Discrete Applied Mathematics 182 (2015) 25–33

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On the complexity of the identifiable subgraph problem

Marcin Kamiński [a], Martin Milanič [b,c,d,∗]

[a] Department of Mathematics, Computer Science, and Mechanics, University of Warsaw, Warsaw, Poland
[b] University of Primorska, UP IAM, Muzejski trg 2, SI6000 Koper, Slovenia
[c] University of Primorska, UP FAMNIT, Glagoljaška 8, SI6000 Koper, Slovenia
[d] IMFM, Jadranska 19, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form 21 October 2014
Accepted 27 October 2014
Available online 20 November 2014

Keywords:
Bipartite graph
Matching
Identifiable graph
Topological minor
Strongly cyclic graph
Polynomial algorithm

ABSTRACT

A bipartite graph $G = (L, R; E)$ with at least one edge is said to be *identifiable* if for every vertex $v \in L$, the subgraph induced by its non-neighbors has a matching of cardinality $|L|-1$. This definition arises in the context of low-rank matrix factorization and is motivated by signal processing applications. An *ℓ-subgraph* of a bipartite graph $G = (L, R; E)$ is an induced subgraph of G obtained by deleting from it some vertices in L together with all their neighbors. The Identifiable Subgraph problem is the problem of determining whether a given bipartite graph $G = (L, R; E)$ contains an identifiable ℓ-subgraph. While the problem of finding a *smallest* set $J \subseteq L$ that induces an identifiable ℓ-subgraph of G is NP-hard and also APX-hard, the complexity of the identifiable subgraph problem is still open.

In this paper, we introduce and study the k-bounded Identifiable Subgraph problem. This is the variant of the Identifiable Subgraph problem in which the input bipartite graphs $G = (L, R; E)$ are restricted to have the maximum degree of vertices in R bounded by k. We show that for $k \geq 3$, the k-bounded Identifiable Subgraph problem is as hard as the general case, while it becomes solvable in linear time for $k \leq 2$. Our proof is based on the notion of *strongly cyclic graphs*, that is, multigraphs with at least one edge such that for every vertex v, no connected component of the graph obtained by deleting v is a tree. We show that a bipartite graph $G = (L, R; E)$ with maximum degree of vertices in R bounded by 2 is a no instance to the Identifiable Subgraph problem if and only if a multigraph naturally associated to it does not contain any strongly cyclic subgraph, and characterize such graphs in terms of finitely many minimal forbidden topological minors.

1. Introduction

The identifiability property of bipartite graphs is based on the classical concept of a *matching* in a graph (a subset of pairwise disjoint edges). A bipartite graph $G = (L, R; E)$ with at least one edge is said to be *identifiable* if for every vertex in L, the subgraph induced by its non-neighborhood has a matching of cardinality $|L| - 1$. This definition arises in the context of low-rank matrix factorization and has applications in data mining [17], signal processing [13], and computational biology [4,14,18]. For further details on applications of notions and problems discussed in this paper, we refer to [9,10].

On the one hand, the recognition problem for identifiable bipartite graphs is clearly polynomial, using bipartite matching algorithms. On the other hand, several natural algorithmic problems concerning identifiable graphs turn out to be NP-complete. For example:

∗ Corresponding author at: University of Primorska, UP IAM, Muzejski trg 2, SI6000 Koper, Slovenia.
E-mail addresses: mjk@mimuw.edu.pl (M. Kamiński), martin.milanic@upr.si (M. Milanič).

http://dx.doi.org/10.1016/j.dam.2014.10.021
0166-218X/

(1) Given an identifiable bipartite graph G, how strongly does G possess the identifiability property with respect to edge modifications (that is, edge additions and/or deletions)? This question results in the notion of the *resilience* of G with respect to identifiability, which measures how much one should change G by means of edge modifications to destroy this property. While computing the resilience is polynomial for edge additions or edge modifications, it is NP-complete for edge deletions [10].

(2) Given an identifiable bipartite graph $G = (L, R; E)$, the problem of selecting a minimum-size set $R' \subseteq R$ of vertices in R such that the subgraph of G induced by $L \cup R'$ is identifiable, is approximable in polynomial time to within a factor of $\ln |L| + 1$, using a greedy algorithm based on the notion of submodular set functions [9]. Its decision version is NP-complete [9].

