Maximum colorful independent sets in vertex-colored graphs

Yannis Manoussakis [1] Hong Phong Pham [1]

LRI, Univ.Paris Sud, CNRS, CentraleSupélec, Université Paris-Saclay
91405 Orsay Cedex, France

Abstract

In this paper we study the problem of finding a maximum colorful independent set in vertex-colored graphs. Specifically, given a graph with colored vertices, we wish to find an independent set containing the maximum number of colors. Here we aim to give a dichotomy overview on the complexity of this problem. We first show that the problem is NP-hard even for some cases where the maximum independent set problem is easy, such as cographs and P_5-free graphs. Next, we provide polynomial-time algorithms for cluster graphs and trees.

Keywords: maximum colorful independent sets, vertex-colored graphs, polynomial-time algorithms

1 Introduction

In this paper we deal with vertex-colored graphs. Given a vertex-colored graph, a *tropical subgraph* is a subgraph where each color of the initial graph appears at least once. There are many cases, however, where tropical subgraphs do not necessarily exist. Hence, one might be interested in the more

[1] Emails: `yannis@lri.fr, phongph.hut@gmail.com`

general question of finding a *maximum colorful* subgraph, i.e., a subgraph with the maximum possible number of colors. Note that in a *colorful* subgraph considered in our paper, two adjacent vertices may have the same color, i.e., the subgraph is not necessarily properly colored. In this paper, we are interested in finding maximum colorful independent sets in vertex-colored graphs.

Throughout, we let $G = (V, E)$ denote a simple undirected graph. Given a set of colors $\mathcal{C}$, $G^c = (V, E)$ denotes a vertex-colored graph whose vertices are (not necessarily properly) colored by one of the colors in $\mathcal{C}$. The number of colors of G^c is $|\mathcal{C}|$. Given a subset of vertices $U \subseteq V$, the set of colors of vertices in U is denoted by $\mathcal{C}(U)$. Moreover, we denote by $c(v)$ the color of vertex v and by $v(H, c)$ the number of vertices of H whose color is c. The set of neighbors of v is denoted by $N(v)$. More formally, in this paper we study the following problem:

Maximum Colorful Independent Set Problem (MCISP)

Given a vertex-colored graph $G^c = (V, E)$, find an independent set with the maximum number of colors of G^c.

Related work

In the special case where each vertex has a distinct color, MCISP reduces to the maximum independent set problem. The maximum independent set problem is known to be NP-hard for general graphs, fixed-parameter intractable and hard to approximate. However, the maximum independent set problem can be efficiently solved in polynomial time for several special classes of graphs, such as claw-free graphs [2], P_5-free graphs [1], and perfect graphs. Especially, this problem can be solved in linear time for chordal graphs and cographs [7].

Tropical subgraph and maximum colorful subgraph problems have been studied only recently. Especially, the maximum colorful matching problem [6], have been recently studied.

Due to space limit, some results, proofs and details are omitted from this extended abstract and are deferred to the appendix.

2 Hardness results for MCISP

In this section, we present some NP-hardness results for MCISP. Specifically, we show that MCISP is NP-hard for P_5-free graphs and cographs which are easy cases for the maximum independent set problem [1], [7].

Lemma 2.1 *The maximum colorful independent set problem is NP-hard for P_5-free graphs.*

Sketch of proof. We reduce from the MAX-3SAT problem. Consider a boolean expression B in CNF with variables $X = \{x_1, \ldots, x_s\}$ and clauses $B = \{b_1, \ldots, b_t\}$. In addition, suppose that B constains exactly 3 literals per clause. We show how to construct a vertex-colored P_5-free graph G^c associated with any such formula B, such that, there exists a truth assignment to the variables of B satisfying t' clauses if and only if G^c contains an independent set with t' distinct colors. Suppose that $\forall i, 1 \leq i \leq s$, the variable x_i appears in clauses $b_{i1}, b_{i2}, \ldots, b_{i\alpha_i}$ and $\overline{x_i}$ appears in clauses $b'_{i1}, b'_{i2}, \ldots, b'_{i\beta_i}$ in which $b_{ij} \in B$ and $b'_{ik} \in B$. Now a vertex-colored P_5-free graph graph G^c is constructed as follows. Vertices are of the form (x_i, b_{ij}) if x_i appears in the clause b_{ij}, or $(\overline{x_i}, b'_{ik})$ if $\overline{x_i}$ appears in the clause b'_{ik}. Now the set of edges are defined as follows. There is an edge between each vertex (x_i, b_{ij}) and each vertex $(\overline{x_i}, b'_{ik})$ for $1 \leq j \leq \alpha_i$ and $1 \leq k \leq \beta_i$. In contrast, there are no edges between (x_i, b_{ij}) and $(x_{i'}, b_{i'k})$, between (x_i, b_{ij}) and $(\overline{x_{i'}}, b'_{i'k})$, between (x_i, b_{ij}) and $(x_i, b_{ij'})$. Additionally, we use the color c_l for the vertex (x_i, b_{ij}), or $(\overline{x_i}, b'_{ik})$ if b_{ij} or b'_{ik} is the clause b_l of B. It is not difficult to check that this graph G^c satisfies the property mentioned above. □

Remark 2.2 Observe that the constructed graph is a bipartite graph, thus MCISP is also NP-hard for bipartite graphs which are easy cases of the maximum independent set problem. Therefore, MCISP is also NP-hard for perfect graphs.

