Also available at http://amc.imfm.si
ISSN 1855-3966 (printed edn.), ISSN 1855-3974 (electronic edn.)
ARS MATHEMATICA CONTEMPORANEA 6 (2013) 83–88

On the minimum rainbow subgraph number of a graph

Ingo Schiermeyer

Institut für Diskrete Mathematik und Algebra, Technische Universität Bergakademie Freiberg, 09596 Freiberg, Germany

Received 19 October 2011, accepted 9 April 2012, published online 1 June 2012

Abstract

We consider the MINIMUM RAINBOW SUBGRAPH problem (MRS): Given a graph G whose edges are coloured with p colours. Find a subgraph $F \subseteq G$ of minimum order and with p edges such that each colour occurs exactly once. This problem is NP-hard and APX-hard.

For a given graph G and an edge colouring c with p colours we define the *rainbow subgraph number* $rs(G, c)$ to be the order of a minimum rainbow subgraph of G with size p. In this paper we will show lower and upper bounds for the rainbow subgraph number of a graph.

Keywords: Edge colouring, rainbow subgraph.

Math. Subj. Class.: 05C15, 05C35

1 Introduction and motivation

We use [2] for terminology and notation not defined here and consider finite and simple graphs only.

Our research was motivated by the following problem from bioinformatics. The problem data consist in a set $\mathcal{G}$ of p *genotypes* $g_1, g_2, \ldots, g_p$ corresponding to p individuals in a population. Each genotype g is a vector with entries in $\{0, 1, 2\}$. Each position where a 2 appears is called *ambiguous* position. For a genotype g we have to determine a pair of haplotypes h_P and h_M (h_P stands for the paternal haplotype and h_M stands for the maternal haplotype), which are binary vectors such that $g = h_P \oplus h_M$.

Given two haplotypes h' and h'', their sum is defined as the vector $g = h' \oplus h''$ with $g[i] = 0$, if $h'[i] = h''[i] = 0$, $g[i] = 1$, if $h'[i] = h''[i] = 1$ and $g[i] = 2$, if $h'[i] \neq h''[i]$.

We say that a set $\mathcal{H}$ of haplotypes *resolves* $\mathcal{G}$ if for every $g \in \mathcal{G}$ there exist $h_1, h_2 \in \mathcal{H}$ such that $g = h_1 \oplus h_2$. Given a set $\mathcal{G}$ of genotypes, the *haplotyping problem* consists

E-mail address: Ingo.Schiermeyer@tu-freiberg.de (Ingo Schiermeyer)

in finding a set $\mathcal{H}$ of haplotypes that resolves $\mathcal{G}$. In the $Pure\ Parsimony\ Haplotyping\ problem$ ($PPH\ problem$) we are interested in finding a set $\mathcal{H}$ of smallest possible cardinality. If each genotype has at most k ambiguous positions, then we denote this problem by PPH(k). The $PPH\ problem$ has been studied in ([3],[4],[7],[9]).

Matos Camacho et al. [8] have shown that the PPH(k) can be transformed to a graph problem, the MINIMUM RAINBOW SUBGRAPH problem (MRS). Note that this edge-colouring need not be proper.

Definition 1.1 (Rainbow subgraph)**.**
Let G be a graph with an edge-colouring. A subgraph H of G is called *rainbow subgraph* if H does not contain two edges of the same colour.

Definition 1.2 (Minimum Rainbow Subgraph problem (MRS))**.**
Given a graph G, whose edges are coloured with p colours, find a subgraph $F \subseteq G$ of minimum order and with p edges such that each colour occurs exactly once.

For a set $\mathcal{G}$ of p genotypes $g_1, g_2, \ldots, g_p$ we will use p colours $1, 2, \ldots, p$. For each haplotype we introduce a vertex. If two haplotypes h' and h'' resolve a genotype g_i ($g_i = h' \oplus h''$), then the corresponding vertices will be joined by an edge which receives colour i. If a genotype is resolved by two identical haplotypes, then the corresponding vertex is joined by an edge which is called a $loop$.

In this way we construct a graph G, whose edges are coloured with p colours. Note that this is a proper edge colouring (no vertex is incident with two edges of the same colour), since a haplotype h can be used at most once in a pair of haplotypes, which resolves a genotype g. Furthermore, every set $\mathcal{H}$ of haplotypes that resolves $\mathcal{G}$ corresponds to a rainbow subgraph F of G.

It has been shown in [8] that a graph G containing loops can be transformed into a graph G' without loops. Hence in the following we may assume that all graphs have no loops.

Matos Camacho et al. [8] proved the MRS problem to be NP-hard and APX-hard. In [5] it has been shown that the MRS problem remains NP-hard and APX-hard even for graphs with maximum degree 2.