(3) The following definition is a slight modification of [10, Definition 2].

Definition 1. *Let $G = (L, R; E)$ be a bipartite graph. For a subset $J \subseteq L$, the ℓ-subgraph of G induced by J is the subgraph $G(J) = G[J, R \setminus N(L \setminus J)]$, where $N(L \setminus J)$ denotes the set of all vertices in R with a neighbor in $L \setminus J$. We say that a graph G' is an ℓ-subgraph of G if there exists a subset $J \subseteq L$ such that $G' = G(J)$.*

In other words, an ℓ-subgraph of G is an induced subgraph of G obtained by deleting from G some (possibly none) vertices in L together with all their neighbors. In [10], the following problem was introduced.

Min-Identifiable Subgraph (min-ids)	
Instance:	A bipartite graph $G = (L, R; E)$ and an integeqr k.
Question:	Does G have an identifiable ℓ-subgraph induced by a set J with $|J| \leq k$?

In [10], this problem was shown to be APX-hard in general but polynomially solvable for trees and for bipartite graphs in which the maximum degree of vertices in L is at most 2.

In the same paper, the following related problems were also introduced:

Max-Identifiable Subgraph (max-ids)	
Instance:	A bipartite graph $G = (L, R; E)$ and an integer k.
Question:	Does G have an identifiable ℓ-subgraph induced by a set J with $|J| \geq k$?

Identifiable Subgraph (ids)	
Instance:	A bipartite graph $G = (L, R; E)$.
Question:	Does G have an identifiable ℓ-subgraph?

While the min-ids problem is APX-hard, the complexity of the max-ids and ids problems in general remains open. Both problems are solvable in polynomial time for trees, as well as for bipartite graphs $G = (L, R; E)$ such that the maximum degree of vertices in L is at most 2 [10]. Clearly, if the ids problem is NP-complete then so is the max-ids problem: by varying the value of k, one could solve the ids problem using an algorithm for the max-ids problem. Similarly, if the ids problem is NP-complete, this would provide an alternative proof for the NP-completeness of min-ids.

In this paper, we study the ids problem from the point of view of parameterizing it according to the maximum degree $\Delta(R)$ of vertices in R. This is formalized by the following problem definition (where k is a positive integer):

k-bounded Identifiable Subgraph (k-ids)	
Instance:	A bipartite graph $G = (L, R; E)$ with $\Delta(R) \leq k$.
Question:	Does G have an identifiable ℓ-subgraph?

While the k-bounded Identifiable Subgraph problem for $k \geq 3$ is as hard as the general case (see Section 3), we show that the ids problem becomes solvable in linear time if the maximum degree of vertices in R is at most 2. This result is stated formally in the following theorem.

Theorem 1. *The* 2-bounded Identifiable Subgraph *problem is solvable in linear time.*

A proof of Theorem 1 will be given in Section 4. Our approach can be described roughly as follows. We first preprocess the input graph G so that we have $d(x) = 2$ for all $x \in R$. Under this assumption, graph G can be obtained from some multigraph H by subdividing every edge exactly once. We prove that G has an identifiable ℓ-subgraph if and only if H has an induced strongly cyclic subgraph, where a graph with at least one edge is said to be *strongly cyclic* if no deletion of a vertex results in a graph with an acyclic component. We then characterize graphs containing an induced strongly cyclic subgraph as precisely the graphs that contain as a topological minor one of five particular graphs on at most 6 vertices (see Fig. 2 on page 6). Finally, we describe how known results and techniques for graphs of bounded treewidth can be used to detect the presence of one of these five graphs as a topological minor in linear time.

The paper is structured as follows. In Section 2, we give the necessary definitions and recall the results from the literature we will need in some of our proofs. In Section 3, we discuss the case with $\Delta(R) = k$ for $k \geq 3$. In Section 4 we prove our main result (Theorem 1) and argue in Section 4.1 that an identifiable ℓ-subgraph in a graph with $\Delta(R) \leq 2$ can also be found in polynomial time if one exists. Section 5 concludes the paper with some open questions.