Maximum colorful cycles in vertex-colored graphs

Giuseppe F. Italiano[2], Yannis Manoussakis[1], Nguyen Kim Thang[3], and Hong Phong Pham[1]

[1] LRI, University Paris-Saclay, Orsay, France
[2] University of Rome Tor Vergata, Rome, Italy
[3] IBISC, University Paris-Saclay, Evry, France

Abstract. In this paper, we study the problem of finding a maximum colorful cycle a vertex-colored graph. Specifically, given a graph with colored vertices, the goal is to find a cycle containing the maximum number of colors. We aim to give a dichotomy overview on the complexity of the problem. We first show that the problem is NP-hard even for simple graphs such as split graphs, biconnected graphs, interval graphs. Then we provide polynomial-time algorithms for classes of vertex-colored threshold graphs and vertex-colored bipartite chain graphs, which are our main contributions.

1 Introduction

In this paper we deal with vertex-colored graphs, which are useful in various situations. For instance, the Web graph may be considered as a vertex-colored graph where the color of a vertex represents the content of the corresponding page (red for mathematics, yellow for physics, etc.) [4]. In a biological population, vertex-colored graphs can be used to represent the connections and interactions between species where different species have different colors. Other applications of vertex-colored graphs can also be found in bioinformatics (Multiple Sequence Alignment Pipeline or for multiple Protein-Protein Interaction networks) [7], or in a number of scheduling problems [17].

Given a vertex-colored graph, a *tropical subgraph* is a subgraph where each color of the initial graph appears at least once. Many graph properties, such as the domination number, the vertex cover number, independent sets, connected components, paths, matchings etc. can be studied in their tropical version. Finding a tropical subgraph in a (biological) population is to look for a subgraph which fully represents the (bio-)diversity of the population. In this paper, we consider a more general question of finding a *maximum colorful* subgraph which is a subgraph with maximum number of colors. Given a vertex-colored graph and some property of a subgraph (for example, paths, cycles, connected components), it could be that the tropical subgraph with the given property does not exist. Hence, one can ask the question of finding a subgraph with the most diverse population. Clearly, a maximum colorful subgraph is tropical if it contains all colors.

This notion of colorful subgraph is close to, but somewhat different from the *colorful* concept considered in [1, 14, 15], where neighbor vertices must have different colors. It is also related to the concepts of *color patterns* or *colorful* used in bio-informatics [8]. Note that in a *colorful* subgraph considered in our paper, two adjacent vertices may have the same color. In this paper, we study maximum colorful cycles in vertex-colored graphs.

Throughout the paper, 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 $U \subset V$, the set of colors of vertices in U is denoted by $\mathcal{C}(U)$. Moreover, we denote the color of the vertex v by $c(v)$ and denote the number of vertices of U whose colors is c by $v(U, c)$. The set of neighbors of v is denoted by $N(v)$. In this paper, we study the following problem:

Maximum Colorful Cycle Problem (MCCP). Given a vertex-colored graph $G^c = (V, E)$, find a simple cycle with the maximum number of colors of G^c.

Related work. In the special case where each vertex has a distinct color, MCCP reduces to the Hamiltonian cycle problem. The Hamiltonian cycle problem has been widely studied in the literature and it is well known that this problem is NP-complete even for specific classes of graphs such as for undirected planar graphs of maximum degree three [10], for 3-connected 3-regular bipartite graphs [2], etc. However, the Hamiltonian cycle problem can be solved in time $O(m+n)$ for proper interval graphs [3, 13].

If a graph G^c is Hamiltonian then it must contain a tropical cycle (which is a maximum colorful cycle) since the set of vertices must contain all colors. The problem of finding a longest cycle has been also studied and this problem can be used to solve the Hamiltonian cycle problem (and thus it is NP-hard). However, for some classes of graphs, there exist polynomial time algorithms for finding the longest cycle in threshold graphs [16], and in bipartite chain graphs [19]. Note that a longest cycle does not necessarily contain the maximum number of colors. However, in our paper, we take advantage of those algorithms to construct a Hamiltonian cycle for a given set of candidate vertices of a maximum colorful cycle.

The tropical subgraph and maximum colorful subgraph problems in vertex-colored graphs have been studied recently. The tropical subgraph problems in vertex-colored graphs such as tropical connected subgraphs, tropical independent sets have been investigated in [9]. Recently, the maximum colorful matching problem [5] and the maximum colorful path problem [6] have been studied, and several hardness results and polynomial-time algorithms were shown for different classes of graphs.

Our contributions. In this paper, we aim to give dichotomy overview on the complexity of MCCP. First, we prove that MCCP is NP-hard even for split graphs, interval graphs and biconnected graphs. Next, we present polynomial-time algorithms for several classes of graphs. First, we show that the MCCP is polynomial for proper interval graphs and split complete graphs. Although those algorithms are not complicated, they provide a sharp separation in term of complexity for interval and split graphs.

