

# Identifying codes in line graphs<sup>☆</sup>

Florent Foucaud<sup>a</sup>, Sylvain Gravier<sup>b</sup>, Reza Naserasr<sup>a,b</sup>, Aline Parreau<sup>b</sup>, Petru Valicov<sup>a</sup>

<sup>a</sup>LaBRI - Université de Bordeaux - CNRS, 351 cours de la Libération, 33405 Talence cedex, France.

<sup>b</sup>Institut Fourier 100, rue des Maths, BP 74, 38402 St Martin d'Hères cedex, France.

## Abstract

An identifying code of a graph is a subset of its vertices such that every vertex of the graph is uniquely identified by the set of its neighbours within the code. We study the edge-identifying code problem, i.e. the identifying code problem in line graphs. If  $\gamma^{\text{ID}}(G)$  denotes the size of a minimum identifying code of an identifiable graph  $G$ , we show that the usual bound  $\gamma^{\text{ID}}(G) \geq \lceil \log_2(n+1) \rceil$ , where  $n$  denotes the order of  $G$ , can be improved to  $\Theta(\sqrt{n})$  in the class of line graphs. Moreover, this bound is tight. We also prove that the upper bound  $\gamma^{\text{ID}}(\mathcal{L}(G)) \leq 2|V(G)| - 5$ , where  $\mathcal{L}(G)$  is the line graph of  $G$ , holds (with two exceptions). This implies that a conjecture of R. Klasing, A. Kosowski, A. Raspaud and the first author holds for a subclass of line graphs. Finally, we show that the edge-identifying code problem is NP-complete, even for the class of planar bipartite graphs of maximum degree 3 and arbitrarily large girth.

*Keywords:* Identifying codes, Dominating sets, Line graphs, NP-completeness.

## 1. Introduction

An *identifying code* of a graph  $G$  is a subset  $C$  of vertices of  $G$  such that for each vertex  $x$ , the set of vertices in  $C$  at distance at most 1 from  $x$ , is nonempty and uniquely identifies  $x$ . In other words,  $C$  is both a *dominating* and a *separating set* of  $G$ , i.e. for each vertex  $v \in V(G)$ ,  $N[v] \cap C \neq \emptyset$  and for each pair  $u, v \in V(G)$ ,  $N[u] \cap C \neq N[v] \cap C$ . Here  $N[v]$  is the closed neighbourhood of  $v$  in  $G$ . This concept was introduced in 1998 in [15] and is a well-studied one (see e.g. [1, 5, 6, 9, 10, 13, 15, 18]).

A vertex  $x$  is a *twin* of another vertex  $y$  if  $N[x] = N[y]$ . A graph  $G$  is called *twin-free* if no vertex has a twin. The first observation regarding the concept of identifying codes is that a graph is *identifiable* if and only if it is twin-free. As usual for many other graph theory concepts, a natural problem in the study of identifying codes is to find one of a minimum size. Given a graph  $G$ , the smallest size of an identifying code of  $G$  is called *identifying code number* of  $G$  and denoted by  $\gamma^{\text{ID}}(G)$ . The main lines of research here are to find the exact value of  $\gamma^{\text{ID}}(G)$  for interesting graph classes, to approximate it and to give lower or upper bounds in terms of simpler graph parameters. Examples of classic results are as follows:

**Theorem 1.** [13] *If  $G$  is a twin-free graph with at least two edges then  $\gamma^{\text{ID}}(G) \leq |V(G)| - 1$ .*

The collection of all twin-free graphs reaching this bound is classified in [9].

A better upper bound in terms of both number of vertices and maximum degree of a graph is also proposed:

**Conjecture 2.** [10] *There exists a constant  $c$  such that for every twin-free graph  $G$ ,  $\gamma^{\text{ID}}(G) \leq |V(G)| - \frac{|V(G)|}{\Delta(G)} + c$ .*

Some support for this conjecture is provided in [9, 10, 11].

The parameter  $\gamma^{\text{ID}}(G)$  is also bounded below by a function of  $|V(G)|$  where equality holds for infinitely many graphs.

<sup>☆</sup>This research is supported by the ANR Project IDEA • ANR-08-EMER-007, 2009-2011.

**Theorem 3.** [15] For any twin-free graph  $G$ ,  $\gamma^{\text{ID}}(G) \geq \lceil \log_2(|V(G)| + 1) \rceil$ .

The collection of all graphs attaining this lower bound is also classified in [18].

From a computational point of view, it is shown that given a graph  $G$ , finding the exact value of  $\gamma^{\text{ID}}(G)$  is in the class of NP-hard problems. It in fact remains NP-hard for many subclasses of graphs [1, 5]. Furthermore, approximating  $\gamma^{\text{ID}}(G)$  is not easy either as shown in [16, 12, 19].

The problem of finding identifying codes in graphs can be viewed as a special case of the more general combinatorial problem of finding *transversals* in hypergraphs (a transversal is a set of vertices intersecting each hyperedge). More precisely, to each graph  $G$  one can associate the hypergraph  $\mathcal{H}(G)$  whose vertices are vertices of  $G$  and whose hyperedges are all the sets of the form  $N[v]$  and  $N[u] \ominus N[v]$  (symmetric difference of  $N[u]$  and  $N[v]$ ). Finding an identifying code for  $G$  is then equivalent to finding a transversal for  $\mathcal{H}(G)$ . Though the identifying code problem is captured by this more general problem, the structural properties of the graph from which the hypergraph is built allow one to obtain stronger results which are not true for general hypergraphs. In this work, we show that even stronger results can be obtained if we consider hypergraphs coming from line graphs. These stronger results follow from the new perspective of identifying edges by edges.

Given a graph  $G$  and an edge  $e$  of  $G$ , we define  $N[e]$  to be the set of edges adjacent to  $e$  together with  $e$  itself. An *edge-identifying code* of a graph  $G$  is a subset  $C_E$  of edges such that for each edge  $e$  the set  $N[e] \cap C_E$  is nonempty and uniquely determines  $e$ . When considering edge-identifying codes we will assume the edge set of the graph is not empty. It is easily observed that the notion of edge-identifying code of  $G$  is equivalent to the notion of identifying code of the line graph of  $G$ ,  $\mathcal{L}(G)$ . Thus a graph  $G$  admits an edge-identifying code if and only if  $\mathcal{L}(G)$  is twin-free. A pair of twins in  $\mathcal{L}(G)$  would correspond in  $G$  to a pair of: 1. parallel edges; 2. adjacent edges whose non common ends are of degree 1; 3. adjacent edges whose non common ends are of degree 2 but they are connected to each other. Since we will consider simple graphs, we will not have parallel edges. Two edges of the type described in the other two cases are called *pendant* and thus a graph is *edge-identifiable* if and only if it is *pendant-free*. The smallest size of an edge-identifying code of an edge-identifiable graph  $G$  is denoted by  $\gamma^{\text{EID}}(G)$  and is called *edge-identifying code number* of  $G$ .

To warm up we notice that five edges of a perfect matching of the Petersen graph  $P$ , form an edge-identifying code of this graph (see Figure 1). The lower bound of Theorem 3 proves that  $\gamma^{\text{EID}}(P) \geq 4$ . Later, by improving this bound for line graphs, we will see that in fact  $\gamma^{\text{EID}}(P) = 5$  (see Proposition 10 and Theorem 14).

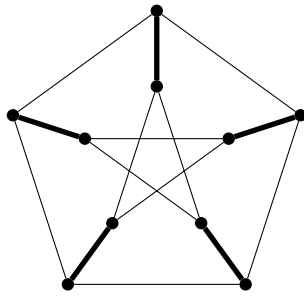


Figure 1: An edge-identifying code of the Petersen graph.

The outline of the paper is as follows: in Section 2, we give the edge-identifying code number of some basic families of graphs. In Section 3, we improve the general lower bound for the class of line graphs, then in Section 4 we improve the upper bound. Finally, in Section 5 we show that determining  $\gamma^{\text{EID}}(G)$  is also in the class of NP-hard problems even when restricted to planar subcubic bipartite graphs of arbitrarily large girth.

## 2. Preliminaries

In this section we first give some easy tools which help for finding the minimum edge-identifying code of graphs. We then apply these tools to determine the exact values of  $\gamma^{\text{EID}}$  for some basic families of graphs. We recall that  $C_n$  is the cycle on  $n$  vertices,  $P_n$  is the path on  $n$  vertices,  $K_n$  is the complete graph on  $n$  vertices and  $K_{n,m}$  is the complete bipartite graph with parts of size  $n$  and  $m$ .

