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

(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.

2. Preliminaries

All graphs considered in this paper are finite, undirected and loopless but may contain parallel edges. For a graph G, we denote by $V(G)$ the vertex set of G and by $E(G)$ its edge (multi)set. A *bipartite graph* is a graph $G = (V, E)$ without parallel edges such that there exists a partition of V into two sets L and R such that $L \cap R = \emptyset$ and $E \subseteq \{\{\ell, r\} ; \ell \in L \text{ and } r \in R\}$. In this paper, we will regard bipartite graphs as already *bipartitioned*, that is, given together with a fixed bipartition (L, R) of their vertex set, and hence use the notation $G = (L, R; E)$. For a graph $G = (V, E)$ and a subset of vertices $X \subseteq V$, $N_G(X)$ denotes the neighborhood of X, i.e., the set of all vertices in $V \setminus X$ that have a neighbor in X. For a vertex $x \in V$, we write $N_G(x)$ for $N_G(\{x\})$, and denote the *degree* of x with $d_G(x) = |N_G(x)|$. In $N_G(X)$, $N_G(x)$, $d_G(x)$, we shall omit the subscript G if the graph is clear from the context. For a bipartite graph $G = (L, R; E)$ and vertex sets $X \subseteq L$, $Y \subseteq R$, we denote by $G[X, Y]$ the subgraph of G induced by $X \cup Y$.

A *cut vertex* of a connected graph G is a vertex whose removal will disconnect the graph. A *block* of G is a maximal connected subgraph of G without cut vertices. The *block tree* of G is a tree $B(G)$ with bipartition (B, S) of its vertex set, where B is the set of blocks of G and S the set of cut vertices of G, a block B and a cut vertex v being adjacent in $B(G)$ if and only if B contains v. An *end block* of G is a block of G that corresponds to a leaf of the block tree of G. A graph G is said to be 2*-connected* if it has at least 3 vertices and for every $x \in V(G)$, the graph $G - x$ is connected. A *cycle* in a graph G is a connected subgraph of G in which every vertex is incident with exactly two edges. A subgraph of G obtained by deleting one edge from a cycle in G is a *path* in G. Also, any 1-vertex subgraph is a path. Given a path P in a graph G, we denote by $\mathrm{int}(P)$, the subpath of P containing only the internal vertices of P. For a graph G, a subgraph H of G, and a vertex $v \in V(G) \setminus V(H)$, a (v, H)*-path* is a path P in G such that one endpoint of P is v and the other endpoint of P belongs to $V(H)$. A (v, X)*-path* where $v \notin X \subseteq V(G)$ is defined similarly, For $X, Y \subseteq V(G)$, and (X, Y)*-path* is a path P in G such that one endpoint of P belongs to X and the other one to Y.

Our proof of Theorem 6 in Section 4 will rely on the following classical results on the structure of 2-connected graphs (see, e.g., Exercise 5.1.4 and Proposition 9.5 in [3]).

Proposition 1. *Let $G = (V, E)$ be a 2-connected graph. Then:*

(i) *If X and Y are two sets of vertices of G, each of cardinality at least two, then there exist in G two disjoint (X, Y)-paths.*

(ii) *Let x be a vertex of G, and let $Y \subseteq V \setminus \{x\}$ be a set with $|Y| \geq 2$. Then there exist two internally disjoint (x, Y)-paths whose endpoints in Y are distinct.*

Subdividing an edge of a graph G means replacing the edge with a path of length two. Given a graph G, the *subdivision graph* of G is the graph $S(G)$ obtained from G by subdividing each edge of it. More generally, a graph H is a *subdivision* of a graph G if H can be obtained from G by replacing its edges by paths of length at least 1. A graph H is a *topological minor* of a graph G if a subdivision of H is isomorphic to a subgraph of G.

The above definitions are typically used for simple graphs, but can be easily generalized to graphs with parallel edges. For definitions not given in the paper, we refer to the books by Bondy and Murty [3] and Diestel [6] (graph theory), to the monographs by Ausiello et al. [1] and Garey and Johnson [11] (computational complexity), and to the monograph by Flum and Grohe [8] (treewidth, MSOL).

3. A reduction of the IDS problem to the k-IDS problem for $k \geq 3$

In this section, we show that the Identifiable Subgraph problem is polynomially equivalent to the 3-bounded Identifiable Subgraph problem. The reduction can be achieved by a repeated application of an operation that preserves the property of containing an identifiable ℓ-subgraph and can be used to decrease the degree of vertices in R down to 3 with only a polynomial blowup of the input graph.

Theorem 2. *For every $k \geq 3$, the* Identifiable Subgraph *problem is polynomially equivalent to the k-bounded* Identifiable Subgraph *problem.*

Proof. Let $k \geq 3$. Clearly, if the IDS problem is polynomially solvable then so is the k-IDS problem. In the rest of the proof, we will describe a polynomial reduction of the IDS problem to the k-IDS problem. That is, given an instance $G = (L, R; E)$ to the IDS problem, we will show how to compute in polynomial time an instance $G' = (L', R'; E')$ to the k-IDS problem such that G is a yes instance to the IDS problem if and only if G' is a yes instance to the k-IDS problem.