Next, we will show that MCISP remains NP-hard for cographs by the following lemma.

Lemma 2.3 *The maximum colorful independent set problem is NP-hard for cographs.*

3 Efficient algorithms for MCISP

In this section, we present our polynomial-time algorithms for MCISP in cluster graphs and trees. We first consider the case of cluster graphs which are sub-class of cographs and claw-free graphs.

3.1 An algorithm for MCISP for cluster graphs

Recall that a cluster graph is a graph formed from the disjoint union of complete graphs. Observe that if we pick one vertex from each complete graph then we obtain an independent set. Therefore, MCISP is equivalent to the problem of choosing one vertex from each complete graph such that the number of colors is maximized. Note that all complete graphs have to be selected, otherwise one could add more vertices. From now on, we denote a set of such vertices as a maximum independent set. Let ℓ be the number of complete graphs. Let denote the set of complete graphs by $K_1, K_2, \ldots, K_\ell$.

Main ideas is to convert MCISP to the maximum colorful matching problem by creating another graph and then applying the algorithm in [6] to find a maximum colorful matching in new graph. From the cluster graph G^c, we first remove all edges of all complete graphs. Next, we add ℓ new vertices $u_1, u_2, \ldots, u_\ell$ to G^c in which the vertex u_i is colored by c'_i which is distinct from all other colors. Additionally, we add an edge from u_i to each vertex of K_i, for all $1 \leq i \leq \ell$. Let denote the new graph by G'^c. Now the following proposition shows the relationship of two problems on two these graphs.

Proposition 3.1 *Let $S = \{v_1, v_2, \ldots, v_\ell\}$ be a maximum independent set of G^c. Then S is a maximum colorful independent set of G^c if and only if $M = \{(u_1, v_1), (u_2, v_2), \ldots, (u_\ell, v_\ell)\}$ is a maximum colorful matching of G'^c.*

Proof. We first assume that S is a maximum colorful independent set of G^c, suppose that M is not a maximum colorful matching of G'^c. Let $M' = \{(u_1, v'_1), (u_2, v'_2), \ldots, (u_\ell, v'_\ell)\}$ be a maximum colorful matching of G'^c and let $S' = \{v'_1, v'_2, \ldots, v'_\ell\}$. Clearly we have that $|\mathcal{C}(M')| > |\mathcal{C}(M)|$. Since the color of each u_i is distinct from all other colors, we easily obtain that $|\mathcal{C}(S')| > |\mathcal{C}(S)|$. This is a contradiction since S is a maximum colorful independent set. Conversely, suppose that M is a maximum colorful matching G'^c. It is possible to show similarly that S must be a maximum colorful independent set of G^c. $\square$

As as result, this proposition allows to use the algorithm of finding a maximum colorful matching in [6] for MCISP. So the first algorithm is presented as follows.

Theorem 3.2 *Let G^c be a vertex-colored cluster graph. Algorithm 1, compute a maximum colorful independent set of G^c in time max$\{O((c+\ell)n), O(\sqrt{n}n)\}$.*

Proof. Algorithm 1 uses the algorithm in [6] to find a maximum colorful matching of a vertex-colored graph G^c, the complexity of this algorithm is $\max\{O(c|E(G)|), O(\sqrt{|V(G)|}|E(G)|)\}$. It is easy to see that in new graph G'^c, we have $|E(G'^c)| = |K_1| + |K_2| + \ldots + |K_\ell| = |V(G^c)| = n$ and $|V(G'^c)| = n + \ell \leq 2n$. Moreover, G'^c uses $c + \ell$ colors for its vertices. So, the complexity of Algorithm 1 is $\max\{O((c + \ell)n), O(\sqrt{n}n)\}$. $\square$

Algorithm 1 Maximum colorful independent sets in vertex-colored cluster graphs.

1: Let G'^c be the graph obtained by removing from G^c all its edge, adding ℓ new vertices $\{u_1, u_2, \ldots, u_\ell\}$ with distinct colors and adding edges between each u_i and all edges of K_i, $1 \le i \le \ell$.
2: $M \leftarrow$ a maximum colorful matching of G'^c (apply the algorihtm in [6]). Let $M = \{(u_1, v_1), (u_2, v_2), \ldots, (u_\ell, v_\ell)\}$.
3: **return** $S = \{v_1, v_2, \ldots, v_\ell\}$ as a maximum colorful independent set of G^c.