Our main contributions are polynomial-time algorithms for threshold graphs and bipartite chain graphs. A graph G is a *threshold* graph if it is constructed from the repetition of two operations: (1) adding an *isolated* vertex to the current graph, or (2) adding a *dominating* vertex to the current graph, i.e.,. one vertex connected to all vertices added earlier. A *bipartite chain* graph $G(X \cup Y, E)$ is a bipartite graph in which vertices in X can be linearly ordered such that $N(x_1) \supseteq N(x_2) \supseteq \ldots \supseteq N(x_{|X|})$. In our approach for both threshold and bipartite chain graphs, we develop connections between maximum colorful cycles and maximum colorful matchings and derive structural properties of maximum colorful cycles. Those properties enable us to identify a small set of candidate vertices for maximum tropical cycles. Subsequently, using longest cycle algorithms [16, 19], on these vertices, we can efficiently compute the corresponding maximum tropical cycles for MCCP. The running times of our algorithms are $O(\max\{|\mathcal{C}| \cdot M(m,n)), n(n+m)\})$ and $O(|\mathcal{C}| \cdot \max\{M(m,n), n^3\})$ for threshold graphs and for bipartite graphs respectively, where $|\mathcal{C}|$ is the total number of colors and $M(m,n)$ is the running time for finding a maximum matching in a general graph with m edges and n vertices. (It is known that $M(m,n) = O(\sqrt{n}m)$ [18].) Due to space limit, some results are put in the appendix.

2 Hardness results for MCCP

Theorem 1. *MCCP is NP-hard for interval graphs and biconnected graphs.*

Proof. We reduce from the SAT problem. Consider a boolean CNF formula B with variables $X = \{x_1, \ldots, x_s\}$ and clauses $B = \{b_1, \ldots, b_t\}$. We construct the following graph. Suppose that $\forall 1 \leq i \leq s$, 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 $b'_{ik} \in B$. An intersection model for our graph is constructed as follows. On the real line, we create $(s+1)$ intervals $v_1, v_2, \ldots, v_{s+1}$ such that v_i intersects only v_{i-1} and v_{i+1} $(1 \leq i \leq s-1)$. Next for each variable x_i of X $(1 \leq i \leq s)$, we create α_i same intervals $b_{i1}, b_{i2}, \ldots, b_{i\alpha_i}$ such that these intervals intersect pairwise each other and intersect only v_i and v_{i+1} among other vertices v_j. Similarly, for each $\overline{x_i}$, β_i same intervals $b'_{i1}, b'_{i2}, \ldots, b'_{i\beta_i}$ are drawn such that they intersect pairwise each other and intersect only v_i and v_{i+1}. Additionally, we create one special interval v_0 such that v_0 intersects with all other intervals, except for the intervals b_{1j} and $b'_{1j}(0 \leq j \leq b_{1\alpha_1}, b'_{1\beta_1})$. Note that this graph is both interval and biconnected.

From this intersection model, we obtain the corresponding interval graph and give colors as follows. Every vertex corresponding to the clause b_l has the same color c_l. Observe that vertices $b_{i1}, b_{i2}, \ldots, b_{i\alpha_i}$ make a clique in G^c, similarly for vertices $b'_{i1}, b'_{i2}, \ldots, b'_{i\beta_i}$. For the vertex v_i, we use the color c'_i such that all colors c'_i are distinct and different from the colors c_l. See an illustration in Figure 1.

Now we claim that there exists a truth assignment to the variables of B satisfying all clauses if and only if G^c contains a cycle with all colors.

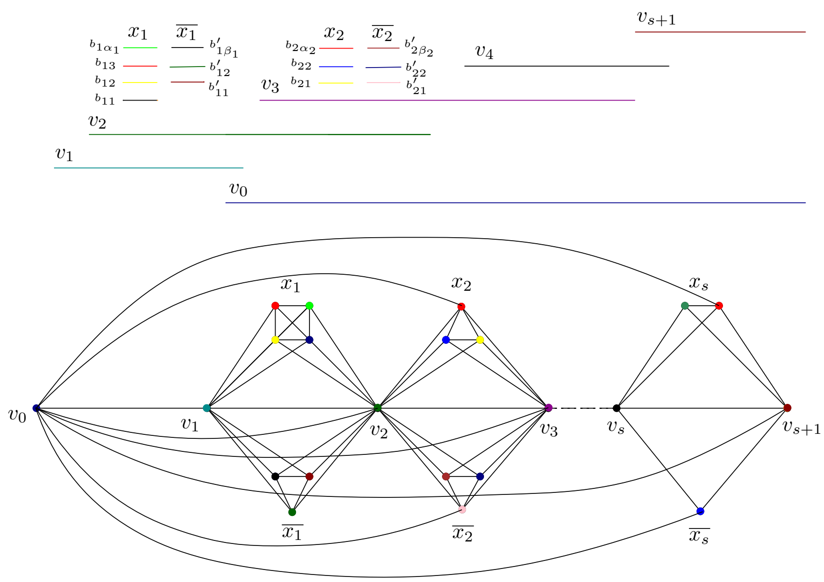

Fig. 1. Reduction of the SAT problem to MCCP for interval graphs.