The idea is to repeatedly apply a graph transformation that preserves the property of containing an identifiable ℓ-subgraph but decreases the degrees of vertices in R. This transformation takes as input a bipartite graph $G = (L, R; E)$ and a vertex $v \in R$ of degree greater than k, and outputs a graph $\varphi(G, v) = G'$, defined as follows. Fix a linear order of neighbors of v in G, say $N_G(v) = \{v_1, \dots, v_r\}$. Then, $G' = (L', R'; E')$ where

- $L' = L \cup \{v_0\}$ where v_0 is a new vertex,
- $R' = (R \setminus \{v\}) \cup \{v', v''\}$ where v' and v'' are two new vertices,
- $E' = (E \setminus \{v_i v \mid 1 \leq i \leq r\}) \cup \{v_1 v', v_2 v', v_0 v', v_0 v''\} \cup \{v_i v'' \mid 3 \leq i \leq r\}$.

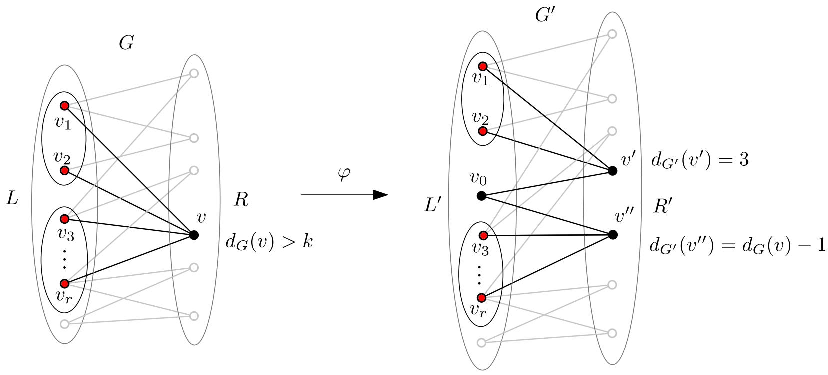

Fig. 1. Transformation φ.

See Fig. 1.

Clearly, $|V(G')| = |V(G)| + 2$ and $|E(G')| = |E(G)| + 2$, and the reduction can be performed in polynomial time. For every $w \in R'$, we have

$$d_{G'}(w) = \begin{cases} d_G(w), & \text{if } w \notin \{v', v''\}; \\ d_G(v) - 1, & \text{if } w = v''; \\ 3, & \text{if } w = v'. \end{cases}$$

In particular, if we define $f(G, R) = \sum_{v \in R} \max\{d_G(v) - k, 0\}$, then $f(G', R') = f(G, R) - 1$.

Given an input graph $G = (L, R, E)$ to the IDS problem, we apply transformation φ iteratively as long as the value of f is positive. Since $f(G, R) = O(|R||L|)$ and the value of $f(G, R)$ reduces by 1 with each transformation, transformation φ will be applied at most $O(|L||R|)$ times, resulting in a graph G' with $|V(G')| = O(|L||R|)$ and $|E(G')| = O(|L||R|)$ such that $\Delta(G') \le k$. Hence, the proof of the theorem will follow from the claim below.

Claim. *For every bipartite graph $G = (L, R; E)$ and every vertex $v \in R$ with $d_G(v) > k$, graph G contains an identifiable ℓ-subgraph if and only if $\varphi(G, v)$ does.*

Proof of claim. Let $\varphi(G, v) = G'$ with $G' = (L', R'; E')$. Suppose first that the ℓ-subgraph of G induced by some $J \subseteq L$ is identifiable. Using the notation as in the definition of the mapping φ, let us define

$$J' = \begin{cases} J \cup \{v_0\}, & \text{if } v \in V(G(J)); \\ J, & \text{otherwise.} \end{cases}$$

Then $J' \subseteq L'$. We claim that the ℓ-subgraph of G' induced by J' is identifiable. We consider two cases, depending on whether $v \in V(G(J))$ or not.

Case 1: $v \in V(G(J))$. In this case, the set $L \setminus J$ contains no neighbor of v, that is, $N_G(v) \subseteq J$. Consequently, the graph $G'(J')$ is equal to $\varphi(G(J), v)$, and we will verify the identifiability of $G'(J')$ by definition. We need to show that for every $w \in J'$, the subgraph of $G'(J') = \varphi(G(J), v)$ induced by the non-neighbors of w has a matching of cardinality $|J'| - 1$.

First, let $w \in J$. Let M be a matching in $G(J)$ of cardinality $|J| - 1$ avoiding the neighborhood of w. We consider two cases based on whether v is covered by M or not. Suppose first that v is covered by M. By the definition of M, we have $w \notin N_G(v)$. Consequently, $w \notin N_{G'}(v') \cup N_{G'}(v'')$. Let $m(v)$ be the vertex matched to v by M. Let $\{v', v''\} = \{z', z''\}$ so that $z' = v'$ if $m(v) \in \{v_1, v_2\}$, and $z' = v''$ otherwise. Then, $(M \setminus \{m(v)v\}) \cup \{m(v)z', v_0z''\}$ is a matching of size $|J'| - 1$ in the graph $G'(J')$ avoiding the neighborhood of w. Suppose now that v is not covered by M. Since w is in J, it is non-adjacent in $G'(J')$ to at least one vertex, say z, in the set $\{v', v''\}$. Furthermore, M is a matching in G', and thus $M \cup \{v_0z\}$ is a matching of size $|J'| - 1$ in the graph $G'(J')$ avoiding the neighborhood of w.