**Lemma 4.** *Let  $G$  be a simple graph with girth at least 5. Let  $C_E$  be a subset of edges of  $G$  that covers vertices of  $G$  (that is, an edge cover of  $G$ ), such that the graph  $(V(G), C_E)$  is pendant-free. Then  $C_E$  is an edge-identifying code of  $G$ . In particular, if  $G$  has a perfect matching  $M$ ,  $M$  is an edge-identifying code of  $G$ .*

*Proof.* Since  $C_E$  covers all the vertices of  $G$ , it is an edge-dominating set. Thus we only need to prove that  $C_E$  is also an edge-separating set. Let  $e_1, e_2$  be two edges of  $G$ . If  $e_1, e_2 \in C_E$ , then  $C_E \cap N[e_1] \neq C_E \cap N[e_2]$  because  $(V(G), C_E)$  is pendant-free. Otherwise, we can assume that  $e_2 \notin C_E$ . If  $e_1 \in C_E$  and  $C_E \cap N[e_1] = C_E \cap N[e_2]$ , then  $e_2$  must be adjacent to  $e_1$ . Let  $u$  be their common vertex and  $e_2 = uv$ . Since  $C_E$  is an edge cover, there is an edge  $e_3 \in C_E$  which is incident to  $v$ . However,  $e_3$  cannot be adjacent to  $e_1$  because  $G$  is triangle-free. Therefore  $e_3$  separates  $e_1$  and  $e_2$ . Finally, we assume neither of  $e_1$  and  $e_2$  is in  $C_E$ . Then there are two edges of  $C_E$ , say  $e_3$  and  $e_4$ , adjacent to the two ends of  $e_1$ . But since  $G$  has neither  $C_3$  nor  $C_4$  as a subgraph,  $e_3$  and  $e_4$  cannot be both adjacent to  $e_2$  and, therefore,  $e_1$  and  $e_2$  are separated.  $\square$

We note that in the previous proof the absence of  $C_4$  is only used at the last part where  $e_1, e_2, e_3, e_4$  could induce a  $C_4$  which is not adjacent to any other edge of  $C_E$ . Thus, we have the following stronger statement:

**Lemma 5.** *Let  $G$  be a graph with girth at least 4. Let  $C_E$  be a subset of edges of  $G$  that covers vertices of  $G$ , such that  $C_E$  is pendant-free. If no pair of isolated edges in  $C_E$  induces a  $C_4$ , then  $C_E$  is an edge-identifying code of  $G$ .*

We will also need the following lemma about pendant-free trees.

**Lemma 6.** *If  $T$  is a pendant-free tree on more than two vertices, then  $T$  has two vertices of degree 1 each adjacent to a vertex of degree 2.*

*Proof.* Take a longest path in  $T$ , then it is easy to verify that both ends of this path satisfy the condition of the lemma.  $\square$

We are now ready to determine  $\gamma^{\text{EID}}$  of some families of graphs.

**Proposition 7.** *We have  $\gamma^{\text{EID}}(K_n) = \begin{cases} 5, & \text{if } n = 4 \text{ or } 5 \\ n - 1, & \text{if } n \geq 6 \end{cases}$ . Furthermore, let  $C_E$  be an edge-identifying code of  $K_n$  of size  $n - 1$  ( $n \geq 6$ ) and let  $G_1, G_2, \dots, G_k$  be the connected components of  $(V(K_n), C_E)$ . Then exactly one component, say  $G_i$ , is isomorphic to  $K_1$  and every other component  $G_j$  ( $j \neq i$ ) is isomorphic to a cycle of length at least 5.*

*Proof.* We note that  $\mathcal{L}(K_4)$  is isomorphic to  $K_6 \setminus M$ , where  $M$  is a perfect matching of  $K_6$ . This graph is known to have identifying code number 5. It is also an easy case analysis that  $K_5$  does not admit an edge-identifying code of size 4 and that edges of a  $C_5$  form an edge-identifying code of size 5 of  $K_5$ . Furthermore, it is not difficult to check that the set of edges of a cycle of length  $n - 1$  ( $n \geq 6$ ) identifies all edges of  $K_n$ . Thus we have  $\gamma^{\text{EID}}(K_n) \leq n - 1$ . The fact that  $\gamma^{\text{EID}}(K_n) \geq n - 1$  would follow from the second part of the theorem which is proved as follows.

Let  $C_E$  be an edge-identifying code of  $K_n$  of size  $n - 1$  or less ( $n \geq 6$ ). Let  $G' = (V(K_n), C_E)$ . Let  $G_1, G_2, \dots, G_k$  be the connected components of  $G'$ . Since  $G'$  has  $n$  vertices but only  $n - 1$  edges, at least one component of  $G'$  is a tree. On the other hand we claim that at most one of these components can be a tree and that such tree would be isomorphic to  $K_1$ . Let  $G_i$  be a tree. First we show that  $|V(G_i)| \leq 2$ . If not, by

Lemma 6 there is a vertex  $x$  of degree 1 in  $G_i$  with a neighbour  $u$  of degree 2. Let  $v$  be the other neighbour of  $u$ . Then the edges  $xv$  and  $uv$  are not identified. If  $V(G_i) = \{x, y\}$  then for any other vertex  $u$ , the edges  $ux$  and  $uy$  are not separated. Finally, if there are  $G_i$  and  $G_j$  with  $V(G_i) = \{x\}$  and  $V(G_j) = \{y\}$ , then the edge  $xy$  is not dominated by  $C_E$ . Thus exactly one component of  $G'$ , say  $G_1$ , is a tree and  $G_1 \cong K_1$ . This implies that  $\gamma^{\text{EID}}(K_n) \geq n - 1$ . Therefore,  $\gamma^{\text{EID}}(K_n) = n - 1$  and furthermore each  $G_i$ , ( $i \geq 2$ ), is a graph with a unique cycle.

It remains to prove that each  $G_i$ ,  $i \geq 2$  is isomorphic to a cycle of length at least 5. By contradiction suppose one of these graphs, say  $G_2$ , is not isomorphic to a cycle. Since  $G_2$  has a unique cycle, it must contain a vertex  $v$  of degree 1. Let  $t$  be the neighbour of  $v$  in  $G_2$  and let  $u$  be the vertex of  $G_1$ . Then the edges  $tv$  and  $tu$  are not separated by  $C_E$ . Finally we note that such cycle cannot be of length 3 or 4, because  $C_3$  is not pendant-free and in  $C_4$ , the two non-edges (which are edges of  $K_n$ ) would not be separated.  $\square$

**Proposition 8.**  $\gamma^{\text{EID}}(K_{n,n}) = \lceil \frac{3n-1}{2} \rceil$  for  $n \geq 3$

*Proof.* Let  $X$  and  $Y$  be the two parts of  $K_{n,n}$ . If  $n$  is even, then let  $\{A_i\}_{i=1}^{\frac{n}{2}}$  be a partition of vertices such that each  $A_i$  has exactly two vertices in  $X$  and two in  $Y$ . Let  $G_i$  be a subgraph of  $K_{n,n}$  isomorphic to  $P_4$  and with  $A_i$  as its vertices. If  $n$  is odd, let  $A_1$  be of size 2 and having exactly one element from  $X$  and one element from  $Y$  and let also  $G_1$  be the subgraph (isomorphic to  $K_2$ ) induced by  $A_1$ . Then we define  $A_i$ 's and  $G_i$ 's ( $i \geq 2$ ) as in the previous case (for  $K_{n-1, n-1}$ ). By Lemma 5, the set of edges in the  $G_i$ 's induces an edge-identifying code of  $K_{n,n}$  of size  $\lceil \frac{3n-1}{2} \rceil$ . To complete the proof we show that there cannot be any smaller edge-identifying code.

Let  $C_E$  be an edge-identifying code of  $G$ . Let  $G_1, G_2, \dots, G_k$  be the connected components of  $(X \cup Y, C_E)$ . The proof will be completed if we show that except possibly one, every  $G_i$  must have at least four vertices. To prove this claim we first note that there is no connected pendant-free graph on three vertices. We now suppose  $G_1$  and  $G_2$  are both of order 2. Then the two edges connecting  $G_1$  and  $G_2$  are not separated. If  $G_1$  is of order 1 and  $G_2$  is of order 2, then the edge connecting  $G_1$  to  $G_2$  is not identified from the edge of  $G_2$ . If  $G_1$  and  $G_2$  are both of order 1, then both of their vertices must be in the same part of the graph as otherwise the edge connecting them is not dominated by  $C_E$ . But now for any vertex  $x$  which is not in the same part as  $G_1$  and  $G_2$ , the edges connecting  $x$  to  $G_1$  and  $G_2$  are not separated.  $\square$

The following examples show that if true, the upper bound of Conjecture 2 is tight even in the class of line graphs. These examples were first introduced in [10] but without using the notion of edge-identifying codes.