Now from a truth assignment B with all satisfied clauses, it is possible to obtain a cycle with all colors as follows. We start with the edge (v_0, v_1). Then, to go from v_i to v_{i+1}, $1 \leq i \leq s$, in the case that the variable x_i is assigned true then we select the sub-path $(v_i \to b_{i1} \to b_{i2} \to \ldots \to b_{i\alpha_i} \to v_{i+1})$ into the final path. Otherwise, i.e., x_i is assigned as false, then the sub-path $(v_i \to b'_{i1} \to b'_{i2} \to \ldots \to b'_{i\beta_i} \to v_{i+1})$ is selected. Finally, the edge (v_{s+1}, v_0) is added. It is clear that we obtain a cycle with all colors of G^c.

Conversely, from a cycle K containing all colors G^c, we obtain an assignment with all satisfied clauses as follows. Observe first that all vertices v_i must be in K since their colors are distinct. Since v_0, v_1, v_2 are in K and v_0 is not connected to any b_{1j} or $b'_{1j}(0 \leq j \leq b_{1\alpha_1}, b'_{1\beta_1})$, so we must have that the edge (v_0, v_1) must be in K (otherwise, both v_0 and v_1 can not be in K together). Note that in the case that v_0 is directly connected to any v_i or b_{ij} or b'_{ik} $(2 \leq i \leq s)$ then both v_1 and v_{s+1} can not be in K together. Thus v_0 must be connected to v_1 and v_{s+1} in K. Now, for each $1 \leq i \leq s$, K must go from v_i to v_{i+1}. In the case K goes from v_i to v_{i+1} through one path at the side of x_i then we assign the value *true* for x_i, if the side of $\overline{x_i}$ is used then *false* is assigned for x_i, otherwise (the edge (v_i, v_{i+1}) is in K) we assign arbitrarily *true* or *false* for x_i. Clearly this assignment is consistent for each x_i . Since except for all the colors of v_i then all colors of clauses of B must be in K, we obtain that this assignment satisfies all clauses of B. This completes our proof. □

3 An algorithm for MCCP in threshold graphs

Recall that a graph G is a *threshold* graph if it is constructed from the repetition of two operations: (1) adding an *isolated* vertex to the current graph, or (2) adding a *dominating* vertex to the current graph, i.e., one vertex connected to all vertices added earlier. In the following, we denote vertices of type (1) as *isolated vertices* and vertices of type (2) as *dominating vertices*. Let G^c be a vertex-colored threshold graph. Without loss of generality, we can assume that the last added vertex v to G^c is a *dominating* vertex (otherwise v would be an isolated vertex and it would not appear in a maximum colorful cycle, unless we are in the trivial case where the maximum colorful cycle has size one). By this assumption, G^c is connected. It follows from the construction of threshold graphs that any edge must contain at least one dominating vertex and any two dominating vertices must be connected to each other.

We denote by X the set of dominating vertices of G^c, and by Y the set of isolated vertices of G^c. The set of vertices $V(G^c)$ is denoted by $\{v_1, v_2, \ldots, v_m\}$, in the order in which they were added to G^c. We also denote the *number of colors* of any maximum colorful cycle and of any maximum colorful matching in a vertex-colored threshold graph G^c respectively by C_c and by C_m. Recall that C_m can be computed by the algorithm in [5]. In this section we first study the structural properties of maximum colorful cycles and develop connections between maximum colorful cycles and maximum colorful matchings. Next, we will use those properties to design an efficient algorithm for finding a maximum colorful cycle.

Lemma 1. *Let G^c be a vertex-colored threshold graph. Then $C_m - 1 \leq C_c \leq C_m + 1$.*

Proof. We first show that $C_c \geq C_m - 1$. Let M be a maximum colorful matching in G^c with C_m different colors. Note that $|M|$ is at least C_m. Since each edge must contain at least one dominating vertex, choose one dominating vertex from each edge of M: denote those vertices by $x_1, x_2, \ldots, x_{|M|}$, such that x_j was added earlier than x_{j+1}, for $1 \leq j \leq |M| - 1$. Let z_j be the neighbor of x_j in the matching M, for $1 \leq j \leq |M|$. Note that z_j can be an isolated vertex or a dominating vertex. By the order of x-vertices, $N(x_1) \subseteq N(x_2) \subseteq \ldots \subseteq N(x_{|M|})$. Thus, (x_j, z_{j-1}) must be an edge in $E(G^c)$, for $1 \leq j \leq |M|$. As a result, $C' = (x_1, z_1, x_2, z_2, \ldots, x_{|M|-1}, z_{|M|-1}, x_{|M|}, x_1)$ is a cycle containing all vertices in M, except for $z_{|M|}$. The number of colors in the cycle C' is at least $C_m - 1$ since we remove $z_{|M|}$, and thus $C_c \geq C_m - 1$.