Second, let $w \notin J$. Then $w = v_0$. For $i \in \{1, 2\}$, let M_i be a matching in $G(J)$ of cardinality $|J| - 1$ avoiding the neighborhood of v_i. Since v_1 and v_2 are adjacent to v in $G(J)$, neither of M_1, M_2 covers v. In particular, this implies that the vertex v_1' matched by M_2 to v_1 satisfies $v_1' \neq v$. Consequently, M_1 is a matching in $G'(J')$ covering neither v_1 nor v_1', and thus $M_1 \cup \{v_1v_1'\}$ is a matching of size $|J'| - 1$ in the graph $G'(J')$ avoiding the neighborhood of v_0.

Case 2: $v \notin V(G(J))$. In this case, we have $J' = J \subseteq L' \setminus \{v_0\}$. Since v_0 belongs to $L' \setminus J'$, vertices v' and v'' (which are adjacent to v_0 in G') do not belong to $G'(J')$, and hence it can be seen that $G'(J')$ equals $G(J)$; the identifiability of $G'(J')$ thus follows immediately from that of $G(J)$.

For the converse direction, suppose that the ℓ-subgraph of G' induced by some $J' \subseteq L'$ is identifiable. Let $J = J' \setminus \{v_0\}$. Then $J \subseteq L$. We claim that the ℓ-subgraph of G induced by J is identifiable.

If $v_0 \notin J'$, then neither v' nor v'' belongs to $V(G'(J'))$, which implies that $G(J) = G'(J')$, and identifiability of $G(J)$ trivially follows. So suppose that $v_0 \in J'$. We claim that in this case we also have $v', v'' \in V(G'(J'))$. Indeed, if this were not the case, then v_0 would be either isolated in $G'(J')$ (if $v', v'' \notin V(G'(J'))$) or of degree one in $G'(J')$ (if $|V(G'(J')) \cap \{v', v''\}| = 1$). In the former case, $G'(J')$ would clearly not be identifiable. In the latter case, suppose first that $V(G'(J')) \cap \{v', v''\} = \{v'\}$. Then $\{v_1, v_2\} \subseteq J$ and hence v_0 would be isolated in the subgraph of $G'(J')$ induced by the non-neighbors of v_1, contrary to the identifiability of $G'(J')$. The case when $V(G'(J')) \cap \{v', v''\} = \{v''\}$ can be handled similarly.

The fact that $v', v'' \in V(G'(J'))$ implies that $N_G(v) \subseteq J'$ and consequently $N_G(v) \subseteq J$. Let us verify the identifiability of $G(J)$ by definition. Let $w \in J$. Let M' be a matching of size $|J'|-1 = |J|$ in the subgraph of $G'(J')$ induced by the non-neighbors of w. Then, a matching of size $|J| - 1$ in the subgraph of $G(J)$ induced by the non-neighbors of w can be obtained from M' by deleting from it the edge covering v_0 and replacing, if necessary, any edge of the form v_iv' for some $i \in \{1, 2\}$, or v_iv'' for some $i \in \{3, \ldots, r\}$ with the edge v_iv. This replacement is only done if such an edge exists in M'; notice that since v_0 is matched in M' to either v' or v'', there can be at most one such edge so the replacement is well defined. Therefore, the subgraph of G induced by J is identifiable also in this case.

This completes the proof of the claim and with it the proof of Theorem 2. □

4. A linear-time algorithm for the 2-bounded Identifiable Subgraph problem

In this section we prove our main result:

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

Let $G = (L, R; E)$ be a bipartite graph such that $\Delta(R) \le 2$. We will describe a linear-time algorithm that will determine whether G contains an identifiable ℓ-subgraph.

We first preprocess the input graph G in time $O(|R|)$ so that after the preprocessing step, we have either found an identifiable ℓ-subgraph, or we have $d(x) = 2$ for all $x \in R$. This can be achieved by an iterative application of the following two obvious preprocessing rules:

- If G has an isolated vertex $v \in R$, replace G with $G - v$.
- If there is a vertex $y \in R$ such that $d(y) = 1$, then the singleton $J = N(y)$ forms a subset of L such that the ℓ-subgraph induced by J is identifiable.

From now on, we work under the assumption that all vertices in R are of degree 2. Under this assumption, it is clear that $G = S(H)$ is the subdivision graph of some other graph H. (The graph H, which can have parallel edges, is given by $V(H) = L$ and $E(H) = \{N_G(y) : y \in R\}$.) In the next lemma, we give a necessary and sufficient condition for the existence of an identifiable ℓ-subgraph in G in terms of the graph H. The condition will rely on the notion of strongly cyclic graphs, which we now introduce.

Definition 2. *We say that a graph K with at least one edge is strongly cyclic if for all $v \in V(K)$, every component of the graph $K - v$ contains a cycle (equivalently: no component of $K - v$ is a tree).*

We start with two auxiliary observations.

