Journal of Discrete Algorithms 23 (2013) 21–34

Contents lists available at SciVerse ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

Identifying path covers in graphs ☆

Florent Foucaud [a,b,*], Matjaž Kovše [c]

[a] Univ. Bordeaux, LaBRI, UMR5800, F-33400 Talence, France
[b] CNRS, LaBRI, UMR5800, F-33400 Talence, France
[c] Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, Univ. of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany

ARTICLE INFO

Article history:
Available online 26 July 2013

Keywords:
Test cover
Identification
Paths
Approximation

ABSTRACT

This paper introduces the problem of identifying vertices of a graph using paths. An *identifying path cover* of a graph G is a set $\mathcal{P}$ of paths such that each vertex belongs to a path of $\mathcal{P}$, and for each pair u, v of vertices, there is a path of $\mathcal{P}$ which includes exactly one of u, v. This new notion is related to a large number of other identification problems in graphs and hypergraphs. We study the identifying path cover problem under both combinatorial and algorithmic points of view. In particular, we derive the optimal size of an identifying path cover for paths, cycles and hypercubes, and give upper bounds for trees. We give lower and upper bounds on the minimum size of an identifying path cover for general graphs, and discuss their tightness. In particular, we show that any connected graph G has an identifying path cover of size at most $\lceil\frac{2|V(G)|}{3}\rceil + 5$. We then study the computational complexity of the associated optimization problem, in particular we show that when the length of the paths is asked to be of bounded value, the problem is APX-complete.

1. Introduction

This paper aims to study the new optimization problem of identifying the vertices of a graph by means of paths, which we call the *identifying path cover problem*. We first relate this problem to a large number of other problems and review a part of the associated literature, before giving its definition.

1.1. Test covers and the identification problem

Identification problems have been addressed many times in the last decades under different denominations and in different contexts. We present two general problems from the literature which have almost the same definition, and which we herein call the minimum test cover problem and the minimum identification problem. Instances of these problems are set systems, i.e. pairs consisting of a set $\mathcal{I}$ of elements ("individuals") and a set $\mathcal{A}$ of subsets of $\mathcal{I}$ ("attributes").

Among these two problems, the *minimum test cover problem*, in short MIN-TC, seems to have been studied first and is probably better known. Given a set system of individuals and attributes, the MIN-TC problem asks for a minimum subset $\mathcal{C}$ of $\mathcal{A}$ such that for each pair I, I' of $\mathcal{I}$, there is an element C of $\mathcal{C}$ such that exactly one of I, I' is *covered* by C, that is,

☆ A shorter version of this paper appeared under the name *On graph identification problems and the special case of identifying vertices using paths* in the Proceedings of the International Workshop on Combinatorial Algorithms, IWOCA 2012 (Foucaud and Kovše, 2012) [11]. The present paper contains additional results, mainly Theorems 8, 10 and 20, and a corrected version of Theorems 17 and 18, which contained some mistakes in the short version.

* Corresponding author at: Univ. Bordeaux, LaBRI, UMR5800, F-33400 Talence, France.
E-mail address: florent.foucaud@gmail.com (F. Foucaud).

1570-8667/$ – see front matter
http://dx.doi.org/10.1016/j.jda.2013.07.006

belongs to C (we say that C *separates* I from I'). The MIN-TC problem appears in a large number of papers under different denominations (*minimum test cover problem* [8], *minimum test collection problem* [14], *minimum test set problem* [21]). In fact, a well-celebrated theorem of J.A. Bondy on *induced subsets* [3] can be seen as the first study of this problem.

In this paper and as in a large portion of the literature dealing with special cases of this kind of problems, we are interested in a slight modification of MIN-TC, where not only each pair of individuals has to be separated, but also, each individual has to be covered. We call this problem the *minimum identification problem*, MIN-ID for short (note that it has been studied under the denomination of *discriminating code problem* in [4], but we use our terminology in order to fit to special cases described later). MIN-TC and MIN-ID are very close to each other, since for any solution to one of them, there is a solution to the other one whose size differs by at most 1: any solution to MIN-ID is also one for MIN-TC, and, given a solution $\mathcal{C}$ to MIN-TC which is not a valid solution to MIN-ID, at most one individual I may not be covered by $\mathcal{C}$. It is then sufficient to add an arbitrary attribute A covering I to $\mathcal{C}$ to get a valid solution to MIN-ID.