Next, suppose by contradiction that $C_c \geq C_m + 2$. Let $K = (v_{i_1}, v_{i_2}, \ldots, v_{i_k}, v_{i_1})$ be a cycle with C_c colors. Let $k = 2t$ if k is even and $k = 2t + 1$ if k is odd. Now, the matching $M = \{(v_{i_1}, v_{i_2}), (v_{i_3}, v_{i_4}), \ldots, (v_{i_{2t-1}}, v_{i_{2t}})\}$ has $|\mathcal{C}(M)| \geq C_m + 1$ colors, a contradiction. Thus $C_c \leq C_m + 1$ and this completes our proof. □

The following observation allows us to reduce the search space for isolated vertices of maximum colorful cycles:

Lemma 2. *Any maximum colorful cycle can be reduced to another maximum colorful cycle in which any isolated vertex has a color different from the colors of other vertices.*

By Lemma 2, we can restrict our attention only to maximum colorful cycles where each isolated vertex has a distinct color. We now introduce some new terminology. Let $\mathcal{C}_1 := \mathcal{C}(Y) \setminus \mathcal{C}(X) = \{c_{11}, c_{12}, \ldots, c_{1k_1}\}$ be the colors in Y but not in X. Denote $\mathcal{C}_2 := \mathcal{C}(X) = \{c_{21}, c_{22}, \ldots, c_{2k_2}\}$ the set of colors in X. By these definitions, the numbers of colors in $\mathcal{C}_1$ and $\mathcal{C}_2$ are k_1 and k_2, respectively. For each color c_{1i} in $\mathcal{C}_1$, let $\min[c_{1i}]$ be the index of the first vertex in Y with color c_i, i.e., $c(v_{\min[c_{1i}]}) = c_{1i}$ and $c(v_j) \neq c_{1i}$ for every $v_j \in Y$ with $j < \min[c_{1i}]$. Without loss of generality, suppose that $1 < \min[c_{11}] < \min[c_{12}] < \ldots < \min[c_{1k_1}]$. Similarly, for each color c_{2i} in $\mathcal{C}_2$, let $\max[c_{2i}]$ be the index of the last vertex in X such that $c(v_{\max[c_{2i}]}) = c_{2i}$ and $c(v_j) \neq c_{2i}$ for every $v_j \in X$ with $j > \max[c_{2i}]$. Without loss of generality, suppose that $\max[c_{21}] > \max[c_{22}] > \ldots > \max[c_{2k_2}]$. Moreover, for each maximum colorful cycle K, let X_K and Y_K be the sets of dominating vertices and isolated vertices in K, respectively, and let us denote their sets of colors by $\mathcal{C}(X_K)$ and $\mathcal{C}(Y_K)$.

We now consider three different cases, depending on whether $C_c = C_m + 1$, $C_c = C_m$ or $C_c = C_m - 1$.

3.1 Case 1: $C_c = C_m + 1$.

Lemma 3. *Suppose that $C_c = C_m + 1$, then any maximum colorful cycle must contain all colors of the given graph G^c.*

Case 1.1: *There exists a maximum colorful cycle K with some edge connecting two dominating vertices.*

Lemma 4. *There exists another maximum colorful cycle K' whose set of vertices $V(K') = V(X) \cup \{v_{\min[c_{11}]}, v_{\min[c_{12}]}, \ldots, v_{min[c_{1k_1}]}\}$.*

Proof. Let (u, v) be an edge of K such that both u and v are dominating vertices. Recall that any two dominating vertices are connected to each other. Therefore in the case that there exists some dominating vertices which are not in K then we can include them into K by adding into K a path from u to v containing all these dominating vertices and remove the edge (u, v) from K. By doing so it is possible to obtain another maximum colorful cycle K' containing the set $V(X)$. Now since K' contains all colors of the original graph (by Lemma 3) and each isolated vertex of K' (also K) has distinct color (by Lemma 2), we obtain that all colors of $\mathcal{C}_1$ must appears exactly once in K' and the number of isolated vertices of K' is equal to k_1. Now observe that for any two isolated vertices w and t such that w was added earlier than t in G^c then $N(w) \supseteq N(t)$. This allows to replace all isolated vertices of K' with distinct colors by the set of vertices $\{v_{\min[c_{11}]}, v_{\min[c_{12}]}, \dots, v_{\min[c_{1k_1}]}\}$. So it is possible to obtain another maximum colorful cycle K' with its set of vertices as $V(X) \cup \{v_{\min[c_{11}]}, v_{\min[c_{12}]}, \dots, v_{\min[c_{1k_1}]}\}$. □

Case 1.2: *For any maximum colorful cycle, there is no edge of this cycle that connects two dominating vertices.*

For each isolated vertex v, let $X^+(v)$ and $X^-(v)$ be the sets of dominating vertices added to G^c after and before v, respectively. Similarly, let $Y^+(v)$ and $Y^-(v)$ be the sets of isolated vertices added after and before v, respectively. Our maximum colorful cycle in this case will use a special vertex to reduce.

Lemma 5. *There exists exactly one isolated vertex v^* such that*

$$|X^+(v^*)| + |\mathcal{C}(X^+(v^*))| = C_m + 1$$

Moreover, the set of dominating vertices $X_K = X^+(v^*)$ and the number of isolated vertices $|Y_K| = |X^+(v^*)|$.

Proof. If we consider isolated vertices v in the order of the construction of the threshold graph G^c then the value of the sum $|X^+(v)| + |\mathcal{C}(X^+(v))|$ will strictly decrease. Thus there exists at most one isolated vertex v^* satisfying the lemma equality.

In the remainder of the proof, we will show the existence of such vertex v^*. Let v be the *first* isolated vertex in K (in the order of the construction of the threshold graph G^c). We will prove that v is v^*.

We claim that all vertices of $X^+(v)$ are in K and no vertex of $X^-(v)$ is in K.

