete for any other H.*

We first discuss the polynomial part, then present a useful reduction which is easier to be formulated in the complementary version (i.e., as a result on clique covering), and complete the proof of the NP–completeness part in Section 4. Some ideas towards sets of more forbidden graphs conclude the last section. A more detailed account on partial results concerning 2 forbidden induced subgraphs will be given in the full version of the paper.

2 Polynomial Cases

All graphs H for which the H-Free Coloring problem is polynomial-time solvable are depicted in Fig. 1. If H is an induced subgraph of H' then every H–free graph is also H'–free, and hence H-Free Coloring $\propto$ H'-Free Coloring. Therefore it suffices to prove the polynomial part only for the maximal graphs from Fig. 1, i.e. for P_4 and $P_3 \oplus K_1$.

It is well known that P_4–free graphs are perfect (they are exactly the so called co-graphs), and hence their chromatic number can be determined in linear time [1].

For graphs which do not contain $P_3 \oplus K_1$ we use the following structural result of Olariu:

Proposition 1. ([7]) *If G is $(P_3 \oplus K_1)$–free then every connected component of its complement $\overline{G}$ is triangle–free or is a complete multipartite graph.*

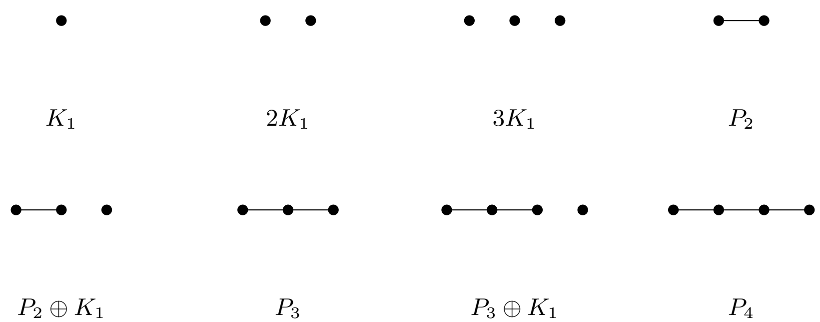


Fig. 1. The graphs H for which H-Free Coloring is polynomially solvable

The chromatic number of a graph equals the clique covering number of its complement. In this way we determine the chromatic number of a $(P_3 \oplus K_1)$–free graph in polynomial time. The clique covering number of each triangle–free connected component of the complement equals the size of the component minus the maximum size of a matching, and thus can be determined in polynomial time (e.g., by Edmonds's algorithm). The clique covering number of complete multipartite graphs can also be computed in polynomial time (it equals the size of the largest part of the multipartition), and so it follows from Proposition 1 that the chromatic number of $(P_3 \oplus K_1)$–free graphs can be computed in polynomial time. This concludes the "easy" part of the theorem.

3 Clique Covering of Sparse Graphs

In this section we present an NP–completeness reduction for coloring graphs without forests which consist of short paths. It is more convenient to state the result for the complementary problem of clique covering. We denote by K_4^- the graph on 4 vertices with 5 edges.

Theorem 2. *The clique covering problem remains NP–complete even when restricted to planar $\{K_4, K_4^-, C_4, C_5\}$–free graphs.*

Proof. We use a reduction from the Satisfiability problem restricted to planar formulas, namely a variant in which every clause of the input formula contains 2 or 3 literals and every variable occurs in exactly 3 clauses, once positive and twice negated. A formula is called planar if its bipartite incidence graph is planar. NP–completeness of planar formulas was first established by Lichtenstein [5], and NP–completeness of the above described restricted version can be found e.g. in [3].

Given a formula Φ with variable set X and clause set C, we construct a graph G as follows. Fix a planar noncrossing drawing of the incidence graph of Φ. For every variable $x \in X$, the variable gadget is the graph

$$G_x = (\{x, x^+, x_1^-, x_2^-\}, \{xx^+, xx_1^-, xx_2^-, x_1^- x_2^-\}).$$

For every clause c, the clause gadget G_c will be a cycle of length 7. Three nonconsecutive vertices of the cycle are used as ports for connecting with variable gadgets. If x, y, z are variables occurring in c such that in the fixed drawing, the edges cx, cy, cz leave the vertex c in this clockwise order, we denote $c(x), c(y), c(z)$ the ports on G_c also in this clockwise order. (If c contains only two literals, we use as ports two vertices of G_c which are at distance two on the cycle.) For every variable x, let c be the clause containing x positive and d_1, d_2 the clauses that contain $\neg x$. Then identify vertices $c(x)$ and x^+ into one vertex, vertices $d_1(x)$ and x_1^- into one vertex and vertices $d_2(x)$ and x_2^- into one vertex.