Remark: If we do not consider edge colourings, the analogous problem is known as the $(t, f(t))$ dense subgraph problem ($(t, f(t))$-DSP), which asks whether there is a t-vertex subgraph of a given graph G which has at least $f(t)$ edges. When $f(t) = \binom{t}{2}$, $(t, f(t))$-DSP is equivalent to the well-known t-clique problem (cf. [1]).

2 Lower bounds for the rainbow subgraph number

Definition 2.1. Let G be a graph and c be its edge colouring with p colours. The *rainbow subgraph number* of G (with respect to the colouring c) is defined as the order of its minimum rainbow subgraph of size p, and denoted by $rs(G, c)$ (or $rs(G)$, when the colouring c is clear from the context).

Improved lower bounds for the rainbow subgraph number $rs(G)$ will be of major importance for the design of approximation algorithms with better approximation ratios for the MRS problem (cf. [8, 5]). So far nothing better than the trivial lower bound $rs(G) \geq \frac{2p}{\Delta(G)}$ is known. We can improve this lower bound by counting the number of distinct colours among all edges incident to a vertex.

Definition 2.2. Given an edge colouring of a graph G with colours $1, 2, \dots, p$, we define $c(e) = i$, if the edge e has colour i for $1 \le i \le p$.
Let $cd(v)$ (*colour degree*) denote the number of distinct colours among all edges incident to the vertex v and let $cd(i) = \max\{cd(v) \mid v \in V(G)$ has an incident edge with colour $i\}$ be the *maximum colour degree* for every colour $i, 1 \le i \le p$.

Using the maximum colour degrees for all colours we can show the following improved lower bound.

Proposition 2.3. *Let G be a graph, whose edges are coloured with p colours. Then*

$$rs(G) \ge \sum_{i=1}^{p} \frac{2}{cd(i)} \ge \frac{2p}{\Delta(G)}.$$

Proof. Let F be a minimum rainbow subgraph of order $k = rs(G)$. Then

$$rs(G) = k = \sum_{v \in V(F)} \frac{d_F(v)}{d_F(v)} = \sum_{e=uw, e \in E(F)} \frac{1}{d_F(u)} + \frac{1}{d_F(w)} \ge \sum_{i=1}^{p} \frac{2}{cd(i)} \ge \frac{2p}{\Delta(G)}.$$

□

The following example shows that this bound is sharp and improves the lower bound of $\frac{2p}{\Delta(G)}$ significantly.

Example 2.4. For $p \ge 4$ and $\Delta \ge 2$ let $G = K_{1,\Delta} + C_{p-1}$ (where $G + H$ denotes the disjoint union of two graphs G and H). All edges of the cycle C_{p-1} are coloured distinctly, say with colours $1, 2, \dots, p-1$, and all edges of $K_{1,\Delta}$ are coloured with colour p. Then $rs(G) = p + 1 = p - 1 + 2 = \sum_{i=1}^{p} \frac{2}{cd(i)} > \frac{2p}{\Delta(G)}$.

We can further improve this lower bound by counting the number of distinct colours among all edges incident to the endvertices of an edge. For this purpose we define $q(i) = \min\{\frac{1}{cd(u)} + \frac{1}{cd(w)} \mid uw \in E(G)$ and $c(uw) = i\}$.

Proposition 2.5. *Let G be a graph, whose edges are coloured with p colours. Then*

$$rs(G) \ge \sum_{i=1}^{p} q(i) \ge \sum_{i=1}^{p} \frac{2}{cd(i)} \ge \frac{2p}{\Delta(G)}.$$

Proof. Let F be a minimum rainbow subgraph of order $k = rs(G)$. For every colour $i, 1 \le i \le p$, let $u_i w_i$ be an edge such that $\frac{1}{cd(u_i)} + \frac{1}{cd(w_i)} = q(i)$. If $uw \in E(F)$ is an edge with $c(uw) = i$, then $\frac{1}{cd(u)} + \frac{1}{cd(w)} \ge \frac{1}{cd(u_i)} + \frac{1}{cd(w_i)} = q(i) \ge 2 \cdot \frac{1}{\max\{cd(u_i), cd(w_i)\}} \ge \frac{2}{cd(i)}$.
Therefore,

$$rs(G) = k = \sum_{v \in V(F)} \frac{d_F(v)}{d_F(v)} = \sum_{e=uw, e \in E(F)} \frac{1}{d_F(u)} + \frac{1}{d_F(w)} \ge$$

$$\sum_{i=1}^{p} q(i) \ge \sum_{i=1}^{p} \frac{2}{cd(i)} \ge \frac{2p}{\Delta(G)}.$$

□

The following example shows that this bound is sharp and improves the previous two lower bounds significantly.