$X^+(v) \subset K$. Assume that there exists a dominating vertex $u \in X^+(v)$ and u is not in K. As u was added after v, u is connected with v by an edge. Let w be a neighbor of v on K. Since v is an isolated vertex we must have that w is a dominating vertex. Now we remove the edge (v, w) from K and add two edges (v, u) and (u, w) on K then one obtains another maximum colorful cycle in which the edge (u, w) connects two dominating vertices (contradiction to the assumption of Case 1.2). Thus u must be in K.

$X^-(v) \cap K = \emptyset$. Assume that there exists a dominating vertex $t \in X^-(v)$ and t is also in K. Let z be a neighbor of t in K then z must be an isolated vertex. Therefore, z must be added earlier than t (by the construction of threshold graphs). Thus z must be added earlier than v, a contradiction since v is the first isolated vertex in K. So t must not be in K.

Hence, the claim follows.

By the claim, we have $X_K = X^+(v)$. As there is no edge of K connecting two dominating vertices, each edge must have an endpoint as dominating vertex and another endpoint as isolated vertex. So $|X_K| = |Y_K|$. From that the number of isolated vertices of K (i.e. $|Y_K|$) is equal to $|X^+(v)|$.

By Lemma 2, each isolated vertex in this cycle has a distinct color, so the number of colors of all isolated vertices in K is $|X^+(v)|$. Moreover, the number of colors of all dominating vertices of K is $|C(X^+(v))|$. Therefore we obtain that $|X^+(v)| + |C(X^+(v))|$ equals the number of colors in K, which is $C_c = C_m + 1$ by the case assumption. Hence, the lemma equality holds at v, i.e., $|X^+(v)| + |\mathcal{C}(X^+(v))| = C_m + 1$. □

Note that one can detect this vertex v^* efficiently by computing C_m and by checking the above identity over all vertices.

Let $\{c'_{11}, c'_{12}, \ldots, c'_{1k'}\} = \mathcal{C}(Y) \setminus \mathcal{C}(X^+(v^*))$ be the set of colors in Y but not in $X^+(v^*)$. As we did before, let us define by $v_{\min[c'_{1i}]}$ the first isolated vertex in G^c with color c'_{1i}. Again, without loss of generality, assume that $\min[c'_{11}] < \min[c'_{12}] < \ldots < \min[c'_{1k'}]$. Now we are ready to show the main structural property of a maximum colorful cycle in this case.

Lemma 6. *Let v^* be the unique vertex such that $|X^+(v^*)| + |C(X^+(v^*))| = C_m + 1$ and let K be a maximum colorful cycle. Then, there exists another maximum colorful cycle K' where $V(K') = X^+(v^*) \cup \{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \ldots, v_{\min[c'_{1|X^+(v^*)|}]}\}$.*

Proof. By Lemma 5, the number of isolated vertices of K (i.e. $|Y_K|$) equals $|X^+(v^*)|$ and each vertex in Y_K has a distinct color. Observe that $\mathcal{C}(Y_K) \subseteq \mathcal{C}(Y) \setminus \mathcal{C}(X^+(v^*))$. This follows from the fact that if $\mathcal{C}(Y_K) \cap \mathcal{C}(X^+(v^*)) \neq \emptyset$ then the total number of colors in K is

$$\begin{aligned} |\mathcal{C}(Y_K) \cup \mathcal{C}(X_K)| &= |\mathcal{C}(Y_K) \cup \mathcal{C}(X^+(v^*))| < |\mathcal{C}(Y_K)| + |\mathcal{C}(X^+(v^*))| \\ &\leq |X^+(v^*)| + |\mathcal{C}(X^+(v^*))| = C_m + 1 = C_c \end{aligned}$$

where the equalities are due to Lemma 5. This contradicts the fact that K is a maximum colorful cycle. Therefore, the observation holds true and so $|X^+(v^*)| \leq k'$.

Recall that for any two isolated vertices w and t such that w was added earlier than t in G^c then $N(w) \supseteq N(t)$. Hence, by replacing $|X^+(v^*)|$ isolated vertices in K by vertices $\{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \ldots, v_{\min[c'_{1|X^+(v^*)|}]}\}$, one gets another cycle K' with the same number of colors as K. This vertex replacing procedure can be done since $|X^+(v^*)| \leq k'$. Note that K' and K may have different sets of colors but their cardinals are the same. Thus, we have a maximum colorful cycle K' where $V(K') = X^+(v^*) \cup \{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \ldots, v_{\min[c'_{1|X^+(v^*)|}]}\}$. □

Since v^* and the sets of vertices $X^+(v^*)$ and $\{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \ldots, v_{\min[c'_{1|X^+(v^*)|}]}\}$ can be immediately identified, in the case that there exists a maximum colorful cycle then we can use the algorithm in [16] to construct a Hamiltonian cycle consisting of all these vertices.

3.2 Case 2: $C_c = C_m$.

The following lemma helps to limit the search space of the set of colors of dominating vertices of a maximum colorful cycle.