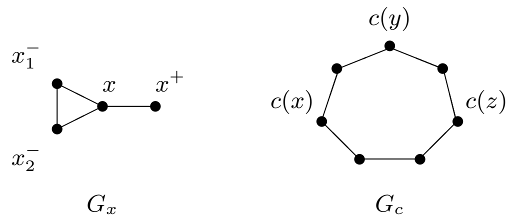


Fig. 2. The construction of variable and clause gadgets

It is clear that the graph G constructed in this way is planar. The only triangles occur in variable gadgets, and hence no two triangles of G share an edge. Thus G contains neither K_4 nor K_4^-. Similarly, G does not contain C_4 as an induced subgraph. In fact, since we may assume that no two clauses share two variables, the shortest cycles of G apart of the triangles in variable gadgets are the 7-cycles in the clause gadgets.

We claim that the clique covering number of G is always $\geq |X| + 3|C|$, with equality holding if and only if Φ is satisfiable.

For every variable x, we need a clique Q_x that would contain vertex $x \in V(G_x)$. This clique may further contain either x^+ (Q_x is the edge xx^+), or x_1^- and x_2^- (if Q_x is the triangle $xx_1^-x_2^-$). Obviously, $x^+ \in Q_x$ implies that Q_x contains neither x_1^- nor x_2^-. This establishes a 1-1 correspondence between truth assignments of X and the ways in which the vertex x is covered:

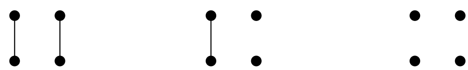
$$\phi(x) = \text{True iff } Q_x = \{xx^+\}.$$

Note that a variable x evaluates True in a clause c if and only if the port $c(x)$ in G_c is covered by Q_x.

For a clause c, we need at least 3 cliques to cover the vertices of G_c which are not ports. Hence $\sigma(G) \geq |X| + 3|C|$. Moreover, if none of the ports in G_c is covered by the appropriate Q_x, we need at least 4 cliques to cover this G_c. Therefore $\sigma(G) > |X| + 3|C|$ if Φ is not satisfiable.

To see the converse, note that C_7 without a vertex is the path P_6, which can be covered by 3 edges. Hence if at least one port in G_c is covered by the variable clique Q_x (i.e., x evaulates True in c), the remaining vertices of such G_c can be covered by 3 cliques. This shows that $\sigma(G) = |X| + 3|C|$ if Φ is satisfiable.

Corollary 1. *The* $\{2K_2, K_2 \oplus 2K_1, 4K_1, C_5\}$-Free Coloring *problem is NP–complete. In particular, the* H-Free Coloring *problem is NP–complete if H is a (not necessarily induced) spanning subgraph of* $2K_2$.



$2K_2$ $K_2 \oplus 2K_1$ $4K_1$

Fig. 3. The minimal linear forests H for which H-Free Coloring is NP–complete

4 NP–Complete Cases

In this section we summarize the NP–completeness part of the proof of Theorem 1. We consider three cases.

Case 1 Suppose H contains a cycle, say C_k. Let G be an instance of vertex 3-colorability, which is well known to be NP–complete.

Let H' be a critical non-3-colorable graph of girth greater than k (i.e., H' does not contain cycles of length at most k; existence of graphs of large girth and chromatic number is a well known result in graph theory, cf. e.g. [2]). Take an edge ab of H', delete it and add a new extra vertex a' pending on b. Call this new graph $H'(a, a')$. Since the girth of H' was greater than k, $H'(a, a')$ does not contain C_k and also every path from a to a' has length greater than k. Since H' was critical non-3-colorable, $H'(a, a')$ is 3-colorable, but in every proper 3-coloring, a and b have the same color, i.e., a and a' have different colors.