**Proposition 9.** Let  $G$  be a  $k$ -regular multigraph ( $k \geq 3$ ). Let  $G_1$  be obtained from  $G$  by subdividing each edge exactly once. Then  $\gamma^{\text{EID}}(G_1) = (k-1)|V(G)| = |E(G_1)| - \frac{|E(G_1)|}{2k-2} = |V(\mathcal{L}(G_1))| - \frac{|V(\mathcal{L}(G_1))|}{\Delta(\mathcal{L}(G_1))}$ .

*Proof.* Let  $x$  be a vertex of  $G_1$  of degree at least 3 (an original vertex from  $G$ ). For each edge  $e_i^x$  incident to  $x$  let  $e_i'^x$  be the edge adjacent to  $e_i^x$  but not incident to  $x$  and let  $A_x = \{e_i'^x\}_{i=1}^k$ . Then  $\{A_x \mid x \in V(G)\}$  is a partition of  $E(G_1)$ . For any edge-identifying code  $C_E$  of  $G_1$ , if two elements of  $A_x$ , say  $e_1'^x$  and  $e_2'^x$ , are both not in  $C_E$ , then  $e_1^x$  and  $e_2^x$  are not separated. Thus  $|C_E \cap A_x| \geq k-1$ . This proves that  $|C_E| \geq (k-1)|V(G)|$ . But then it is easy to build an edge-identifying code of this size.  $\square$

Hypercubes, being the natural ground of code-like structures, have been a center of focus for determining the smallest size of their identifying codes. The problem has proved to be a challenging one from both theoretical and computational points of view. Today the precise identifying code number is known for only seven hypercubes [4]. In contrast, we show here that finding the edge-identifying code number of a hypercube is not so difficult. We first introduce the following general proposition.

**Proposition 10.** Let  $G$  be a connected pendant-free graph. We have:

$$\gamma^{\text{EID}}(G) \geq \frac{|V(G)|}{2}$$

*Proof.* Let  $C_E$  be an edge-identifying code of  $G$ . Let  $G'$  be the subgraph induced by  $C_E$  and let  $G_1, \dots, G_s$  be the connected components of  $G'$ . Let  $n_i$  be the order of  $G_i$  and  $k_i$  be its size (thus  $\sum_{i=1}^s k_i = |C_E|$ ). Let  $X = V(G) \setminus V(G')$  and  $n'_i$  be the number of vertices in  $X$  that are joined to a vertex of  $G_i$  in  $G$ . We show that  $n'_i + n_i \leq 2k_i$ . If  $k_i = 1$ , then clearly  $n'_i = 0$  and  $n'_i + n_i = 2 = 2k_i$ . If  $G_i$  is a tree, then  $n_i = k_i + 1$  and, by Lemma 6,  $G_i$  must have two vertices of degree 2 each having a vertex of degree 1 as a neighbour. Then no vertex of  $X$  can be adjacent to one of these two vertices in  $G$ . Moreover, each other vertex of  $G_i$  can be adjacent to at most one vertex in  $X$ . So  $n'_i \leq k_i - 1$ , and finally  $n_i + n'_i \leq 2k_i$ . If  $G_i$  is not a tree, we have  $n_i \leq k_i$  and  $n'_i \leq n_i$  and, therefore,  $n'_i + n_i \leq 2k_i$ . Finally, since  $G$  is connected, each vertex in  $X$  is connected to at least one  $G_i$ . Hence by counting the number vertices of  $G$  we have:

$$|V(G)| \leq \sum_{i=1}^s (n_i + n'_i) \leq 2 \sum_{i=1}^s k_i \leq 2|C_E|$$

□

Proposition 10 together with Lemma 4 leads to the following result:

**Corollary 11.** *Let  $G$  be an edge-identifiable graph with girth at least 5. If  $G$  has a perfect matching  $M$ , then  $M$  is an optimal edge-identifying code and  $\gamma^{\text{EID}}(G) = \frac{|V(G)|}{2}$ .*

This shows for example that the edge-identifying code of the Petersen graph given in Figure 1 is optimal. This result can be extended to graphs with girth 4 which have a perfect matching with no pair of edges of the matching that induces a  $C_4$ . This is the case for the hypercube of dimension  $d \geq 4$ . The hypercube of dimension  $d$ , denoted  $\mathcal{H}_d$ , is a graph whose vertices are elements of  $\mathbb{Z}_2^d$  with two vertices being adjacent if their difference is in the standard basis:  $\{(1, 0, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, \dots, 0, 1)\}$ . The hypercube of dimension  $d$  can also be viewed as the cartesian product of the hypercube of dimension  $d - 1$  and  $K_2$ . In this way of building  $\mathcal{H}_d$  we add a new coordinate to the left of the vectors representing the vertices of  $\mathcal{H}_{d-1}$ .

**Proposition 12.** *For  $d \geq 4$ , we have  $\gamma^{\text{EID}}(\mathcal{H}_d) = 2^{d-1}$ .*

*Proof.* By Proposition 10, we have  $\gamma^{\text{EID}}(\mathcal{H}_d) \geq 2^{d-1}$ . We will construct by induction a perfect matching  $M_d$  of  $\mathcal{H}_d$  such that no pair of edges induces a  $C_4$ , for  $d \geq 4$ . By Lemma 5,  $M_d$  will be an edge-identifying code of  $\mathcal{H}_d$ , proving the result. Two such matchings of  $\mathcal{H}_4$ , which are also disjoint, are presented in Figure 2. The matching  $M_5$  can now be built using each of these two matchings of  $\mathcal{H}_4$  - one matching per copy of  $\mathcal{H}_4$  in  $\mathcal{H}_5$ . It is easily verified that  $M_5$  has the required property. Furthermore,  $M_5$  has the extra property that for each edge  $uv$  of  $M_5$ ,  $|u - v| \neq (1, 0, 0, 0, 0)$ . We now build the matching  $M_d$  of  $\mathcal{H}_d$  ( $d \geq 6$ ) from  $M_{d-1}$  in such a way that no two edges of  $M_d$  belong to a 4-cycle in  $\mathcal{H}_d$  and that for each edge  $uv$  of  $M_d$ ,  $|u - v| \neq (1, 0, 0, \dots, 0)$ . To do this let  $\mathcal{H}'_1$  and  $\mathcal{H}'_2$  be the two copies of  $\mathcal{H}_{d-1}$  in  $\mathcal{H}_d$ . We then choose an exact copy of  $M_{d-1}$  in  $\mathcal{H}'_1$  and an image of  $M_{d-1}$  in  $\mathcal{H}'_2$  obtained from the following automorphism of  $\mathcal{H}_{d-1}$ :  $\sigma(x) = x + (1, 0, 0, \dots, 0)$ . It is now easy to check that the new matching has both properties we need. □

We note that the formula of Proposition 12 does not hold for  $d = 2$  and  $d = 3$ . For  $d = 2$  the hypercube  $\mathcal{H}_2$  is isomorphic to  $C_4$  and thus  $\gamma^{\text{EID}}(\mathcal{H}_2) = 3$ . For  $d = 3$ , we note that an identifying code of size 4, if it exists, must be a matching with no pair of edges belonging to a 4-cycle. But this is not possible. An identifying code of size 5 is shown in Figure 3, therefore  $\gamma^{\text{EID}}(\mathcal{H}_3) = 5$ .