Proposition 2. *Let $K = (A, B; E)$ be a bipartite graph such that $d_K(y) \ge 2$ for all $y \in B$. If K contains a matching of size $|A|$ then every connected component of K has a cycle.*

Proof. Let $K = (A, B; E)$ be a bipartite graph such that $d_K(y) \ge 2$ for all $y \in B$ such that K contains a matching of size $|A|$. Suppose for a contradiction that there is a connected component $K' = (A', B'; E')$ of K that is a tree. Then, $|E'| = |A'|+|B'|-1$. On the other hand, $|A'| \le |B'|$ since K' contains a matching of size $|A'|$. We thus obtain the following contradictory chain of inequalities:

$$|A'| + |B'| \le 2|B'| \le |E'| = |A'| + |B'| - 1.$$

This contradiction shows that every connected component of K contains a cycle. □

Proposition 3. *Let $K = (A, B; E)$ be a bipartite graph such that $d_K(y) \le 2$ for all $y \in B$. If every connected component of K has a cycle then K contains a matching of size $|A|$.*

Proof. Let $K = (A, B; E)$ be a bipartite graph with $d_K(y) \le 2$ for all $y \in B$ and such that every connected component of K has a cycle. By induction on the number of connected components, it is enough to consider the case when K is connected. Consider the multigraph K' defined by $V(K') = A$ and $E(K') = \{N_K(y) : y \in B \text{ and } d_K(y) = 2\}$. Since K is connected and contains a cycle, the same holds for K'. Let F denote a spanning subgraph of K' that contains precisely one cycle. Fix an orientation $\tilde{E}$ of the edges of F such that each vertex has out-degree one in $(A, \tilde{E})$: such an orientation can be obtained, for example, by orienting the edges in the unique cycle C in F in one of the two directions following the cycle, and orienting all the other edges toward C. By construction, the set $\{\{x, m(x)\} : x \in A\}$, where $m(x)$ is the element of B corresponding to the arc in $\tilde{E}$ with x as initial endpoint, forms a matching of size $|A|$ in K. □

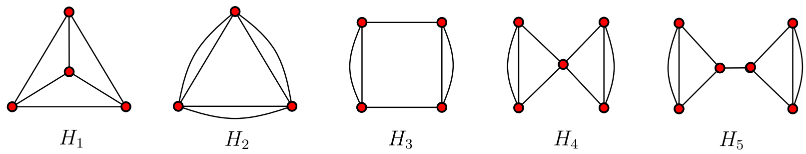

Fig. 2. Minimal strongly cyclic graphs.

Lemma 3. *Let $G = (L, R; E)$ be a bipartite graph such that $d_G(y) = 2$ for all $y \in R$, and let H denote the graph such that $G = S(H)$. Then, for every subset $J \subseteq L$, the following two conditions are equivalent:*

1. *set J induces an identifiable ℓ-subgraph of G;*
2. *the subgraph of H induced by J is strongly cyclic.*

Proof. First, suppose that $J \subseteq L$ induces an identifiable ℓ-subgraph of G. Let H' denote the subgraph of H induced by J, and let $v \in J$. Then, the subgraph $H' - v$ coincides with the graph $(J - \{v\}, E(v))$ where $E(v)$ denotes the multiset $\{N_G(u) : u \in R \wedge N(u) \subseteq J \setminus \{v\}\}$. Since J induces an identifiable ℓ-subgraph, the subgraph G' of $G(J)$ induced by the non-neighbors of v contains a matching M of size $|J| - 1$. Every vertex in $R \cap V(G')$ is of degree 2 in G'. Thus, Proposition 2 applies and we conclude that every connected component of G' has a cycle. Since G' is exactly the subdivision graph of $H' - v$, it follows that every connected component of $H' - v$ has a cycle. Since the choice of v was arbitrary, H' is strongly cyclic.

The converse direction can be proved similarly. Suppose that the subgraph of H induced by $J \subseteq V(H) = L$ (call it H') is strongly cyclic. Since H' is strongly cyclic, for every $v \in J$, every component of the subgraph $H' - v$ contains a cycle. This implies that every component of the subdivision graph of $H' - v$ contains a cycle. Notice that $S(H' - v)$ is exactly the subgraph G' of $G(J)$ induced by the non-neighbors of v. By Proposition 3, the graph G' contains a matching of size $|J| - 1$. Therefore, $J \subseteq L$ induces an identifiable ℓ-subgraph of G. □

Lemma 3 shows that in order to solve the IDS problem for G in linear time, it suffices to show that one can determine in linear time whether a given graph H contains an induced strongly cyclic subgraph. In what follows, we prove that this is indeed the case. This will be a consequence of a characterization of graphs containing an induced strongly cyclic subgraph in terms of topological minors, and some general results and techniques for graphs of bounded treewidth.

We start with the following simple property of strongly cyclic graphs.

Observation 4. *The class of strongly cyclic graphs is closed under edge addition.*

It follows that H contains an induced strongly cyclic subgraph if and only if H contains a (not necessarily induced) strongly cyclic subgraph. Fig. 2 shows five minimal strongly cyclic subgraphs.

It is easy to see that for every $i \in \{1, \ldots, 5\}$, the graph H_i is strongly cyclic while no proper subgraph of it is strongly cyclic.

Next, we observe that the class of strongly cyclic graphs is closed also under edge subdivision.

Lemma 5. *If a graph G is strongly cyclic and G' is the graph obtained from G by subdividing an edge of it, then G' is also strongly cyclic.*

Proof. Suppose that G' is obtained from G by replacing edge $uv \in E(G)$ by a path uwv. To show that G' is strongly cyclic, we need to show that for every vertex $x \in V(G')$, every component of $G' - x$ has a cycle.