Lemma 7. *For any maximum colorful cycle K, there exists at most one dominating vertex v such that $v \notin K$ and $c(v) \notin \mathcal{C}(K)$.*

By this lemma, the set of colors $\mathcal{C}(X_K)$ is either $\mathcal{C}(X)$ or $\mathcal{C}(X) \setminus c$ for some color c. We distinguish two corresponding sub-cases.

Case 2.1: *There exists a maximum colorful cycle K such that there exists exactly one dominating vertex v^{**} such that $v^{**} \notin K$ and $c(v^{**}) \notin \mathcal{C}(K)$.*

Denote $\{c'_{11}, c'_{12}, \ldots, c'_{1k'}\} = (\mathcal{C}(Y) \setminus \mathcal{C}(X)) \cup \{c(v^{**})\}$. Similarly as previously, let $v_{\min[c'_{1i}]}$ be the first isolated vertex with the color c'_{1i} in G^c. Without loss of generality, suppose that $\min[c'_{11}] < \min[c'_{12}] < \ldots < \min[c'_{1k'}]$. Note that, in contrast to the previous case, the vertex v^{**} can not be immediately identified. However, our algorithm will loop over all vertices by considering each as v^{**}. The following lemma helps to replace vertices to obtain another maximum colorful cycle from a maximum colorful cycle in this situation.

Lemma 8. *Let K be a maximum colorful cycle of G^c and v^{**} be a dominating vertex such that $v^{**} \notin K$ and $c(v^{**}) \notin \mathcal{C}(K)$. Then, there exists another colorful cycle K' such that the set of vertices $V(K') = V(X) \setminus \{v^{**}\} \cup \{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \ldots, v_{\min[c'_{1\ell}]}\}$ where $\ell = C_m - |\mathcal{C}(X)| + 1$.*

Proof. By the case assumption, $|\mathcal{C}(X_K)| = |\mathcal{C}(X)| - 1$. Since $|\mathcal{C}(K)| = C_m$, we obtain that $|\mathcal{C}(Y_K)| = C_m - |\mathcal{C}(X)| + 1$. Moreover, by Lemma 2, the color of any isolated vertex of K is different to other vertex's color. So $\mathcal{C}(Y_K) \subseteq \mathcal{C}(Y) \setminus \mathcal{C}(X_K)$. Therefore, $C_m - |\mathcal{C}(X)| + 1 \leq |\mathcal{C}(Y) \setminus \mathcal{C}(X) \cup \{c(v^{**})| = k'$. Recall the observation that for any two isolated vertices w and t, if w was added earlier than t in G^c then $N(w) \supseteq N(t)$. Denote $\ell = C_m - |\mathcal{C}(X)| + 1$. By replacing ℓ vertices of K by vertices $\{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \ldots, v_{\min[c'_{1\ell}]}\}$, one obtains another maximum colorful cycle K' with the same number of colors. Note that $\ell < k'$ so the replacing procedure can always be done. Now the set of vertices $V(K') = V(X) \setminus \{v^{**}\} \cup \{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \ldots, v_{\min[c'_{1\ell}]}\}$ as required by the lemma. □

Case 2.2: *For any maximum colorful cycle K, there does not exist any dominating vertex v such that $v \notin K$ and $c(v) \notin C(K)$.*

Recall that $\mathcal{C}_2 := \{c_{21}, c_{22}, \ldots, c_{2k_2}\}$ is the set of colors in X. Let $\max[c_{2i}]$ be the index of the last vertex in X such that $c(v_{\max[c_{2i}]}) = c_{2i}$ and $c(v_j) \neq c_{1i}$ for every $v_j \in X$ with $j > \max[c_{2i}]$. Without loss of generality, suppose that $\max[c_{21}] > \max[c_{22}] > \ldots > \max[c_{2k_2}]$. Let $X_t(G^c)$ be the set of t last dominating vertices of the set $V(X) \setminus \{v_{\max[c_{21}]}, v_{\max[c_{22}]}, \ldots, v_{\max[c_{2k_2}]}\}$. Now the following lemma helps to reduce a maximum colorful cycle to another maximum colorful cycle which is easier to find.

Lemma 9. *Let K be a maximum colorful cycle of G^c. Then, there exists another colorful cycle K' where $V(K') = X_t(G^c) \cup \{v_{\max[c_{21}]}, v_{\max[c_{22}]}, \ldots, v_{\max[c_{2k_2}]}\}$ $\cup\{v_{\min[c_{11}]}, v_{\min[c_{12}]}, \ldots, v_{\min[c_{1\ell}]}\}$ where $\ell = C_m - |\mathcal{C}(X)|$ and $t = |V(K)| - k_2 - C_m + |\mathcal{C}(X)|$.*

By Lemma 9, given the value $|V(K)|$, all elements in the cycle K' are immediately identified. Therefore in our final algorithm we will vary the value of $|V(K)|$ to find the maximum colorful cycle if this case holds.

3.3 Case 3: $C_c = C_m - 1$.

In this case, it is possible to obtain easily a cycle with $C_m - 1$ colors from any colorful matching, based on the first part of the proof of Lemma 1.

3.4 Algorithm for threshold graphs.

Based on the structural properties of maximum colorful cycles according to different cases, we derive the following algorithm for finding a maximum colorful cycle. The algorithm makes use of the algorithm computing a maximum colorful matching [5] and the algorithm computing a Hamiltonian cycle in threshold graphs [16].