### 3. Lower Bounds

Recall from Theorem 3 that  $\gamma^{\text{ID}}(G)$  is bounded below by a function of the order of  $G$ . As mentioned before, this bound is tight. Let  $C$  be a set of  $c$  isolated vertices. We can build a graph  $G$  of order  $2^c - 1$  such that  $C$  is an identifying code of  $G$ . To this end, for every subset  $X$  of  $C$  with  $|X| \geq 2$ , we associate a new vertex which is joined to all vertices in  $X$  and only to those vertices. Then, it is easily seen that  $C$  is an identifying code of this graph. However, the graph built in this way is far from being a line graph as it contains  $K_{1,t}$  for even large values of  $t$ . In fact this lower bound turns out to be far from being tight

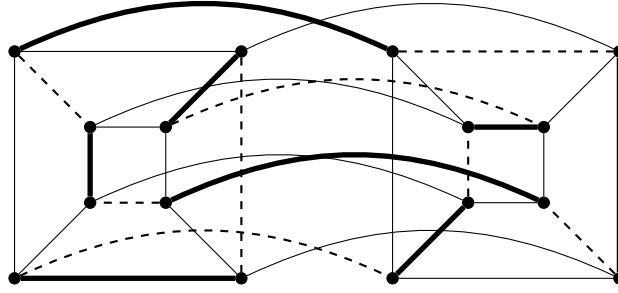


Figure 2: Two disjoint edge-identifying codes of  $\mathcal{H}_4$ .

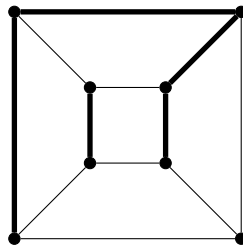


Figure 3: An optimal edge-identifying code of  $\mathcal{H}_3$ .

for the family of line graphs. In this section we give a tight lower bound on the size of an edge-identifying code of a graph in terms of the number of its edges. Equivalently we have a lower bound for the size of an identifying code in a line graph in terms of its order. This lower bound is of the order  $\Theta(\sqrt{n})$  and thus is a much improved lower bound with respect to the general bound of Theorem 3.

Let  $G$  be an edge-identifiable graph and let  $C_E$  be an edge-identifying code of  $G$ . To avoid trivialities such as having isolated vertices we may assume  $G$  is connected. We note that this does not mean that the subgraph induced by  $C_E$  is also connected, in fact we observe almost the contrary, i.e. in most cases, an edge-identifying code of a minimum size will induce a disconnected subgraph of  $G$ . We first prove a lower bound for the case when an edge-identifying code induces a connected subgraph.

**Theorem 13.** *If an edge-identifying code  $C_E$  of a nontrivial graph  $G$  induces a connected subgraph of  $G$  which is not isomorphic to  $K_2$ , then  $G$  has at most  $\binom{|C_E|+2}{2} - 4$  edges. Furthermore, inequality can only hold if  $C_E$  induces a path.*

*Proof.* Let  $G'$  be the subgraph induced by  $C_E$ . Since we assumed  $G'$  is connected, and since  $G'$  is pendant-free, it cannot have three vertices. Since we assumed  $G' \not\cong K_2$ , we conclude that  $G'$  has at least four vertices. For each vertex  $x$  of  $G'$ , let  $C_x$  be the set of all edges incident to  $x$  in  $G'$ . Let  $e = uv$  be an edge of  $G$ , then one or both of  $u$  and  $v$  must be in  $V(G')$  and, therefore,  $e$  is uniquely determined by either  $C_u$ , or  $C_v$ , or  $C_u \cup C_v$ . The total number of sets of this form is  $\binom{|V(G')|+1}{2}$ , thus if  $|V(G')| \leq |C_E|$  we are done. Otherwise, since  $G'$  is connected  $|V(G')| = |C_E| + 1$  and  $G'$  is a pendant-free tree on at least 4 vertices. If  $v$  is a vertex of degree 1 adjacent to  $u$ , then we have  $C_v = \{uv\}$  but  $uv \in C_u$  and, therefore,  $C_u = C_u \cup C_v$ . On the other hand, by Lemma 6, there are two vertices of degree 2 that have neighbours of degree 1. Let  $u$  be such a vertex, let  $v$  be its neighbour of degree 1 and  $x$  be its other neighbour. Then  $C_v = \{uv\}$  and  $C_u = \{uv, ux\}$  and, therefore,  $C_u \cup C_x = C_v \cup C_x$ . Thus the total number of distinct sets of the form  $C_y$  or  $C_y \cup C_z$  is at most  $\binom{|C_E|+2}{2} - 4$ . But if equality holds there can only be two vertices of degree 1 in  $G'$  and hence  $C_E$  is a path.  $\square$

We note that if this bound is tight, then  $G'$  is a path. Furthermore, for each path  $P_{k+1}$  one can build

many graphs which have  $P_{k+1}$  as an edge-identifying code and have  $\binom{k+2}{2} - 4$  edges. The set of all these graphs will be denoted by  $\mathcal{J}_k$ . An example of such a graph is obtained from  $K_{k+2}$  by removing a certain set of four edges as shown in Figure 4. Note that every other member of  $\mathcal{J}_k$  is obtained from the previous example by splitting the vertex that does not belong to  $P_{k+1}$  (but without adding any new edge).

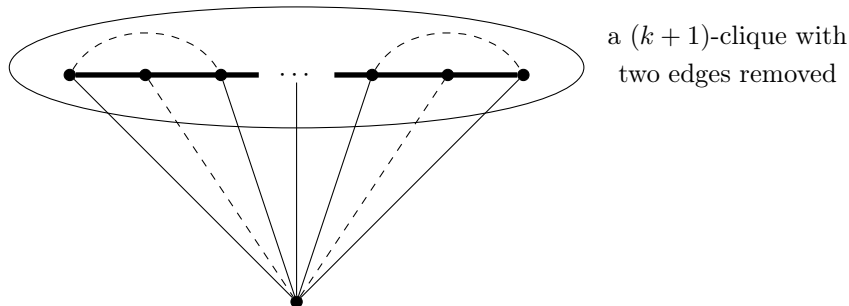


Figure 4: An extremal graph of  $\mathcal{J}_k$  with its connected edge-identifying code

Next we consider the case when the subgraph induced by  $C_E$  is not necessarily connected.

**Theorem 14.** *Let  $G$  be an edge-identifiable graph and let  $C_E$  be an edge-identifying code of  $G$  with  $|C_E| = k$ . Then we have:*

$$E(G) \leq \begin{cases} \binom{\frac{4}{3}k}{2}, & \text{if } k \equiv 0 \pmod{3} \\ \binom{\frac{4}{3}(k-1)+1}{2} + 1, & \text{if } k \equiv 1 \pmod{3} \\ \binom{\frac{4}{3}(k-2)+2}{2} + 2, & \text{if } k \equiv 2 \pmod{3} \end{cases}$$

*Proof.* Let  $G$  be a graph with maximum number of edges among all graphs with  $\gamma^{\text{EID}}(G) = k$ . It can be easily checked that for  $k = 1, 2$  or  $3$ , the maximum number of edges of  $G$  is  $1, 3$  or  $6$  respectively. For  $k \geq 4$ , we prove a slightly stronger statement: given a  $k$ -edge-identifying code  $C_E$  of  $G$ , all but at most two of the connected components of the subgraph induced by  $C_E$  must be isomorphic to  $P_4$ . When there is only one component not isomorphic to  $P_4$ , it must be isomorphic to a  $P_2$ , a  $P_5$  or a  $P_6$ . If there are two such components, then they can be two copies of  $P_2$ , a  $P_2$  with a  $P_5$ , or just two copies of  $P_5$ . This would depend on the value of  $k \pmod{3}$ .

To prove our claim let  $G$  be a graph as defined above, let  $C_E$  be an edge-identifying code of size  $k$  of  $G$  and let  $G'$  be the subgraph induced by  $C_E$ . For each vertex  $u \in V(G) \setminus V(G')$ , we can assume that  $u$  has degree 1: if  $u$  has degree  $d > 1$ , with neighbours  $v_1, \dots, v_d$  necessarily in  $V(G')$ , then replace  $u$  by  $d$  vertices of degree 1:  $u_1, \dots, u_d$ , connecting  $u_i$  to  $v_i$ . Then the number of edges does not change, and the code  $C_E$  remains an edge-identifying code of size  $k$ , thus it would suffice to prove our claim for this new graph. Let  $G'_1, G'_2, \dots, G'_r$  be the connected components of  $G'$  with  $|V(G'_i)| = n'_i$ . For each  $i \in \{1, \dots, r\}$ , let  $G_i$  be the graph induced by the vertices of  $G'_i$  and the vertices connected to  $G'_i$  only. To each vertex  $x$  of  $G'$  we assign the set  $C_x$  of edges in  $G'$  incident to  $x$ .