If $x \in \{u, v\}$ then the components of $G' - x$ are exactly the same as the components of $G - x$, except that one component (the one containing the vertex in $\{u, v\} \setminus \{x\}$) has an additional vertex of degree 1. Thus, since G is strongly cyclic, every component of $G - x$ has a cycle, and so does every component of $G' - x$.

If $x = w$ then the graph $G' - x$ is the same as G with edge uv removed. Hence, every component of $G' - x$ contains, as a subgraph, a component of $G - u$. Together with the strong cyclicity of G, this implies that every component of $G' - x$ has a cycle.

If $x \notin \{u, v, w\}$ then every component of $G' - x$ is isomorphic to a subdivision of a component of $G - x$. Again, the conclusion easily follows from the assumption that G is strongly cyclic. □

Lemma 5 implies that all subdivisions of any of the five graphs from Fig. 2 are also strongly cyclic. In the next theorem, we prove that the presence of a subdivision of one of these graphs is not only a sufficient condition for the presence of a strongly cyclic subgraph, but also a necessary one. Recall that a graph G_1 is a *topological minor* of a graph G_2 if a subdivision of G_1 is isomorphic to a subgraph of G_2.

Theorem 6. *A graph H contains a strongly cyclic subgraph if and only if it contains one of the graphs $H_1, \ldots, H_5$ as a topological minor.*

Proof. Since sufficiency was already established, it is enough to show that every strongly cyclic graph contains a subgraph isomorphic to a subdivision of one of the graphs depicted in Fig. 2.

Suppose for a contradiction that this is not the case, and let H be a strongly cyclic graph that contains no subdivision of H_1, H_2, H_3, H_4 or H_5. Without loss of generality, we may assume that H is connected.

Claim 1: *H is 2-connected.*

Suppose not. Then, H contains at least two end blocks. Let B_1 and B_2 be two distinct end blocks of H, and let v_1 and v_2 be the unique cut vertices of H belonging to B_1 and B_2, respectively. Furthermore, for $i = 1, 2$, let C_i be a cycle in $B_i - v_i$. (Notice that C_1 and C_2 exist since H is strongly cyclic.) For $i = 1, 2$, since B_i is 2-connected, Proposition 1(ii) implies that B_i contains two internally vertex-disjoint (v_i, C_i)-paths, say P_i and Q_i, whose endpoints on C_i are distinct. However, the four paths P_1, Q_1, P_2, Q_2 together with the cycles C_1 and C_2 and a shortest (v_1, v_2)-path in H form a subdivision of either H_4 or H_5, depending on whether $v_1 = v_2$ or not; a contradiction. This proves Claim 1.

Claim 2: *Every two cycles in H have a vertex in common.*

Suppose not, and let C_1 and C_2 be two vertex-disjoint cycles in H. Since H is 2-connected, Proposition 1(i) implies that there exist in H two vertex-disjoint (C_1, C_2)-paths. These two paths, together with C_1 and C_2 form a subdivision of H_3; a contradiction. This proves Claim 2.

Now, fix an arbitrary vertex $v \in V(H)$ and a cycle C in $H - v$. Since H is 2-connected, Proposition 1(ii) implies there exist two internally disjoint (v, C)-paths P, Q, whose endpoints on C are distinct. Without loss of generality, P and Q are assumed to have no internal vertex in common with $V(C)$, which can be ensured simply by choosing them as minimal paths. Let x and y denote the endpoints of P and Q on C. Moreover, let P_1 and P_2 denote the two (x, y)-paths in C, and P_3 the (x, y)-path obtained as the union of P and Q. Then:

Claim 3: *Every cycle C' in $H - x$ contains y.*

Suppose not, and let C' be a cycle in $H - x$ such that $y \notin V(C')$. Since every two cycles in H have a vertex in common (by Claim 2), C' has a vertex in common with each of the cycles $C, P_1 \cup P_3, P_2 \cup P_3$. In particular, since $x, y \notin V(C')$, we infer that C' has a vertex in common with at least two out of the three paths $\mathrm{int}(P_1)$, $\mathrm{int}(P_2)$, and $\mathrm{int}(P_3)$. This implies that the following is well defined: let $i, j \in \{1, 2, 3\}$ and $\tilde{P}$ be chosen so that $|E(\tilde{P})|$ is minimized where $i \neq j$ and $\tilde{P}$ is a $\mathrm{int}(P_i)$–$\mathrm{int}(P_j)$ path in $H - \{x, y\}$. The minimality of $\tilde{P}$ implies that the graph $P_1 \cup P_2 \cup P_3 \cup \tilde{P}$ forms a subdivision of H_1 in H; a contradiction. This proves Claim 3.

Let $X = P_1 \cup P_2 \cup P_3$.