Return to G and construct a graph G' by replacing every edge uv by a copy of $H'(u, v)$. Thus G' has girth greater than k and it is 3-colorable if and only if G is 3-colorable. Hence 3-coloring H–free graphs is NP–complete.

Case 2 Let H be a forest, and suppose H contains a vertex of degree at least 3. Then H contains $K_{1,3}$ (the claw). Since 3-coloring is NP–complete for line graphs by Holyer [4] and since line graphs are claw–free, $K_{1,3}$-Free Coloring is NP–complete, and so is H-Free Coloring in this case.

Case 3 Suppose finally that H is a linear forest, i.e., the disjoint union of paths.

If H has at least 4 connected components then H contains $4K_1$ and H-Free Coloring is NP–complete by Corollary 1.

If H has exactly 3 connected components then either $H = 3K_1$ (in which case $H \leq P_3 \oplus K_1$ and H-Free Coloring is polynomial by Proposition 1) or $K_2 \oplus 2K_1 \leq H$ (and H-Free Coloring is NP–complete by Corollary 1).

Suppose H has exactly 2 connected components. If none of them is an isolated vertex then $2K_2 \leq H$. If one of the components is a path of length at least 3 then $K_2 \oplus 2K_1 \leq H$. In these two cases H-FREE COLORING is NP–complete by Corollary 1. If, on the other hand, $H \leq P_3 \oplus K_1$ then H-Free Coloring is polynomial by Proposition 1.

If H is connected then either $H \leq P_4$ and H-Free Coloring is polynomial since every H–free graph is a co-graph, or $P_5 \leq H$ and H-Free Coloring is NP–complete by Corollary 1 since $2K_2 \leq H$.

5 Towards Two Forbidden Subgraphs

A natural generalization of the question of coloring H–free graphs is as follows.

Meta-Problem Given a finite set of graphs $\mathcal{A}$, what is the computational complexity of deciding the chromatic number of $\mathcal{A}$–free graphs?

We have completely classified the case of one-element sets $\mathcal{A}$, and the nearest goal may be the case of two-element sets, i.e., coloring graphs that do not contain two specified induced subgraphs. Here the situation seems more complex and we have only obtained partial results. Our proof of Theorem 1 was based on two different NP–completeness constructions (our construction in the proof of Theorem 2 and Holyer's reduction) which so far resist attempts to be combined.

We classify graphs H into four types (Type D is disjoint from Types A,B and C, but the latter three are not mutually disjoint):

Type A graphs containing cycles
Type B graphs containing a claw (induced copy of $K_{1,3}$)
Type C graphs containing an induced copy of a spanning subgraph of $2K_2$ (these are depicted in Fig. 3)
Type D graphs which are induced subgraphs of P_4 or of $P_3 \oplus K_1$.

The following observation says that for determining the complexity of the $\{H_1, H_2\}$-Free Coloring problem, we only need to care about cases when H_1 and H_2 are of different types among A, B and C. The three cases are then discussed in the subsequent subsections.

Proposition 2. *Let $\mathcal{A}$ be a finite set of graphs.*

(i) If at least one of the graphs of $\mathcal{A}$ is of Type D then $\mathcal{A}$-FREE COLORING is polynomially solvable.

*(ii) If all graphs of $\mathcal{A}$ are of the same type A or B or C, then $\mathcal{A}$-*Free Coloring *is NP–complete.*

Proof. If one of the graphs $H \in \mathcal{A}$ determines a polynomially solvable instance, then so does the set $\mathcal{A}$, since every $\mathcal{A}$–free graph is also H–free. This implies (i) by Theorem 1.

If all graphs of $\mathcal{A}$ are of type B (or C), then every claw–free ($\{2K_2, K_2 \oplus 2K_1, 4K_1\}$–free) graph is $\mathcal{A}$–free and Holyer's reduction (or our Proposition 1) apply.

If all graphs of $\mathcal{A}$ are of type A, we perform the reduction described in Case 1 of Section 4, with k being the length of a longest cycle among all graphs of $\mathcal{A}$.