We first note that no  $G'_i$  can be of order 3, because there is no connected pendant-free graph on three vertices. If  $u$  and  $v$  are vertices from two disjoint components of  $G'$  with each component being of order at least 4, then the pair  $u, v$  is uniquely determined by  $C_u \cup C_v$ , thus by maximality of  $G$ ,  $uv$  is an edge of  $G$ . If a component of  $G'$  is isomorphic to  $K_2$ , assuming  $u$  and  $u'$  are vertices of this component, then for any other vertex  $v$  of  $G'$  exactly one of  $uv$  or  $u'v$  is an edge of  $G$ .

We now claim that each  $G'_i$  with  $n'_i \geq 4$  is a path. By contradiction, if a  $G'_i$  is not a path, we replace  $G_i$  by a member  $J_{n'_i-1}$  of  $\mathcal{J}_{n'_i-1}$  with  $P_{n'_i}$  being its edge-identifying code. Then we join each vertex of  $P_{n'_i}$  to each vertex of each  $G'_j$  (with  $j \neq i$  and  $n'_j \geq 4$ ) and to exactly one vertex of each  $G'_j$  with  $n'_j = 2$ . We note that the new graph is still  $k$ -edge-identifiable. However, it has more edges than  $G$ . Indeed, while the

number of edges connecting  $G'_i$  and the  $G'_j$ 's ( $j \neq i$ ) is not decreased, the number of edges in  $G_i$  is increased when we replace  $G_i$  by  $J_{n'_i-1}$ . This can be seen by applying Theorem 13 on  $G_i$ .

We now show that none of the  $G'_i$ 's can have more than six vertices. By contradiction, suppose  $G'_1$  is a component with  $n'_1 \geq 7$  vertices (thus  $n'_1 - 1$  edges). We build a new graph  $G_1^*$  from  $G$  as follows. We take disjoint copies of  $J_3 \in \mathcal{J}_3$  and  $J_{n'_1-4} \in \mathcal{J}_{n'_1-4}$  with  $P_4$  and  $P_{n'_1-3}$  being, respectively, their edge-identifying codes. We now let  $V(G_1^*) = V(J_3) \cup V(J_{n'_1-4}) \cup (V(G) \setminus V(G_1))$ . The edges of  $J_3$ ,  $J_{n'_1-4}$  and  $G - G_1$  are also edges of  $G_1^*$ . We then add edges between these three parts as follows. We join every vertex of  $P_4$  to each vertex of  $P_{n'_1-3}$ . For  $i = 2, 3, \dots, r$  if  $n'_i \geq 4$ , join every vertex of  $G'_i$  to each vertex of  $P_4 \cup P_{n'_1-3}$ . If  $n'_i = 2$ , we choose exactly one vertex of  $G'_i$  and join it to each vertex of  $P_4 \cup P_{n'_1-3}$ . The construction of  $G_1^*$  ensures that it still admits an edge-identifying code of size  $k$ , but it has more edges than  $G$ . In fact, the number of edges is increased in two ways. First, because  $P_4 \cup P_{n'_1-3}$  has one more vertex than  $G'_1$ , the number of edges connecting  $P_4 \cup P_{n'_1-3}$  to  $G - G_1$  has increased (unless  $r = 1$ ). More importantly, the number of edges induced by  $J_3 \cup J_{n'_1-4}$  is  $6 + \binom{n'_1-2}{2} - 4 + 4 \times (n'_1 - 3) = \frac{n'_1^2}{2} + \frac{3n'_1}{2} - 7$  which is strictly more than  $|E(G'_1)| = \frac{n'_1^2}{2} + \frac{n'_1}{2} - 4$  for  $n'_1 \geq 3$ . Since  $n'_1 \geq 7$ , this contradicts with the maximality of  $G$ .

With a similar method, the following transformations strictly increase the number of edges while the new graph still admits an edge-identifying code of size  $k$ :

1. Two components of  $G'$  each on six vertices transform into two graphs of  $\mathcal{J}_3$  and a graph of  $\mathcal{J}_4$ .
2. One component of  $G'$  on six vertices and another component on five vertices transform into three graphs of  $\mathcal{J}_3$ .
3. One component of  $G'$  on six vertices and one on two vertices transform into two graphs of  $\mathcal{J}_3$ .
4. Three components of  $G'$  each on five vertices transform into four graphs of  $\mathcal{J}_3$ .
5. Two components of  $G'$  on five vertices and one on two vertices transform into three graphs of  $\mathcal{J}_3$ .
6. A component of  $G'$  on five vertices and two on two vertices transform into two graphs of  $\mathcal{J}_4$ .
7. Three components of  $G'$  each isomorphic to  $P_2$  transform into a graph of  $\mathcal{J}_3$ .

For the proof of case 7, we observe that the number of edges identified by the three  $P_2$ 's would be the same as the number of edges identified by the  $P_4$ . However, since  $k \geq 4$ , there must be some other component in  $G'$ . Moreover, the number of vertices of the three  $P_2$ 's, which are joined to the vertices of the other components of  $G'$ , is three, whereas the number of these vertices of the  $P_4$ , is four. Hence the maximality of  $G$  is contradicted.

We note that cases 1, 2 and 3 imply that if a component of  $G'$  is isomorphic to  $P_6$ , every other component is isomorphic to  $P_4$ . Then cases 4, 5 and 6 imply that if a component is isomorphic to  $P_5$ , then at most one other component is not isomorphic to  $P_4$  and such component is necessarily either a  $P_2$  or a  $P_5$ . Finally, case 7 shows that there can be at most two components both isomorphic to  $P_2$ .

We conclude that each of the components of  $G'$  is isomorphic to  $P_4$  except for possibly two of them. These exceptions are dependent on the value of  $k \bmod 3$  as we described. Hence, the formulas of the theorem can be derived using these structural properties of  $G$ .  $\square$

We note that this bound is tight and the examples were in fact built inside the proof. More precisely, for  $k \equiv 0 \pmod 3$  we take  $\frac{k}{3}$  disjoint copies of elements of  $\mathcal{J}_3$  each having a  $P_4$  as an edge-identifying code. We then add an edge between each pair of vertices coming from two distinct such  $P_4$ 's. We note that the union of these  $P_4$ 's is a minimum edge-identifying code of the graph. If  $k \not\equiv 0 \pmod 3$ , then we build a similar construction. This time we use elements from  $\mathcal{J}_3$  with at most two exceptions that are elements of  $\mathcal{J}_4$  or  $\mathcal{J}_5$ .

The above theorem can be restated in the language of line graphs as follows.

**Corollary 15.** *Let  $G$  be a line graph. Then we have  $\gamma^{ID}(G) \geq \frac{3}{8}(1 + \sqrt{1 + 8|V(G)|}) > \frac{3\sqrt{2}}{4}\sqrt{|V(G)|}$ .*



**Remark.** Note that the lower bound of  $\gamma^{\text{ID}}(G) \geq \Theta(\sqrt{|V(G)|})$ , which holds for the class of line graphs, is also implied by Proposition 10. In [2], Beineke characterized line graphs by a list of nine forbidden induced subgraphs. Considering Beineke's characterization, the lower bound of Corollary 15 can be restated as follows:  $\gamma^{\text{ID}}(G) \geq \Theta(\sqrt{|V(G)|})$  holds if  $G$  has no induced subgraph from Beineke's list. It is then natural to ask what is a minimal list of forbidden induced subgraphs for which a similar claim would hold. Note that the claw graph,  $K_{1,3}$ , belongs to Beineke's list of forbidden subgraphs. However, we remark that our bound does not hold for the class of claw-free graphs. Examples can be built as follows: let  $A$  be a set of size  $k$  and let  $B$  be the set of nonempty subsets of  $A$ . Let  $G$  be the graph built on  $A \cup B$ , where  $A$  and  $B$  each induce a complete graph and a vertex  $a$  of  $A$  is joined to a vertex  $b$  of  $B$  if  $a \in b$ . This graph is claw-free and it is easy to find an identifying code of size at most  $2k = \Theta(\log |V(G)|)$  in  $G$ .