Claim 4: *Every cycle C' in $H - x$ satisfies $V(C') \cap V(X) = \{y\}$.*

Suppose not, and let C' be a cycle in $H - x$ such that $(V(C') \cap V(X)) \setminus \{y\} \neq \emptyset$. We claim that there exists a $(y, V(X) \setminus \{y\})$-path P' in $H - x$ such that no internal vertex of P' belongs to X, and $E(P') \cap E(X) = \emptyset$. Indeed, since $x \notin V(C')$ and $X - x$ is acyclic, C' contains an edge not in X. Follow C' from y in any of the two directions until an edge $e = v_1v_2 \notin E(X)$ is reached, and let $\hat{P} = (v_1, \ldots, v_\ell)$ be a maximal subpath of C' starting with e, not containing any edge of X and such that no internal vertex of $\hat{P}$ belongs to X. Notice that by maximality of $\hat{P}$, we have $v_\ell \in V(X)$. Also, $v_\ell \neq y$ by the choice of C'. If $v_1 = y$ then we can take $P' = \hat{P}$ and we are done. Otherwise, we distinguish two cases. If there exists an $i \in \{1, 2, 3\}$ such that $v_1, v_\ell \in V(P_i)$, then H would contain two vertex-disjoint cycles, namely the cycle formed by $\hat{P}$ together with the subpath of P_i from v_1 to v_ℓ, and the cycle formed by $P_j \cup P_k$ where $\{i, j, k\} = \{1, 2, 3\}$. This would contradict Claim 2. Suppose now that $v_1 \in V(P_i)$ and $v_\ell \in V(P_j)$ where $i, j \in \{1, 2, 3\}$ are distinct. In this case, the fact that no internal vertex of $\hat{P}$ belongs to $V(X)$ implies that the graph $P_1 \cup P_2 \cup P_3 \cup \hat{P}$ forms a subdivision of H_1 in H; a contradiction. This shows that there exists a $(y, V(X) \setminus \{y\})$-path P' in $H - x$ such that no internal vertex of P' belongs to X, and $E(P') \cap E(X) = \emptyset$.

Let z denote the endpoint of P' on $X \setminus \{y\}$, and let $i \in \{1, 2, 3\}$ be such that $z \in V(P_i)$. Let P'' be the unique (z, y)-path in $X \setminus \{x\}$, and let $\hat{C}$ denote the cycle obtained as the union of P' and P''. A symmetric argument to the one given in the above paragraph shows that there exists an $(x, V(X) \setminus \{x\})$-path Q' in $H - y$ such that no internal vertex of Q' belongs to X, and $E(Q') \cap E(X) = \emptyset$. In particular, this implies that there exists an $(x, (V(X) \setminus \{x\}) \cup V(P'))$-path Q'' in $H - y$ such that no internal vertex of Q'' belongs to $V(X) \cup V(P')$, and $E(Q'') \cap E(X) = \emptyset$ (to obtain Q'', just follow Q' from x, stopping as soon as a vertex of $(V(X) \setminus \{x\}) \cup V(P')$ is reached). Let w be the endpoint of Q'' other than x. We analyze several cases according to where w is. Since the path Q'' is in $H - y$, we have $w \neq y$. If $w = z$, then the graph $P_1 \cup P_2 \cup P_3 \cup P' \cup Q''$ forms a subdivision of H_2 in H; a contradiction. If $w \in V(\hat{C}) \setminus \{z, y\}$, then the graph $P_i \cup P_j \cup P' \cup Q''$ where $j \in \{1, 2, 3\} \setminus \{i\}$ forms a subdivision of H_1 in H; a contradiction. If $w \in V(P_i) \setminus V(\hat{C})$, then the graph $P_i \cup P_j \cup P' \cup Q''$ where $j \in \{1, 2, 3\} \setminus \{i\}$ forms a subdivision of H_3 in H; a contradiction. Finally, if $w \in V(P_j)$ for some $j \in \{1, 2, 3\} \setminus \{i\}$, then again the graph $P_i \cup P_j \cup P' \cup Q''$ where $j \in \{1, 2, 3\} \setminus \{i\}$ forms a subdivision of H_3 in H; a contradiction. In either case, we have a contradiction. This proves Claim 4.

We are now ready to complete the proof of Theorem 6. Let C' be a cycle in $H - x$, and let C'' be a cycle in $H - y$. By Claim 4 and symmetry, we have $V(C') \cap V(X) = \{y\}$ and $V(C'') \cap V(X) = \{x\}$. By Claim 2, C' and C'' have a vertex in common. We claim that C' and C'' have *precisely* one vertex in common. Indeed, if this is not the case, then following C'' from x in each of the two directions until C' is reached yields two internally disjoint (x, C')-paths R_1 and R_2 whose endpoints in C' are distinct, implying that the graph $P_1 \cup R_1 \cup R_2 \cup C'$ forms a subdivision of K_4 in H; a contradiction. But now, since the cycles C, C' and C'' intersect pairwise in a single vertex, the graph $C \cup C' \cup C''$ forms a subdivision of H_2 in H; a contradiction. □

It can be easily verified that the set $\{H_1, \ldots, H_5\}$ forms an antichain in the set of finite multigraphs partially ordered with respect to the topological minor relation. Thus, Theorem 6 above essentially characterizes the graphs H naturally associated to the NO instances of the 2-IDS problem (that is, graphs H such that $G = S(H)$ contains no ℓ-identifiable subgraph) in terms of minimal forbidden topological minors.