Example 2.6. Let $G \cong K_{1,p}$ for some $p \geq 2$. Let the edges of G be coloured with p colours. Then $cd(i) = p$ and $q(i) = 1 + \frac{1}{p}$ for $1 \leq i \leq p$. Thus $rs(G) = p + 1 = p \cdot (1 + \frac{1}{p}) = \sum_{i=1}^{p} q(i) > 2 = p \cdot \frac{2}{p} = \sum_{i=1}^{p} \frac{2}{cd(i)} = \frac{2p}{\Delta(K_{1,p})}$.

3 Upper bounds for the rainbow subgraph number

First observe that the trivial upper bound $rs(G) \leq 2p$ is achieved if the rainbow subgraph F is a matching. This upper bound has been improved towards $rs(G) \leq 2p + 1 - \Delta(G)$ by Koch [6] for properly edge-coloured graphs and this bound is sharp. For instance, let $G = K_{1,\Delta} + (p - \Delta)K_2$, where $p \geq \Delta$, and all edges of G are coloured distinctly. Then $rs(G) = 2p + 1 - \Delta(G)$.

Similar to Brooks' Theorem (cf. [2]) we can characterize all graphs achieving this bound.

Theorem 3.1. *Let G be a graph with maximum degree $\Delta \geq 2$, whose edges are properly coloured with p colours. If $rs(G) = 2p + 1 - \Delta(G)$, then G has the following properties:*

1. *G contains a star $K_{1,\Delta}$ with center vertex v_0 and leaves $v_1, \ldots, v_\Delta$ and $G[N(v_0)]$ is edgeless. Let $c(v_0 v_i) = i$ for $1 \leq i \leq \Delta$ and $H_0 \cong G[N[v_0]]$.*
2. *If $p > \Delta$, then let H_i be the subgraph spanned by the edges with colour i for $\Delta + 1 \leq i \leq p$. The subgraphs $H_{\Delta+1}, H_{\Delta+2}, \ldots, H_p$ are pairwise vertex-disjoint and $V(H_0) \cap V(H_i) = \emptyset$ for $\Delta + 1 \leq i \leq p$.*
3. *$E(H_i, H_j) = \emptyset$ for $\Delta + 1 \leq i < j \leq p$ (where $E(H_i, H_j)$ is the set of all edges having one vertex in $V(H_i)$ and the other vertex in $V(H_j)$).*
4. *$E(v_i, H_j) = \emptyset$ for $1 \leq i \leq \Delta$ and $\Delta + 1 \leq j \leq p$ (where $E(v_i, H_j)$ is the set of all edges incident with v_i and a vertex in $V(H_j)$).*
5. *If $uv \in E(H_i)$ for some $\Delta + 1 \leq i \leq p$, then $N(u) \cap N(v) = \emptyset$.*
6. *$N(v_i) \cap N(v_j) = \emptyset$ for $v_i \in V(H_i), v_j \in V(H_j), \Delta + 1 \leq i < j \leq p$.*

Proof.

1. Suppose there is an edge $v_i v_j$ for some $1 \leq i < j \leq \Delta$. If $c(v_i v_j) = k$ for some k with $1 \leq k \leq \Delta, k \neq i, j$, then $rs(G) \leq (\Delta+1) - 1 + (2p - 2\Delta) = 2p - \Delta < 2p + 1 - \Delta$, a contradiction. If $c(v_i v_j) = k$ for some k with $\Delta + 1 \leq k \leq p$, then $rs(G) \leq (\Delta + 1) + (2p - 2\Delta - 2) = 2p - \Delta - 1 < 2p + 1 - \Delta$, a contradiction as well.
2. Suppose there are integers i, j with $\Delta + 1 \leq i < j \leq p$ and two adjacent edges e, f with $c(e) = i, c(f) = j$. Then $rs(G) \leq (\Delta + 1) + (2p - 2\Delta - 1) = 2p - \Delta < 2p + 1 - \Delta$, a contradiction. Suppose there are integers i, j with $1 \leq i \leq \Delta, \Delta + 1 \leq j \leq p$ and two adjacent edges e, f with $c(e) = i, c(f) = j$. Then $rs(G) \leq (\Delta + 1) + (2p - 2\Delta - 1) = 2p - \Delta < 2p + 1 - \Delta$, a contradiction as well.
3. Suppose there is an edge $v_i v_j$ with $v_i \in V(H_i), v_j \in V(H_j), \Delta + 1 \leq i < j \leq p$. Then $c(v_i v_j) = k$ for some $1 \leq k \leq \Delta$. Hence $rs(G) \leq (\Delta + 1) - 1 + (2p - 2\Delta) = 2p - \Delta < 2p + 1 - \Delta$, a contradiction.
4. Suppose there is an edge $v_i v_j$ for two vertices $v_i \in V(H_0)$ and $v_j \in V(H_j), \Delta + 1 \leq j \leq p$. Then $rs(G) \leq (\Delta+1) + (2p - 2\Delta - 1) = 2p - \Delta < 2p + 1 - \Delta$, a contradiction.