Both MIN-TC and MIN-ID can be seen as special cases of the well-known *minimum set cover problem* [14,16], MIN-SC for short, where, given a base set $\mathcal{X}$ and a set $\mathcal{S}$ of subsets of $\mathcal{X}$, it is asked to find a minimum subset $\mathcal{C}$ of $\mathcal{S}$ covering all elements of $\mathcal{X}$ [8]. MIN-TC and MIN-ID enjoy the same computational complexity. It is known that both problems are $O(\ln(|\mathcal{I}|))$-approximable (where $\mathcal{I}$ denotes the set of individuals of the input) using a reduction to MIN-SC [21]. On the other hand, both problems are not only NP-hard [4,14] but have also been shown to be NP-hard to approximate within a factor of $o(\ln(|\mathcal{I}|))$ by reduction from MIN-SC [2,8].

A natural restriction of MIN-ID is, given some integer k, the one where the sets of $\mathcal{A}$ all have exactly k elements. We will call this problem MIN-ID-k.

1.2. Related problems

In this paper, we study a special case of MIN-ID. Just as some particular cases of MIN-SC arising from specific structures have gained a lot of interest (consider for example all variants of the minimum dominating set problem, or the minimum vertex cover problem), it is of interest to investigate special cases of the MIN-ID problem having a particular structure. In this line of research, many specific cases arising from graph theory are of particular interest since graphs model networks of all kinds and are found in real world applications. For example, in the *identifying code problem* [9,13,18], one wants to identify each vertex v using vertices at distance at most 1 from v. This problem can be seen as MIN-ID where $\mathcal{I} = V(G)$ and $\mathcal{A}$ is the family of the balls around each vertex. This problem has been generalized to digraphs [6,12], and to the case where also *sets* of at most ℓ vertices are to be separated and where vertices can identify at some prescribed distance $r \geqslant 1$ [13]. One may also ask to identify the edges of G using edges, i.e. $\mathcal{I} = E(G)$ and $\mathcal{A}$ is the set of all edge-balls around each edge of G [10]. Rather than considering full balls, also partial balls may be considered, as in the case of *watching systems* [1], where $\mathcal{I} = V(G)$, and $\mathcal{A}$ is the family of all subgraphs of stars in G. Finally, the case where $\mathcal{I} = V(G)$ and $\mathcal{A}$ is the set of all cycles in G has been considered in [15,23].

1.3. The identifying path cover problem

Given a graph G, a *path* is an ordered set of distinct vertices such that any two consecutive vertices in the ordering are adjacent. A path containing k vertices is a k-*path*; a 1-path is just a single vertex. We will consider the path graph P_k, consisting of a single path on k vertices; similarly, C_k denotes the cycle on k vertices.

In this paper, we study MIN-ID when $\mathcal{I} = V(G)$ and $\mathcal{A}$ is the set of all paths of G. We call it *minimum identifying path cover problem*, MIN-IDPC for short and it studies the following notion:

Definition 1. Given a graph G, a set $\mathcal{P}$ of paths of G is an *identifying path cover* if each vertex of G belongs to a path of $\mathcal{P}$ (it is *covered*) and if for each pair u, v of vertices, there is a path of $\mathcal{P}$ which contains exactly one of u, v (u, v are *separated*).

We point out that the covering condition is not implied by the separation condition, since even when all pairs are separated, one vertex of the graph may remain uncovered. We denote by $p^{\mathrm{ID}}(G)$ the minimum number of paths required in any identifying path cover of G. Then, MIN-IDPC is the problem, given a graph G, of determining the value of $p^{\mathrm{ID}}(G)$. An example of an identifying path cover $\mathcal{P}$ of the cube H_3 is given in Fig. 1, where the four thick paths belong to $\mathcal{P}$ (the full, the densely dotted, the loosely dotted and the dashed–dotted path). Note that an identifying path cover of G always exists: consider the set of all 1-paths of G, that is, $\mathcal{P} = V(G)$.

Given a fixed integer $k \geqslant 1$, we will also discuss the natural variant MIN-IDPC-k of MIN-IDPC, where one wants to find a minimum *identifying k-path cover of G*, that is, a set of paths of exactly k vertices forming an identifying path cover of G. We denote by $p_k^{\mathrm{ID}}(G)$ the size of a minimum identifying k-path cover of G. Unlike for the general MIN-IDPC problem, not all graphs admit an identifying k-path cover. We call a graph admitting an identifying k-path cover, k-*path identifiable*. This is the case if, first of all, each vertex of G lies on a k-path, and if for each pair u, v of vertices, there is a k-path covering exactly one of u, v. For example, the path graphs P_{k-1} and P_{2k-2} are not k-path identifiable. Observe that these two conditions are also sufficient: if both are fulfilled, taking all k-paths of G gives a valid identifying k-path cover of G. For fixed k, being k-path identifiable is polynomial-time checkable since there are at most $\binom{n}{k} = O(n^k)$ k-paths in G.