An immediate consequence of Lemmas 3 and 5 and Theorem 6 is the following:

Corollary 7. *A bipartite graph $G = (L, R; E) = S(H)$ with $d(x) = 2$ for all $x \in R$ contains an identifiable ℓ-subgraph if and only if H contains one of the graphs $H_1, \dots, H_5$ as a topological minor.*

It remains to show how to detect one of $H_1, \dots, H_5$ as a topological minor in linear time. Using the linear-time algorithm of Bodlaender [2], we first test whether H is of treewidth at most 2. Then, we proceed as follows:

(1) If the treewidth of H is at least 3, then H contains $H_1 = K_4$ as a topological minor (see, e.g., [6], Propositions 1.7.2(ii) and 12.4.2). Corollary 7 implies that in this case G contains an identifiable ℓ-subgraph.
(2) If the treewidth of H is at most 2, then the algorithm by Bodlaender will have output a tree decomposition of width at most 2. This decomposition can then be used to test in linear time whether H contains an H_i as a topological minor (for $i = 2, 3, 4, 5$), using the following two results:
   - The class of graphs that contain a fixed graph H_i as a topological minor is MSOL-definable (see, e.g., [7]).
   - By Courcelle's Theorem [5], every MSOL-definable property can be tested in linear time for graphs given by a tree decomposition of bounded width.

We can slightly simplify the description of the algorithm by performing the algorithms by Bodlaender [2] and Courcelle [5] directly on G. This can be achieved using the following two observations:

(1) Subdividing an edge does not change the treewidth of the graph (see, e.g., [15]). Hence H is of treewidth at most 2 if and only if G is of treewidth at most 2.
(2) For every $i = 1, \dots, 5$, graph H contains H_i as a topological minor if and only if $G = S(H)$ contains $S(H_i)$ as a topological minor.

Thus, Algorithm 1 below will correctly solve the IDS problem on graphs with $\Delta(R) \leq 2$. Lines 1–4 can be implemented in $O(|R|)$ time. Lines 5–7 can be implemented in $O(|L|+|R|+|E|)$ time [2]. Lines 8–13 can also be implemented in $O(|L|+|R|+|E|)$ time by applying the algorithm of Courcelle [5] at most 4 times. Hence, altogether Algorithm 1 can be implemented to run in time $O(|L| + |R| + |E|)$. This completes our proof of Theorem 1.

Algorithm 1: Solving the 2-BOUNDED IDENTIFIABLE SUBGRAPH problem.

Input: A bipartite graph $G = (L, R; E)$ such that $\Delta(R) \leq 2$.
Output: YES if G contains an identifiable ℓ-subgraph, NO otherwise.

1 **if** *there is a vertex $y \in R$ such that $d(y) = 1$* **then**
2   return YES;
3 **while** *there is an isolated vertex $y \in R$* **do**
4   delete y from G;
5 run the algorithm of Bodlaender [2] on G to test whether G is of treewidth at most 2;
6 **if** *the treewidth of G is at least* 3 **then**
7   return YES;
8 **else**
9   **for** $i = 2, \dots, 5$ **do**
10   using a tree-decomposition of width at most 2 (obtained in line 5) verify, using Courcelle's algorithm [5], whether G contains $S(H_i)$ as a topological minor;
11   **if** *G contains $S(H_i)$ as a topological minor* **then**
12   return YES;
13   return NO;

4.1. Finding an identifiable ℓ-subgraph in polynomial time

An identifiable ℓ-subgraph in a bipartite graph $G = (L, R; E)$ with $\Delta(R) \leq 2$ can also be *found* in polynomial time (if one exists). This is clearly the case if there is a vertex of degree 1 in R. Otherwise, by Lemma 3, a subset $J \subseteq L$ inducing an identifiable ℓ-subgraph will be given by the vertices belonging to L in a subgraph of G isomorphic to a subdivision of $S(H_i)$ (for some $i \in \{1, \dots, 5\}$). We can find $S(H_i)$ (for some $i \in \{1, \dots, 5\}$) as a topological minor in G by solving $O(|L|^{|V(H_i)|})$ instances of the disjoint paths problem with $k = |E(H_i)|$. As shown by Robertson and Seymour, this is a polynomially solvable task:

Theorem 8 (*[16]*)**.** *For every positive integer k, there is an $O(|V(G)|^3)$ time algorithm that solves the following* "DISJOINT PATHS" *problem: Given a simple graph G and k vertex pairs $(s_1, t_1), \dots, (s_k, t_k)$, do there exist paths $P_1, \dots, P_k$ in G, pairwise internally disjoint, such that P_i joins s_i and t_i ($1 \leq i \leq k$)?*

Since graphs $H_1, \dots, H_5$ have up to 6 vertices, the overall complexity of this approach is $O(|L|^9 + |R| + |E|)$.

An alternative solution is to test, for each $i \in \{1, \ldots, 5\}$, for the presence of $S(H_i)$ as a topological minor in G using the recent cubic algorithm of Grohe et al. [12]. First, we test whether G contains $S(H_i)$ as a topological minor. If so, we iteratively delete vertices and edges from G and test, at each step, for the presence of $S(H_i)$ as a topological minor in the resulting graph, restoring the deleted vertex or edge if its deletion resulted in a graph with no $S(H_i)$ as a topological minor. At the end of the algorithm, we will have a subgraph of G isomorphic to a subdivision of $S(H_i)$. The complexity of this approach is $O(|V(G)|^3(|V(G)| + |E(G)|))$.