#### 4. Upper bounds

The most natural question in the study of identifying codes in graphs is to find an identifying code as small as possible. A general bound, only in terms of the number of vertices of a graph, is provided by Theorem 1. Furthermore, the class of all graphs with  $\gamma^{\text{ID}}(G) = |V(G)| - 1$  is classified in [9]. It is easy to check that none but six of these graphs are line graphs. Thus we have the following corollary (where  $G \bowtie H$  denotes the complete join of graphs  $G$  and  $H$ ):

**Corollary 16.** *If  $G$  is a line graph with  $G \notin \{P_3, P_4, C_4, P_4 \bowtie K_1, C_4 \bowtie K_1, \mathcal{L}(K_4)\}$ , then we have  $\gamma^{\text{ID}}(G) \leq |V(G)| - 2$ .*

Since  $\gamma^{\text{EID}}(K_{2,n}) = 2n - 2$ ,  $\gamma^{\text{ID}}(\mathcal{L}(K_{2,n})) = |V(\mathcal{L}(K_{2,n}))| - 2$  and the bound of Corollary 16 is tight for an infinite family of graphs. Conjecture 2 proposes a better bound in terms of both the number of vertices and the maximum degree of a graph. As pointed out in Proposition 9, most of the known extremal graphs for Conjecture 2 are line graphs. In this section, after proving some general bound for the edge-identifying code number of a pendant-free graph we will show that Conjecture 2 holds for the class of line graphs of high enough density.

We recall that a graph on  $n$  vertices is *2-degenerated* if its vertices can be ordered  $v_1, v_2, \dots, v_n$  such that each  $v_i$  is joined to at most two vertices in  $\{v_1, v_2, \dots, v_{i-1}\}$ . Our main idea of proving upper bounds is to show that given a pendant-free graph  $G$ , any minimal edge-identifying code  $C_E$  induces a 2-degenerated subgraph of  $G$  and hence  $|C_E| \leq 2|V(G)| - 3$ . Our proofs are constructive and one could build such small edge-identifying codes.

**Theorem 17.** *Let  $G$  be a pendant-free graph and let  $C_E$  be a minimal edge-identifying code of  $G$ . Then  $G'$ , the subgraph induced by  $C_E$ , is 2-degenerated.*

*Proof.* Let  $uv$  be an edge of  $G'$  with  $d_{G'}(u), d_{G'}(v) \geq 3$ . By minimality of  $C_E$  the subset  $C' = C_E - uv$  of  $E(G)$  is not an edge-identifying code of  $G$ . By the choice of  $uv$ ,  $C'$  is still an edge-dominating set, thus there must be two edges,  $e_1$  and  $e_2$ , that are not separated by  $C'$ . Hence one of them, say  $e_1$ , is incident to either  $u$  or  $v$  and the other one ( $e_2$ ) is not.

We consider two cases: either  $e_1 = uv$  or  $e_1$  is incident to only one of  $u$  and  $v$ . In the first case,  $e_2$  is adjacent to every edge of  $C'$  which  $uv$  is adjacent to. Since for each vertex of  $uv$  there are at least two edges in  $C'$  incident to this vertex, the subgraph induced by  $u, v$  and the vertices of  $e_2$  must be isomorphic to  $K_4$  and there should be no other edge of  $C'$  incident to any vertex of this  $K_4$  (see Figure 5(a)).

In the other case, suppose  $e_1$  is adjacent to  $uv$  at  $u$ . Let  $x$  and  $y$  be two neighbours of  $u$  in  $G'$  other than  $v$ . Then it follows that  $e_2 = xy$  and, therefore,  $d_{G'}(u) = 3$ . Let  $z$  be the other end of  $e_1$ . We consider two subcases: either  $z \notin \{x, y\}$ , or, without loss of generality,  $z = x$ . Suppose  $z \notin \{x, y\}$ . Recall that  $uv$  is the only edge separating  $e_1$  and  $e_2$ , but  $e_1$  must be separated from both  $ux$  and  $uy$ . Thus  $zy \in C_E$  and by a similar argument  $zx \in C_E$ . Furthermore,  $d_{G'}(x) = d_{G'}(y) = d_{G'}(z) = 2$  and  $\{x, y, z, u\}$  induces a  $C_4$  in  $G'$  (see Figure 5(b)). Now suppose  $e_1 = ux$ , since  $uv$  is the only edges separating  $e_1$  and  $e_2$ , then  $uy$  and possibly  $xy$  are the only edges in  $G'$  incident to  $y$ , so  $d_{G'}(y) \leq 2$  and  $d_{G'}(u) = 3$  (see Figures 5(c) and 5(d)).

To summarize, we proved that given an edge  $uv$ , in a minimal edge-identifying code  $C_E$ , we have one of the following cases.

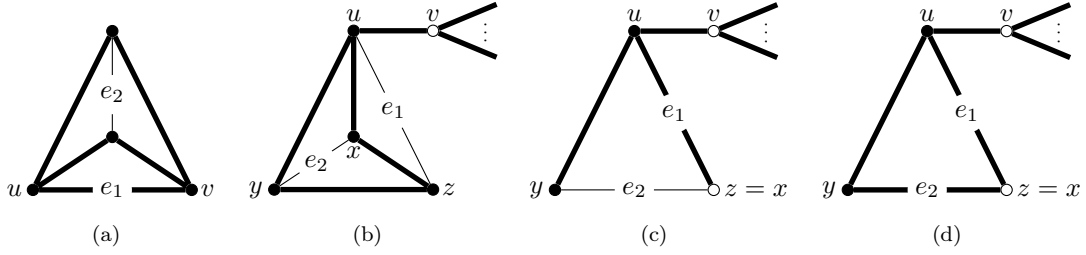


Figure 5: Case distinctions in the proof of Theorem 17. Black vertices have fixed degree in  $G'$ .

- One of  $u$  or  $v$  is of degree at most 2 in  $G'$ .
- Edge  $uv$  is an edge of a connected component of  $G'$  isomorphic to  $K_4^-$  (that is  $K_4$  with an edge removed), see Figure 5(a).
- $d_{G'}(u) = 3$  (considering the symmetry between  $u$  and  $v$ ) in which case either  $u$  is incident to a  $C_4$  whose other vertices are of degree 2 in  $G'$  (Figure 5(b)), or to a vertex of degree 1 in  $G'$  (Figure 5(c)) or to a triangle with one vertex of degree 2 in  $G'$  (Figure 5(d)).

In either case, there exists a vertex  $x$  of degree at most 2 in  $G'$  such that when  $x$  is removed, either  $u$  or  $v$  becomes of degree at most 2 in the remaining subgraph of  $G'$ . In this way we can define an order of elimination of the vertices of  $G'$  showing that  $G'$  is 2-degenerated.  $\square$

By further analysis of our proof we prove the following:

**Corollary 18.** *If  $G$  is a pendant-free graph on  $n$  vertices not isomorphic to  $K_4$  or  $K_4^-$ , then  $\gamma^{\text{EID}}(G) \leq 2n - 5$ .*

*Proof.* We first prove that if  $G$  is a pendant-free graph on  $n$  vertices not isomorphic to  $K_4$ , then  $\gamma^{\text{EID}}(G) \leq 2n - 4$ . Let  $C_E$  be a minimal edge-identifying code and let  $G'$  be the subgraph induced by  $C_E$ . Then, by Theorem 17,  $G'$  is 2-degenerated. Let  $v_n, v_{n-1}, \dots, v_1$  be a sequence of vertices of  $G'$  obtained by a process of eliminating vertices of degree at most 2. Since  $v_1$  and  $v_2$  can induce at most a  $K_2$ , we notice that there could only be at most  $2n - 3$  edges in  $G'$ . But if there are exactly  $2n - 3$ , at every step of the vertex-elimination process, the vertex to be removed must have degree exactly 2. Hence, the subgraph induced by  $v_1, v_2, v_3$  and  $v_4$  is isomorphic to  $K_4^-$ . Considering symmetries, there are three possibilities for the subgraph induced by  $\{v_1, \dots, v_5\}$  (recall that in this graph  $v_5$  is of degree 2). In each of these three cases, one can easily find an edge with both ends of degree at least 3 which also does not belong to any of the four configurations described in the proof of Theorem 17 (see Figure 5), which is a contradiction.

Now we show that if  $\gamma^{\text{EID}}(G) = 2n - 4$ , then  $G \cong K_4^-$ . This can be easily checked if  $G$  has at most four vertices, so we may assume  $n \geq 5$ . Let  $G''$  be the subgraph of  $G'$  induced by  $\{v_1, v_2, v_3, v_4, v_5\}$ . If  $G''$  has seven edges, then we could repeat the argument of the previous paragraph, to obtain a contradiction. Thus we may assume that  $G''$  has exactly six edges and, since it is 2-degenerated, by an easy case analysis, it must be isomorphic to one of the graphs of Figure 6.

