Theoretical Computer Science 533 (2014) 15–25

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Hardness and inapproximability of convex recoloring problems ☆

Manoel Campêlo [a,∗], Cristiana Huiban [b], Rudini M. Sampaio [c], Yoshiko Wakabayashi [d]

[a] Dep. Estatística e Matemática Aplicada, Universidade Federal do Ceará, Fortaleza, Brazil
[b] Centro de Informática, Universidade Federal de Pernambuco, Recife, Brazil
[c] Dep. Computação, Universidade Federal do Ceará, Fortaleza, Brazil
[d] Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 15 July 2013
Received in revised form 31 December 2013
Accepted 10 March 2014
Communicated by V.Th. Paschos

Keywords:
Convex recoloring
Complexity
Inapproximability
Linear-time algorithm
Grid graph
$(q, q-4)$-graph

ABSTRACT

Given a graph with an arbitrary vertex coloring, the Convex Recoloring Problem (CR) consists in recoloring the minimum number of vertices so that each color induces a connected subgraph. We focus on the complexity and inapproximability of this problem on k-colored graphs, for fixed $k \geqslant 2$. We prove a strong complexity result showing that, for each $k \geqslant 2$, CR is already NP-hard on k-colored grids, and therefore also on planar graphs with maximum degree 4. For each $k \geqslant 2$, we prove that, for a positive constant c, there is no $c \ln n$-approximation algorithm for k-colored n-vertex (bipartite) graphs, unless $\mathrm{P} = \mathrm{NP}$. We also prove that CR parameterized by the number of color changes is $W[2]$-hard. For 2-colored $(q, q-4)$-graphs, a class that includes cographs and P_4-sparse graphs, we present linear-time algorithms for fixed q. The same complexity and inapproximability results are obtained for two relaxations of the problem, where only one fixed color or any color is required to induce a connected subgraph, respectively.

1. Introduction

Consider a game in which all players receive an $n \times n$ chessboard (grid) in which all the n^2 squares are occupied by either a *black* or a *white* pebble in an arbitrary manner. Every pebble has one face colored black and the other one colored white (as in Reversi or Othello). The initial configuration of the pebbles is the same for all players. The goal of each player is to reverse (turn to the opposite face) a least number of pebbles in order to reach a configuration where each color occupies a unique 'connected' region of his board. A winner is a player who makes the least number of reversals.

The game we have just described may be seen as the convex recoloring problem on a 2-colored grid. It is an easy-to-state combinatorial problem, for which it is natural to ask whether an optimal solution can be easily found or not. Surprisingly, it is a hard problem, as we prove in this paper. In fact, we shall prove that the problem is hard on k-colored grids, for fixed $k \geqslant 2$.

A *k-coloring* of graph G is a function $C : V(G) \to \{1, 2, \dots, k\}$ that assigns to each vertex v of G a natural number $C(v)$, called the *color* of v. Note that the coloring considered here differs from the classical (proper) vertex coloring of graphs,

☆ Partially supported by FUNCAP (PRN-0040-00040.01.00/10, DCR-0024-00044.01.00/10), FACEPE/CNPq (DCR-0007-1.03/13), FAPESP (Proc. 2013/03447-6), CNPq (Proc. 303987/2010-3, 307627/2010-1, 480608/2011-3 and 477203/2012-4), CAPES (STIC-AmSud 040/2013) and USP (MaCLinC/NUMEC Project).
∗ Corresponding author.
E-mail addresses: mcampelo@lia.ufc.br (M. Campêlo), cmngh@cin.ufpe.br (C. Huiban), rudini@ufc.br (R.M. Sampaio), yw@ime.usp.br (Y. Wakabayashi).

http://dx.doi.org/10.1016/j.tcs.2014.03.017
0304-3975/

where $C(u) \neq C(v)$ for all $uv \in E(G)$. A graph is *k-colored* if it is assigned a k-coloring that uses k colors. A coloring C is *convex* if each color class (i.e. set of vertices with the same color) induces a connected subgraph. The *convex recoloring problem* (CR) consists in recoloring (changing the color of) the minimum number of vertices in G so that the resulting coloring is convex. (We remark that, in a recoloring process, some of the initial colors may disappear.)

This problem was introduced by Moran and Snir [17,18] in 2005, motivated by studies on phylogenetic trees. They showed that CR (i.e. convex recoloring without fixing k) is NP-hard even on paths, and presented the first approximation results. Since then CR and some of its variants have been intensively investigated. More recently, some applications on routing problems, transportation networks and protein–protein interaction networks have also been mentioned in the literature [9,13]. In general, the idea behind these applications is that the given colored graph represents some situation, and a convex coloring of such a graph represents a desirable (perfect) configuration, which one wants to achieve after a least number of color changes (such a number is called the *recoloring distance*).