5.1 Types B versus C (Forests versus Forests)

Say H_1 is of Type B and H_2 is of Type C. Let us assume further that none of H_1, H_2 is of Type A (the open problems of these cases will be stated below). Every forest which contains a claw and at least one more vertex contains a spanning subgraph of $2K_2$, regardless of how and if the other vertex is connected to the claw. So if $H_1 \neq K_{1,3}$ then both H_1 and H_2 are of Type C and the $\{H_1, H_2\}$-Free Coloring problem is NP–complete. Thus the only open question remains when $H_1 = K_{1,3}$:

Problem 1. Complexity of $\{H, K_{1,3}\}$-Free Coloring in case H is a linear forest which contains a spanning subgraph of $2K_2$ as an induced subgraph.

5.2 Types A versus B (Cycles versus Claws)

Suppose $C_k \leq H_1$ and $K_{1,3} \leq H_2$. We first show that all cases with $k \geq 4$ determine difficult instances:

Proposition 3. *The problem $\{C_k, K_{1,3}\}$-Free Coloring is NP–complete for every $k \geq 4$. Hence it is NP–complete for every pair H_1, H_2 such that $C_k \leq H_1$, $k \geq 4$, and $K_{1,3} \leq H_2$.*

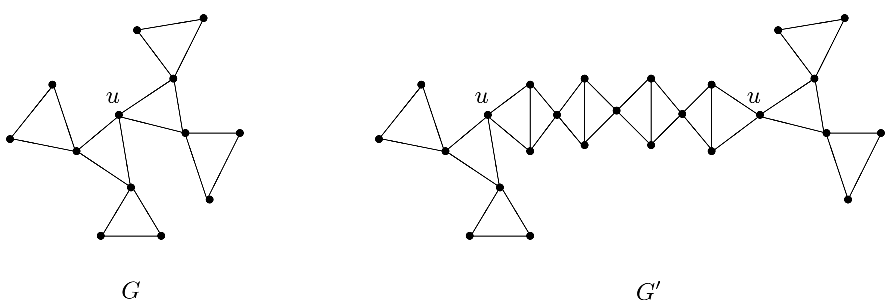


Fig. 4. Chain of K_4^-'s.

Proof. Holyer proved that 3-colorability of line graphs of cubic triangle–free graphs is NP–complete. Hence 3-colorability is NP–complete for 4-regular claw–free graphs such that any two triangles share at most one vertex. Take such a graph G, split each vertex into two (in such a way that both triangles containing the vertex survive) and connect these two vertices by a sufficiently long chain of copies of K_4^- as indicated in Fig. 4. The new graph G' does not contain short cycles (apart of triangles) nor it contains claws. The chain of K_4^-'s enforces that in any 3-coloring the split vertices receive the same color, and so the new graph is 3-colorable if and only if the original one was.

So it remains to consider the case when H_1 contains triangles but no cycle of greater length. Take again Holyer's reduction. Line graphs of cubic triangle–free graphs are claw–free and any two triangles share at most one vertex. Hence we can conclude:

Observation 3. The problem $\{K_4,K_4^-,K_{1,3}\}$-Free Coloring is NP–complete.

This leaves the following open case:

Problem 2. Complexity of $\{H_1, H_2\}$-Free Coloring for H_1 being a claw–free graph containing triangles but no longer cycles, such that every two cycles share at most one vertex, and H_2 being a forest of maximum degree at least 3.

Note that graphs H_1 from Problem 2 have a tree-like structure. If H_2 has more than 4 vertices it contains a spanning subgraph of $2K_2$, and hence the problem is NP–complete whenever H_1 is of Type B as well. So the interesting cases are if H_1 is a triangle, or a triangle plus an isolated vertex, or a triangle with one or two pendant edges. However, if $H_2 = K_{1,3}$ then the number of open cases is larger. Though many of them are shown NP–complete by the reduction in the proof of Proposition 3, many still remain open. A detailed description of the NP–complete and polynomial–time solvable cases known to us will appear in the full version of the paper. Note also that only cases with forests H_2 of maximum degree at most 4 remain open, because of Observation 5 below.

It may be surprising that excluding either C_3 or C_3^+ together with any other graph is essentially equivalent:

Observation 4. The problems $\{C_3, H\}$-Free Coloring and $\{C_3^+, H\}$-Free Coloring are polynomially equivalent.