If  $G''$  is a graph in part (a), (b) or (c) of Figure 6, then one could repeat the proof of Theorem 17 with the edge  $uv$  of the corresponding figure, to obtain a contradiction.

If  $G''$  is isomorphic to the graph in part (d), since it is not pendant-free there must be a vertex  $v_6$  which is joined to exactly two vertices of  $G''$ . Without loss of generality, let  $u$  be a vertex of degree 2 in  $G''$  which is adjacent to  $v_6$ . We first repeat the argument of the proof of Theorem 17 using the edge  $uv$ . Of the four possible end-configurations of Figure 5, one can check that only the last one is possible, in which case  $v_6$  must also be adjacent to  $v'$ . But then we repeat the same argument, this time on the edge  $uv'$ , the only possible case being the one of Figure 5(a). The connected component of  $G'$  containing  $uv$  should be isomorphic to  $K_4^-$  — a contradiction.

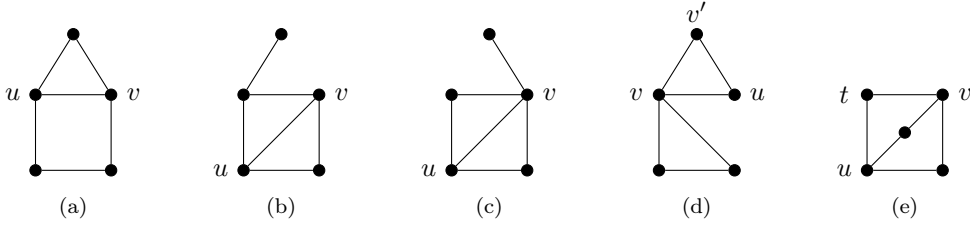


Figure 6: The five possibilities of 2-degenerated graphs on five vertices with six edges.

Finally, let  $G''$  be isomorphic to the graph of Figure 6(e). If there is a vertex  $v_6$  and it is adjacent to a vertex other than  $u$  and  $v$ , say to  $t$ , then, by repeating the argument of the proof of Theorem 17, we imply that  $v_6$  must be adjacent to  $t$  and  $u$ . But if we apply the same argument on  $tv$  we will get a contradiction. Hence  $v_6$  and, therefore, by symmetry, every other vertex is joined, in  $G'$ , to only  $u$  and  $v$ . Furthermore, since  $|E(G')| = |C_E| = 2n - 4$ ,  $G'$  is a spanning subgraph of  $G$ . But then it is easy to verify that  $C_E \setminus \{xu, xv\}$  is an edge-identifying code of  $G$  — a contradiction.  $\square$

We note that  $\gamma^{\text{EID}}(K_{2,n}) = 2n - 2 = 2|V(K_{2,n})| - 6$  thus this bound cannot be improved much.

Corollary 18 implies that Conjecture 2 holds for a large subclass of line graphs:

**Corollary 19.** *If  $G$  is a pendant-free graph on  $n$  vertices and with average degree  $\bar{d}(G) \geq 5$ , then we have  $\gamma^{\text{ID}}(\mathcal{L}(G)) \leq n - \frac{n}{\Delta(\mathcal{L}(G))}$ .*

*Proof.* Let  $u$  be a vertex of degree  $d(u) \geq \bar{d}(G) \geq 5$ . Since  $G$  is pendant-free there is at least one neighbour  $v$  of  $u$  that is of degree at least 2. Thus there is an edge  $uv$  in  $G$  with  $d(u) + d(v) \geq \bar{d}(G) + 2$  and, therefore,  $\Delta(\mathcal{L}(G)) \geq \bar{d}(G)$ . Hence, considering Corollary 18, it is enough to show that  $2|V(G)| - 5 \leq |E(G)| - \frac{|E(G)|}{\bar{d}(G)}$ .

To this end, since  $\bar{d}(G) \geq 5$ , we have  $4|V(G)| \leq (\bar{d}(G) - 1)|V(G)|$ , therefore,

$$4|V(G)| - 10 \leq (\bar{d}(G) - 1)|V(G)|$$

Multiplying both sides by  $\frac{\bar{d}(G)}{2}$  we have:

$$(2|V(G)| - 5)\bar{d}(G) \leq (\bar{d}(G) - 1)\frac{\bar{d}(G)}{2}|V(G)| = (\bar{d}(G) - 1)|E(G)|$$

$\square$

## 5. Complexity

This section is devoted to the study of the decision problem associated to the concept of edge-identifying codes. Let us first define the decision problems we use. The IDCODE problem is defined as follows:

### IDCODE

INSTANCE: A graph  $G$  and an integer  $k$ .

QUESTION: Does  $G$  have an identifying code of size at most  $k$ ?

IDCODE was proved to be NP-complete even when restricted to the class of bipartite graphs of maximum degree 3 (see [5]) or to the class of planar graphs of maximum degree 4 and arbitrarily large girth (see [1]). The EDGE-IDCODE problem is defined as follows:

### EDGE-IDCODE

INSTANCE: A graph  $G$  and an integer  $k$ .

QUESTION: Does  $G$  have an edge-identifying code of size at most  $k$ ?

We will prove that EDGE-IDCODE is NP-hard in some restricted class of graphs by reduction from PLANAR  $(\leq 3, 3)$ -SAT, which is a variant of the SAT problem and is defined as follows [8]:

**PLANAR  $(\leq 3, 3)$ -SAT**

INSTANCE: A collection  $\mathcal{Q}$  of clauses over a set  $X$  of boolean variables, where each clause contains at least two and at most three distinct literals (a variable  $x$  or its negation  $\bar{x}$ ). Moreover, each variable appears in exactly three clauses: twice in its non-negated form, and once in its negated form. Finally, the bipartite incidence graph of  $\mathcal{Q}$ , denoted  $B(\mathcal{Q})$ , is planar ( $B(\mathcal{Q})$  has vertex set  $\mathcal{Q} \cup X$  and  $Q \in \mathcal{Q}$  is adjacent to  $x \in X$  if  $x$  or  $\bar{x}$  appears in clause  $Q$ ).

QUESTION: Can  $\mathcal{Q}$  be satisfied, i.e. is there a truth assignment of the variables of  $X$  such that each clause contains at least one true literal?

PLANAR  $(\leq 3, 3)$ -SAT is known to be NP-complete [8]. We are now ready to prove the main result of this section.

**Theorem 20.** *EDGE-IDCODE is NP-complete even when restricted to bipartite planar graphs of maximum degree 3 and arbitrarily large girth.*

*Proof.* The problem is clearly in NP: given a subset  $C$  of edges of  $G$ , one can check in polynomial time whether it is an edge-identifying code of  $G$  by computing the sets  $C \cap N[e]$  for each edge  $e$  and comparing them pairwise.

We now reduce PLANAR  $(\leq 3, 3)$ -SAT to EDGE-IDCODE. We first give the proof for the case of girth 8 and show that it can be easily extended to an arbitrarily large girth.

We first need to define a generic sub-gadget (denoted  $P$ -gadget) that will be needed for the reduction. In order to have more compact figures, we will use the representation of this gadget as drawn in Figure 7. We will say that a  $P$ -gadget is *attached* at some vertex  $x$  if  $x$  is incident to edge  $a$  of the gadget as depicted in the figure.

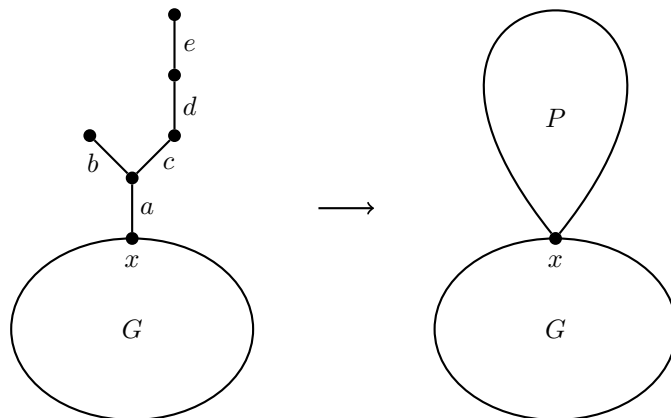


Figure 7: The generic  $P$ -gadget

We make the following claims.

**Claim 1.** *In any graph containing a  $P$ -gadget, at least three edges of this gadget must belong to any edge-identifying code.*

Claim 1 is true because  $d$  is the only edge separating  $b$  and  $c$ . Similarly  $c$  is the only edge separating  $d$  and  $e$ . Finally, in order to separate  $d$  and  $c$ , one has to take at least one of  $a$ ,  $b$  or  $e$ .