5. Suppose there is an edge $uv \in E(H_i)$ for some $\Delta+1 \leq i \leq p$ with $N(u) \cap N(v) \neq \emptyset$. By 3. and 4. we conclude that $N(u) \cap N(v) \cap V(H_0) = \emptyset$. Furthermore, for a vertex $w \in N(u) \cap N(v)$, we have $c(uw) = j, c(vw) = k$ for some $1 \leq j < k \leq \Delta$. Then $rs(G) \leq (\Delta+1) - 2 + (2p - 2\Delta + 1) = 2p - \Delta < 2p + 1 - \Delta$, a contradiction.

6. Suppose $N(v_i) \cap N(v_j) \neq \emptyset$ for two vertices $v_i \in V(H_i), v_j \in V(H_j), \Delta+1 \leq i < j \leq p$. By 3. and 4. we conclude that $N(v_i) \cap N(v_j) \cap V(H_0) = \emptyset$. Furthermore, for a vertex $w \in N(v_i) \cap N(v_j)$, we have $c(uw) = k, c(vw) = l$ for some $1 \leq k < l \leq \Delta$. Then $rs(G) \leq (\Delta+1) - 2 + (2p - 2\Delta + 1) = 2p - \Delta < 2p + 1 - \Delta$, a contradiction.

□

Another upper bound for the rainbow subgraph number follows from an approach presented in [8]. Observe that two adjacent edges of different colours together have three vertices, whereas two edges of different colours in a matching have four vertices. Based on this obervation the following algorithm has been proposed in [8].

Algorithm
Input: A graph G of order n whose edges are coloured with p colours

1. Construct a graph G' with $V(G') = \{v_1, v_2, \ldots, v_p\}$ (v_i corresponds to colour i) and $v_i v_j \in E(G')$ if there exist two adjacent edges $e, f \in E(G)$ with $c(e) = i$ and $c(f) = j$ ($c(x)$ denotes the colour of the edge x).

2. Now compute a maximum matching M of order $\beta(G')$ in G'. This can be done in polynomial time.

3. Next construct a graph H with $V(H) \subseteq V(G)$ as follows: For each matching edge of M choose two adjacent edges in G with these two colours. For each vertex of $V(G')$ not in M choose an edge in G with this colour. In this way we obtain a rainbow subgraph $H \subseteq G$ with $|E(H)| = p$.

Correctness of the algorithm: Edges of the matching correspond to pairs of adjacent edges in the original graph. Colours that are left out by this procedure are added greedily at the end.

Claim 3.2. $|V(H)| \leq 2p - \beta(G')$

Proof. For each matching edge of G' three vertices appear in H. Hence

$$|V(H)| \leq 3\beta(G') + 2(p - 2\beta(G')) = 2p - \beta(G')$$

□

Corollary 3.3. $rs(G) \leq 2p - \beta(G')$.

Acknowledgement

We thank the referees for some valuable comments.

References

[1] Y. Asahiro, R. Hassin and K. Iwama, Complexity of finding dense subgraphs, *Discrete Appl. Math.* **121** (2002), 15–26.

[2] J. A. Bondy and U.S.R. Murty, *Graph Theory*, Springer, London, 2008.

[3] D. Catanzaro and M. Labbé, The pure parsimony haplotyping problem: overview and computational advances, *International Transactions in Operational Research* **16** (2009), 561–584.

[4] D. Gusfield, Haplotype inference by pure Parsimony, in: *Proceedings ot the 14th annual conference on Combinatorial pattern matching (CPM'03)*, Springer, Berlin, Heidelberg, 2003, 144–155.

[5] J. Katrenič and I. Schiermeyer, Improved approximation bounds for the minimum rainbow subgraph problem, *Inf. Process. Lett.* **111** (2011), 110–114.

[6] M. Koch, *Das Population-Haplotyping-Problem: Graphentheoretische Ansätze*, Diploma thesis, TU Bergakademie Freiberg, 2008.

[7] G. Lancia, M. C. Pinotti and R. Rizzi, Haplotyping Populations by Pure Parsimony: Complexity of Exact and Approximation Algorithms, *INFORMS Journal on Computing* **16** (2004), 348–359.

[8] S. Matos Camacho, I. Schiermeyer and Z. Tuza. Approximation algorithms for the minimum rainbow subgraph problem, *Discrete Math.* **310** (2010), 2666–2670.

[9] L. S. Wang, Hapolotype inference by maximum parsimony, *Bioinformatics* **19** (2003), 1773–1780.