5. Conclusions and further research

We showed in this paper that the identifiable subgraph problem is solvable in linear time if the maximum degree of vertices in R is at most 2. Our proof was based on a characterization of graphs containing a strongly cyclic subgraph in terms of topological minors, which may be of independent interest. It can also be derived from this characterization (Theorem 6) that *a graph H contains a strongly cyclic subgraph if and only if it has an H_i minor for some $i \in \{1, 2, 3, 4\}$*. (We omit the proof. The main idea is to verify that every graph such that contracting one of its edges results in a subdivision of one of H_1, H_2, H_3, H_4, contains one of H_1, H_2, H_3, H_4, H_5 as a topological minor.) In particular, this implies that the class of graphs not containing any strongly cyclic subgraphs is closed under minors, and that the class of strongly cyclic graphs is closed under vertex splitting (an operation inverse to edge contraction and generalizing edge subdivision).

There are several open questions and possibilities for further research in this area. For example:

- Compared to the linear-time algorithm from Theorem 1 for the decision version of the 2-IDS problem, the complexities of the two algorithms described in Section 4.1 to *find* an identifiable ℓ-subgraph in such a graph (if one exists) are rather high. Can an identifiable ℓ-subgraph in a bipartite graph $G = (L, R; E)$ with $\Delta(R) \leq 2$ (if one exists) also be found in linear time?
- What is the computational complexity of the IDENTIFIABLE SUBGRAPH problem? Equivalently, what is the computational complexity of the 3-BOUNDED IDENTIFIABLE SUBGRAPH problem?
- What is the complexity of the MAX-IDENTIFIABLE SUBGRAPH problem? What is the complexity of the "2-BOUNDED" MAX-IDENTIFIABLE SUBGRAPH problem?

Finally, as a more general scope for future research, we also think it would be interesting to find further uses of the natural notion of ℓ-subgraphs of (bipartitioned) bipartite graphs.

Acknowledgments

We are grateful to an anonymous reviewer for the many valuable comments. This work is supported in part by the Slovenian Research Agency, research program P1-0285 and research projects J1-5433, J1-6720, and J1-6743.

References

[1] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, Complexity and Approximation – Combinatorial Optimization Problems and Their Approximability Properties, 2nd edn., Springer, Heidelberg, 2002.
[2] H.L. Bodlaender, A linear-time algorithm for finding tree-decompositions of small treewidth, SIAM J. Comput. 25 (1996) 1305–1317.
[3] J.A. Bondy, U.S.R. Murty, Graph Theory, Springer, 2008.
[4] R. Boscolo, C. Sabatti, J.C. Liao, V.P. Roychowdhury, A generalized framework for network component analysis, IEEE Transactions on Computational Biology and Bioinformatics 2 (2005) 289–301.
[5] B. Courcelle, The monadic second-order logic of graphs. I. Recognizable sets of finite graphs, Inform. and Comput. 85 (1990) 12–75.
[6] R. Diestel, Graph Theory, 3rd edn., Springer, 2005.
[7] E. Fischer, T. Kotek, J.A. Makowsky, Application of logic to combinatorial sequences and their recurrence relations, in: M. Grohe, J. Makowsky (Eds.), Model Theoretic Methods in Finite Combinatorics, vol. 558, American Mathematical Society, 2011, pp. 1–42.
[8] J. Flum, M. Grohe, Parameterized Complexity Theory (Texts in Theoretical Computer Science. An EATCS Series), Springer-Verlag, New York, 2006, Inc., Secaucus, NJ, USA.
[9] E. Fritzilas, M. Milanič, S. Rahmann, Y. Rios-Solis, Structural identifiability in low-rank matrix factorization, Algorithmica 56 (2010) 313–332.
[10] E. Fritzilas, J. Monnot, M. Milanič, Y. Rios-Solis, Resilience and optimization of identifiable bipartite graphs, Discrete Appl. Math. 161 (2013) 593–603.
[11] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-completeness, Freeman, San Francisco, 1979.
[12] M. Grohe, K. Kawarabayashi, D. Marx, P. Wollan, Finding topological subgraphs is fixed-parameter tractable, in: Proceedings of the 43rd annual ACM symposium on Theory of computing, STOC '11, ACM, New York, NY, USA, 2011, pp. 479–488.
[13] D. Lee, H. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.
[14] J.C. Liao, R. Boscolo, Y.L. Yang, L.M. Tran, C. Sabatti, V.P. Roychowdhury, Network component analysis: Reconstruction of regulatory signals in biological systems, Proc. Natl. Acad. Sci. USA 100 (2003) 15,522–15,527.
[15] V. Lozin, D. Rautenbach, The tree- and clique-width of bipartite graphs in special classes, Australas. J. Combin. 34 (2006) 57–67.
[16] N. Robertson, P. Seymour, Graph minors. XIII. The disjoint paths problem, J. Combin. Theory Ser. B 63 (1995) 65–110.
[17] F. Shahnaz, M. Berry, V. Pauca, R. Plemmons, Document clustering using nonnegative matrix factorization, Inform. Process. Manag. 42 (2006) 373–386.
[18] H. Wang, E. Hubbell, J. Hu, G. Mei, M. Cline, G. Lu, T. Clark, M. Siani-Rose, M. Ares, D. Kulp, D. Haussler, Gene structure-based splice variant deconvolution using a microarray platform, Bioinformatics 19 (2003) i315–i322.