Algorithm 1 Maximum colorful cycle in vertex-colored threshold graphs.

$C_m \leftarrow$ the number of colors of a maximum colorful matching (using algorithm [5])
$\mathcal{C}_1 := \mathcal{C}(Y) \setminus \mathcal{C}(X) = \{c_{11}, c_{12}, \dots, c_{1k_1}\}$ and $\mathcal{C}_2 := \mathcal{C}(X) = \{c_{21}, c_{22}, \dots, c_{2k_2}\}$
if $\exists$ a Hamiltonian cycle K of $V(X) \cup \{v_{\min[c_{11}]}, \dots, v_{\min[c_{1k_1}]}\}$ **then** `# Case 1.1`
    return K as the maximum colorful cycle `# Lemma 4`
else `# Case 1.2`
    $v^* \leftarrow$ the unique vertex satisfying $|X^+(v^*)| + |C(X^+(v^*))| = C_m + 1$
    $X^+(v^*) \leftarrow$ set of dominating vertices added to the graph after v^*
    $\{c'_{11}, c'_{12}, \dots, c'_{1k'}\} \leftarrow \mathcal{C}(Y) \setminus \mathcal{C}(X^+(v^*))$
    if $\exists$ a Hamiltonian cycle K of $X^+(v^*) \cup \{v_{\min[c_{11}]}, v_{\min[c_{12}]}, \dots, v_{\min[c_{1k_1}]}\}$ **then**
        return K as the maximum colorful cycle `# Lemma 6`
    end if
end if
for $v^{**} \in V(G^c)$ **do** `# Case 2.1`
    $\{c'_{11}, c'_{12}, \dots, c'_{1k'}\} \leftarrow \mathcal{C}(Y) \setminus \mathcal{C}(X) \cup \{c(v^{**})\}$ and $\ell \leftarrow C_m - |\mathcal{C}(X)| + 1$
    if $\exists$ a Hamiltonian cycle K of $V(X) \setminus \{v^{**}\} \cup \{v_{\min[c'_{11}]}, v_{\min[c'_{12}]}, \dots, v_{\min[c'_{1\ell}]}\}$ **then**
        return K as the maximum colorful cycle `# Lemma 8`
    end if
end for
for $0 \le t \le |V(X) \setminus \{v_{\max[c_{21}]}, v_{\max[c_{22}]}, \dots, v_{\max[c_{2k_2}]}\}|$ **do** `# Case 2.2`
    $X_t(G^c) \leftarrow t$ last vertices in $V(X) \setminus \{v_{\max[c_{21}]}, v_{\max[c_{22}]}, \dots, v_{\max[c_{2k_2}]}\}$
    if $\exists$ a Hamiltonian cycle K of $X_t(G^c) \cup \{v_{\max[c_{21}]}, v_{\max[c_{22}]}, \dots, v_{\max[c_{2k_2}]}\}$ $\cup \{v_{\min[c_{11}]}, v_{\min[c_{12}]}, \dots, v_{\min[c_{1\ell}]}\}$ where $\ell = C_m - |\mathcal{C}(X)|$ **then**
        return K as the maximum colorful cycle `# Lemma 9`
    end if
end for
return K as a maximum colorful cycle constructed from any maximum colorful matching based on Lemma 1 `# Case 3`

Theorem 2. *Algorithm 1 computes a maximum colorful cycle of G^c in time $O(\max\{|\mathcal{C}| \cdot M(m,n), n(n+m)\})$ where $|\mathcal{C}|$ is the number of colors in G^c and $M(m,n)$ is the time for finding a maximum matching in a general graph with m edges and n vertices.*

4 An algorithm for bipartite chain graphs

A bipartite graph $G = (X, Y, E)$ is said to be a *bipartite chain* graph if its vertices can be linearly ordered such that $N(x_1) \supseteq N(x_2) \supseteq \ldots \supseteq N(x_{|X|})$. As a consequence, we also immediately obtain a linear ordering over Y such that $N(y_1) \supseteq N(y_2) \supseteq \ldots \supseteq N(y_{|Y|})$. It is known that these orderings over X and Y can be computed in $O(n)$ time. Here we will look for a maximum colorful cycle in a vertex-colored bipartite chain graph $G^c = (X, Y, E)$.

Algorithm 2 Maximum colorful cycle in vertex-colored bipartite chain graphs.