In the case of phylogenetic trees (i.e. a tree which represents the course of evolution for a given set of species), the leaves are labelled with the given species and the internal vertices correspond to hypothesized, extinct species. The colors of the vertices correspond to character states (a character is a biological attribute shared by the species). The natural assumption that the reconstructed phylogeny has the property that each of the characters should have evolved without reverse or convergent transitions is captured, in graph theoretical terms, by the requirement that each of the colors should induce a connected subtree. Thus the recoloring distance captures the distance from a perfect phylogeny. For a more detailed biological explanation, the reader is referred to [18].

Approximation algorithms for CR were first designed for trees [17,18,4], and more recently generalized to graphs with bounded treewidth [13], for which a $(2+\varepsilon)$-ratio has been shown. When each color appears at most twice, the problem remains NP-hard on paths [14], but admits a $\frac{3}{2}$-approximation [16]. For an n-vertex arbitrary graph, however, CR cannot be approximated within a factor of $(1-o(1))\ln\ln n$ in polynomial time, unless $\mathrm{NP} \subset \mathrm{DTIME}(n^{O(\log\log n)})$ [13]. On the other hand, some of the approximation algorithms for trees apply to weighted versions of CR and other variants as well (on partial coloring). These more general variants can all be treated (as the cardinality case) using a polyhedral approach [8].

The *parameterized* complexity has been studied for CR on trees with respect to different parameters: the number of colors k, the number of bad colors β (initial colors that do not induce subtrees), the maximum number of recolorings r (to achieve a convex coloring), the maximum degree Δ of the input graph, etc. The first results showing the fixed-parameter tractability of the problem appear in [18], an $O((r/\log r)^r r.poly(n))$ and an $O(\Delta^k \cdot kn)$ algorithms. Improving these results we have, respectively, the $O(256^r poly(n))$ algorithm of [23] and the $O(3^\beta \cdot \beta n)$ algorithm of [21] (as $\beta \leqslant k$). We also mention the polynomial-time kernelization for the problem obtained by [5], defining a kernel with $O(r^2)$ vertices.

Here, we focus on a restricted version of the convex recoloring problem and on two closely related problems, as defined below.

- **Convex Recoloring Problem with k colors (CRk)**: Given a k-colored graph G, determine the minimum number of vertex recolorings that are needed to obtain a convex coloring of G.
- **OneClass-Convex Recoloring Problem with k colors (CRkONE)**: Given a k-colored graph G and a specific color r, determine the minimum number of vertex recolorings that are needed to guarantee that the color class of r induces a connected subgraph of G.
- **AnyClass-Convex Recoloring Problem with k colors (CRkANY)**: Given a k-colored graph G, determine the minimum number of vertex recolorings that are needed to guarantee that at least one color class induces a connected subgraph of G.

We note that, in the three problems above, the number of colors k in the initial coloring is fixed. Only CR2 has been investigated in the literature, and for this problem, it has been proved recently that it is NP-hard on arbitrary graphs [19]. The other two problems have not been treated in the literature. It is very intriguing that only two colors make the convex recoloring already hard. This fact motivated us to investigate nontrivial classes of graphs for which CR2 can be solved in polynomial time. While it is easy to see that CR2 on trees (resp. ladders) can be solved in linear (resp. polynomial) time, for the well-structured grid graphs, our intuition that the problem would be easy turned out to be wrong. For trees (resp. ladders), it suffices to note that there is a linear (resp. polynomial) number of cuts that separate the input graph into two connected subgraphs (the recoloring cost of which, to make them monochromatic, can be easily tested).

The results we present in this paper were motivated by the investigations to understand the polynomial solvability or approximability threshold of the CR2 problem. We show that for any fixed $k \geqslant 2$, CRk is NP-hard on grids, which clearly implies NP-hardness on (bipartite) planar graphs with maximum degree 4.

This hardness result motivated us to define the two relaxed problems CRkONE and CRkANY. These problems, if polynomially solvable, could possibly help in the design of an approximation algorithm for CRk, as they would provide a lower bound for the optimal value of CRk. Unfortunately, it turned out that these relaxations are also NP-hard. This is the subject of Section 2.

In Section 3, we focus on approximability thresholds. Clearly, all approximation algorithms for CR are approximation algorithms for CRk. On the other hand, we show that, for each $k \geqslant 2$, if $\mathrm{P} \neq \mathrm{NP}$, the three studied problems are not approximable within a factor of $c \ln n$ on n-vertex bipartite graphs (for a constant $c > 0$) with k colors. To our knowledge, this is the strongest inapproximability result for CRk (and therefore for CR), under the hypothesis that $\mathrm{P} \neq \mathrm{NP}$. Kammer and