3.2 An algorithm for MCISP for trees

We remark here that the maximum colorful independent set can be solved in linear time for trees by using the dynamic programming method. However, we can not use this approach for MCISP since the repetition of colors does not allow to compute the optimal solution of main problem based on optimal solution of sub-problems. Instead of that, we still apply the algorithm of finding maximum colorful matchings in [6] for MCISP for trees.

Specially, we also create a vertex-colored tree $G^{'c}$ from the original vertex-colored tree G^c as follows. For each leaf v of G^c, we create another vertex t_v and use the color of v for t_v (i.e. $c(t_v) = c(v)$). Moreover, we also add the edge (v, t_v) to G^c. Clearly the obtained graph $G^{'c}$ is also a vertex-colored tree. Note that $G^{'c}$ is said to be the *extended graph* of G^c. Now the following proposition allows to find a maximum colorful independent set of G^c based on computing a maximum colorful matching of $G^{'c}$.

Proposition 3.3 *Let S be an independent set of G^c, then there exists a matching $M(S)$ of $G^{'c}$ that contains all vertices of S.*

So, our first step is to apply the algorithm in [6] to finding a maximum colorful matching M of the extended graph $G^{'c}$. After that, we must identify the maximum colorful independent set from this matching M. Clearly M is also a cluster graph, so it is possible to apply our algorithm in previous section for MCISP for M. Note that the obtained maximum colorful independent set can contain vertices like t_v (but can not contain both v and t_v), in this case clearly we can replace t_v by v since they have the same color. In summary, our algorithm is formally presented in Algorithm 2.

Theorem 3.4 *Let G^c be a vertex-colored tree. Then, Algorithm 2 computes a maximum colorful independent set of G^c in time $O((c+n)n)$.*

Proof. Clearly, the first step of adding new vertices runs in time $O(n)$. Next, the second step using the algorihtm in [6] has complexity of $max\{O(cn), O(\sqrt{n}n)\}$.

Algorithm 2 Maximum colorful independent sets in vertex-colored trees.

1: G'^c is the graph obtained from G^c by adding a vertex t_v (colored by $c(v)$) and an edge (v, t_v) for each leaf vertex v of G^c .
2: $M \leftarrow$ a maximum colorful matching of G'^c (apply the algorihtm in [6])
3: $S \leftarrow$ a maximum colorful independent set of M (apply Algorithm 1).
4: Replace each vertex t_v of S by v.
5: **return** S as a maximum colorful independent set of G^c.

As considering the cluster graph M, the work of finding a maximum colorful independent set of M by applying Algorihtm 1 runs in time $max\{O((c+|E(M)|)|V(M)|), O(\sqrt{|V(M)|}|V(M)|)\}$. Therefore, it is easy to see that the complexity of Algorithm 2 is $max\{O((c+n)n), O(\sqrt{n}n)\} = O((c+n)n)$. □

References

[1] D. Lokshtanov, M.Vatshelle, *Independent sets in P5-free graphs in polynomial time*, In SODA: 570-581, 2014.

[2] D.Nakamura, A.Tamura, *A revision of Minty's algorithm for finding a maximum weight stable set of a claw-free graph*, Journal of Operations Research Society Japan **44**, 194-204, 2001.

[3] J.-A. Anglès d'Auriac, N. Cohen, A. El Maftouhi, et al, *Connected Tropical Subgraphs in Vertex-Colored Graphs*, DMTCS, **17(3)**, 327-348, 2016.

[4] J.-A. Anglès d'Auriac, H. El Mafthoui, M. Karpinski, Y. Manoussakis, et al, *Tropical Dominating Sets in Vertex-Colored Graphs*, LNCS **9627**, 17-27, 2016.

[5] J. Edmonds *Paths, trees, and flowers*, Canad. J. Math **17**, 449-467, 1965.

[6] J.Cohen, Y. Manoussakis, H.P. Phong, Zs. Tuza, *Tropical matchings in vertex-colored graphs*, ENDM **62**, 219-224, 2017.

[7] Robert E.Tarjan, *Decomposition by clique separators*, Discrete Mathematics **55(2)**, 221-232, 1985.

[8] S. Akbari, V. Liaghat and A. Nikzad, *Colorful paths in vertex-colorings of graphs*, Electronic Journal of Combinatorics **18(1)**, 17-26, 2011.

[9] S. Micali and V. Vazirani, *An $O(\sqrt{n}m)$ algorithm for finding maximum matching in general graphs.* In 21st FOCS: 434-443, 1980.