$C_m \leftarrow$ the number of colors of a maximum colorful matching (using algorithm [5])
for $C_m \geq C_c \geq C_m - 2$ **do**
  for $1 \leq m \leq |X|, 1 \leq n \leq |Y|$ **do**
    $X_{m,n} \leftarrow \{x_i \in X | c(x_i) \notin \mathcal{C}(\{x_1, x_2, \ldots, x_m\}) \cup \mathcal{C}(\{y_1, y_2, \ldots, y_n\})\}$
    $Y_{m,n} \leftarrow \{y_j \in Y | c(y_j) \notin \mathcal{C}(\{x_1, x_2, \ldots, x_m\}) \cup \mathcal{C}(\{y_1, y_2, \ldots, y_n\})\}$
    Denote $\mathcal{C}(X_{m,n}) := \{c_{11}, c_{12}, \ldots, c_{1k_1}\}$ and $\mathcal{C}(Y_{m,n}) := \{c_{21}, c_{22}, \ldots, c_{2k_2}\}$.
    for $0 \leq \ell \leq C_c$ **do**
      $\ell' \leftarrow \max\{C_c - \ell - |\mathcal{C}(x_1, x_2, ..., x_m)| - |\mathcal{C}(y_1, y_2, ..., y_n)|, 0\}$
      $X^{\ell}_{m,n} \leftarrow \{x_{\min[c_{11}]}, x_{\min[c_{12}]}, \ldots, x_{\min[c_{1\ell}]}\}$ the set of ℓ first vertices (in the ordering of x-vertices) with distinct colors in $X_{m,n}$.
      $Y^{\ell'}_{m,n} \leftarrow \{y_{\min[c_{21}]}, y_{\min[c_{22}]}, \ldots, y_{\min[c_{2\ell'}]}\}$ the set of ℓ' first vertices (in the ordering of y-vertices) with distinct colors in $Y_{m,n}$
      if $\exists$ a Hamiltonian cycle K of $\{x_1, x_2, \ldots, x_m\} \cup X^{\ell}_{m,n} \cup \{y_1, y_2, \ldots, y_n\} \cup Y^{\ell'}_{m,n}$ **then**
        return K as the maximum colorful cycle
      end if
    end for
  end for
end for

Theorem 3. *Algorithm 2 computes a maximum colorful cycle of G^c in $O(|\mathcal{C}| \cdot \max\{M(m,n), n^3\})$ where $M(m,n)$ is the best known complexity for finding a maximum matching in a general graph with m edges and n vertices.*

References

[1] Akbari, S., Liaghat, V., Nikzad, A.: Colorful paths in vertex coloring of graphs. the electronic journal of combinatorics 18(1), P17 (2011)

[2] Akiyama, T., Nishizeki, T., Saito, N.: Np-completeness of the hamiltonian cycle problem for bipartite graphs. Journal of Information Processing 3(2), 73–76 (1979)

[3] Bertossi, A.A.: Finding hamiltonian circuits in proper interval graphs. Information Processing Letters 17(2), 97–101 (1983)

[4] Bruckner, S., Hüffner, F., Komusiewicz, C., Niedermeier, R.: Evaluation of ILP-based approaches for partitioning into colorful components. In: International Symposium on Experimental Algorithms. pp. 176–187 (2013)

[5] Cohen, J., Manoussakis, Y., Pham, H., Tuza, Z.: Tropical matchings in vertex-colored graphs. In: Latin and American Algorithms, Graphs and Optimization Symposium (2017)

[6] Cohen, J., Italiano, G., Manoussakis, Y., Thang, N.K., Pham, H.P.: Tropical paths in vertex-colored graphs. In: Proc. 11th Conference on Combinatorial Optimization and Applications (2017, to appear)

[7] Corel, E., Pitschi, F., Morgenstern, B.: A min-cut algorithm for the consistency problem in multiple sequence alignment. Bioinformatics 26(8), 1015–1021 (2010)

[8] Fellows, M.R., Fertin, G., Hermelin, D., Vialette, S.: Upper and lower bounds for finding connected motifs in vertex-colored graphs. Journal of Computer and System Sciences 77(4), 799–811 (2011)

[9] Foucaud, F., Harutyunyan, A., Hell, P., Legay, S., Manoussakis, Y., Naserasr, R.: Tropical homomorphisms in vertex-coloured graphs. Discrete Applied Mathematics (to appear)

[10] Garey, M.R., Johnson, D.S., Stockmeyer, L.: Some simplified NP-complete problems. In: Proc. 6th Symposium on Theory of Computing. pp. 47–63 (1974)

[11] Golumbic: Algorithmic graph theory and perfect graphs. Annals of Discrete Mathematics 57 (2009)

[12] Ibarra, L.: The clique-separator graph for chordal graphs. Discrete Applied Mathematics 157(8), 1737–1749 (2009)

[13] Ibarra, L.: A simple algorithm to find hamiltonian cycles in proper interval graphs. Information Processing Letters 109(18), 1105–1108 (2009)

[14] Li, H.: A generalization of the Gallai–Roy theorem. Graphs and Combinatorics 17(4), 681–685 (2001)

[15] Lin, C.: Simple proofs of results on paths representing all colors in proper vertex-colorings. Graphs and Combinatorics 23(2), 201–203 (2007)

[16] Mahadev, Peled: Longest cycles in threshold graphs. Discrete Mathematics 135(1-3), 169–176 (1994)

[17] Marx, D.: Graph colouring problems and their applications in scheduling. Periodica Polytechnica Electrical Engineering 48(1-2), 11–16 (2004)

[18] Micali, S., Vazirani, V.V.: An $O(\sqrt{|V|}|E|)$ algorithm for finding maximum matching in general graphs. In: Proc. 21st Symposium on Foundations of Computer Science. pp. 17–27 (1980)

[19] Uehara, R., Valiente, G.: Linear structure of bipartite permutation graphs and the longest path problem. Information Processing Letters 103(2), 71–77 (2007)