Proof. Every C_3^+–free connected graph is a C_3–free graph or a complete multipartite graph due to Proposition 1. Both complete multipartite graphs and C_3–free graphs can be easily recognized in polynomial time. Thus if the problem $\{C_3, H\}$-Free Coloring can be solved in polynomial time for a fixed k, we split a graph given as an instance of the problem $\{C_3^+, H\}$-Free Coloring to its connected components, and we apply the algorithm of the problem $\{C_3, H\}$-Free Coloring to the components which are C_3–free. The chromatic number of its other connected components can be determined easily in polynomial time, since each of them forms a complete multipartite graph. The chromatic number of the original graph is clearly the maximum chromatic number of its connected components. Thus the problem $\{C_3^+, H\}$-Free Coloring can be solved in polynomial time in this case, too.

On the other hand, C_3 is an induced subgraph of C_3^+ and hence it holds that $\{C_3, H\}$-Free Coloring $\propto$ $\{C_3^+, H\}$-Free Coloring, since any instance of the problem $\{C_3^+, H\}$-Free Coloring is also an instance of the problem $\{C_3, H\}$-Free Coloring.

We believe that at least the particular case of $H_1 = C_3$ in Problem 2 deserves separate interest. In other words, we investigate triangle–free graphs. The case of a triangle against a star is fully characterized:

Observation 5. The problems $\{C_3, K_{1,k}\}$-Free Coloring and $\{C_3^+, K_{1,k}\}$-Free Coloring are polynomially solvable for $k \leq 4$ and NP-complete otherwise.

Proof. It is enough to restrict to $\{C_3, K_{1,k}\}$-Free Coloring due to Observation 4. Every triangle–free $K_{1,4}$–free graph has maximum degree 3 and by Brooks's theorem, it is 3-colorable, unless it has K_4 as a connected component, an easily distinguishable instance. The NP–completeness of $\{C_3, K_{1,5}\}$-Free Coloring is proved in [6].

5.3 Types A versus C (Cycles versus Linear Forests)

In the last case we assume that $C_k \leq H_1$ and H_2 contains a spanning subgraph of $2K_2$ as induced subgraph. This case is solved if cycles of length greater than 4 appear in H_1:

Proposition 4. *The problem* $\{C_k, H_2\}$-Free Coloring *is NP–complete if* $k \geq 5$ *and* H_2 *contains a spanning subgraph of* $2K_2$ *as an induced subgraph.*

Proof. If $k \geq 6$ then $2K_2 \leq C_k$ and the NP–completeness follows from the fact that both H_1 and H_2 are of the same Type C. For $k = 5$, NP–completeness follows from Corollary 1.

The last open case, however, includes a large number of subproblems and its complexity is not well understood. We note that many cases are polynomially solvable because of structural reasons (including Ramsey's Theorem), e.g., $\{K_4, 4K_1\}$-Free Coloring allows only finite number of input graphs.

Problem 3. Complexity of $\{H_1, H_2\}$-Free Coloring if H_1 contains only induced cycles of length 3 and/or 4, and H_2 contains a spanning subgraph of $2K_2$ as induced subgraph.

References

1. Chvátal, V.: Perfectly ordered graphs, in: Topics on Perfect Graphs (C.Berge, V.Chvátal, eds.), Annals of Discrete Mathematics 21 (1984) 63-65
2. Erdős, P.: Graph theory and probability, Canad. J. Math. 11 (1959) 34-38.
3. Fellows, M.R., Kratochvíl, J., Middendorf, M., Pfeiffer, F.: The complexity of induced minors and related problems, Algorithmica 13 (1995) 266-282
4. Holyer, I.: The NP–completeness of edge-coloring, SIAM J. Comput. 10 (1981) 718-720.
5. Lichtenstein, D.: Planar formulae and their uses, SIAM J. Comput. 11 (1982) 329-343.
6. Maffray, F., Preissmann, M.: On the NP–completeness of the k-colorability problem for triangle–free graphs, Discr. Math. 162 (1996), 313-317
7. Olariu, S.: Paw–free graphs, Inform. Process. Lett. 28 (1988), 53-54
8. Randerath, B., Schiermeyer, I.: 3-colorability in P for P_6–free graphs, Rutcor Research Report 39-2001, ext. abstract in Electronic Notes in Discrete Math., Vol. 8 (2001).