**Claim 2.** *If  $G$  is a pendant-free graph obtained from a graph  $H$  with a  $P$ -gadget attached at a vertex  $x$  of  $H$ , then any edge-identifying code of  $G$  must contain an edge of  $H$  incident to  $x$ .*

Claim 2 follows from the fact that edge  $a$  must be separated from edge  $b$ .

We are now ready to describe the reduction.

Given an instance  $\mathcal{Q} = \{Q_1, \dots, Q_m\}$  of PLANAR ( $\leq 3, 3$ )-SAT over the set of boolean variables  $X = \{x_1, \dots, x_n\}$  together with an embedding of its bipartite incidence graph  $B(\mathcal{Q})$  in the plane, we build the graph  $G_{\mathcal{Q}}$  as follows.

For each variable  $x_j$  and clause  $Q_i$  we build the subgraphs  $G_{x_j}$  and  $G_{Q_i}$  respectively, as shown in Figure 8. We recall that a given variable  $x_j$  appears in positive form in exactly two clauses, say  $Q_p, Q_q$ , and in negative form in exactly one clause, say  $Q_r$ . We then unify vertex  $x_j^1$  of  $G_{x_j}$  with vertex  $l_{pk}$  of  $G_{Q_p}$  which corresponds to  $x_j$ . We do a similar unification for vertices  $x_j^2$  and  $\bar{x}_j^{-1}$  with corresponding vertices from  $G_{Q_q}$  and  $G_{Q_r}$ . This can be done while ensuring the planarity of  $G_{\mathcal{Q}}$ , using the given planar embedding of  $B(\mathcal{Q})$ . Moreover,  $G_{\mathcal{Q}}$  is bipartite since  $B(\mathcal{Q})$  is bipartite and there are no odd cycles in the variable and clause gadgets. Finally, it is easy to see that  $G_{\mathcal{Q}}$  has maximum degree 3 and girth 8. Since a clause gadget has forty-five vertices and a variable gadget, forty-two vertices,  $G_{\mathcal{Q}}$  has  $45m + 42n$  vertices and, therefore, the construction has polynomial size in terms of the size of  $\mathcal{Q}$ .

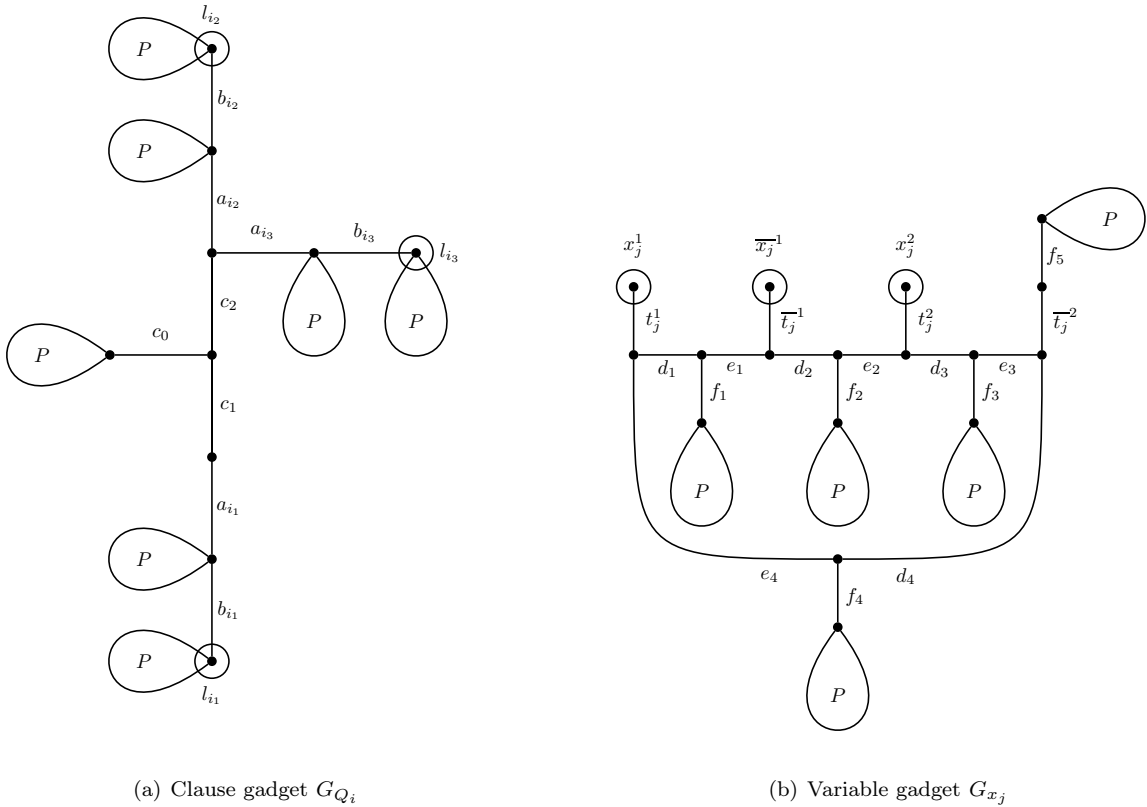


Figure 8: Reduction gadgets for clause  $Q_i$  and variable  $x_j$ .

We will need two additional claims in order to complete the proof.

**Claim 3.** *In a variable gadget  $G_{x_j}$ , in order to separate the four pairs of edges  $\{d_i, e_i\}$  for  $1 \leq i \leq 4$ , at least two edges of  $A = \{d_i, e_i \mid 1 \leq i \leq 4\} \cup \{t_j^1, \bar{t}_j^1, t_j^2, \bar{t}_j^2\}$  belong to any edge-identifying code  $C$ . Moreover, if  $|C \cap A| = 2$ , then either  $C \cap A = \{t_j^1, t_j^2\}$  or  $C \cap A = \{\bar{t}_j^1, \bar{t}_j^2\}$ .*

The first part of Claim 3 follows from the fact that the two edges of each of the pairs  $\{d_1, e_1\}$  and  $\{d_3, e_3\}$  must be separated. The second follows from an easy case analysis.

The following claim follows directly from Claim 2.

**Claim 4.** *Let  $v_1 v_2 v_3 v_4$  be a path of four vertices where each of the vertices  $v_2$  and  $v_3$  has its own  $P$ -gadget attached and both  $v_2$  and  $v_3$  have degree 3. Then, at least one of the three edges of the path belong to any identifying code of the graph. If exactly one belongs to a code, it must be  $v_2 v_3$ .*

We now claim that  $\mathcal{Q}$  is satisfiable if and only if  $G_{\mathcal{Q}}$  has an edge-identifying code of size at most  $k = 25m + 22n$ .

For the sufficient side, given a truth assignment of the variables satisfying  $\mathcal{Q}$ , we build an edge-identifying code  $C$  as follows. For each  $P$ -gadget, edges  $a, c, d$  are in  $C$ . For each clause gadget  $G_{Q_i}$ , edge  $c_0$  belongs to  $C$ . For each literal  $l_{i_k}$  of  $Q_i$ ,  $1 \leq k \leq 3$ , if  $l_{i_k}$  is true, edge  $a_{i_k}$  belongs to  $C$ ; otherwise, edge  $b_{i_k}$  belongs to  $C$ . If  $Q_i$  has only two literals and vertex  $l_{i_k}$  is the vertex not corresponding to a literal of  $Q_i$ , then edge  $b_{i_k}$  belongs to  $C$ . Now, one can see that all edges of  $G_{Q_i}$  are dominated. Furthermore, with the exception of the pair  $\{c_1, c_2\}$ , every edge of  $G_{Q_i}$  is separated from every other edge of  $G_{\mathcal{Q}}$ . Next, in each variable gadget  $G_{x_j}$ , if  $x_j$  is true, edges  $t_j^1$  and  $t_j^2$  belong to  $C$ . Otherwise, edges  $\bar{t}_j^1$  and  $\bar{t}_j^2$  belong to  $C$ . Edges  $f_1, f_2, f_3, f_4$  and  $f_5$  also belong to  $C$ . Because of this choice, all edges of  $G_{x_j} \setminus \{t_j^1, t_j^2, \bar{t}_j^1\}$  are dominated. Since each of the three edges  $t_j^1, t_j^2, \bar{t}_j^1$  is incident to a vertex of a  $P$ -gadget of some clause gadget, they are also dominated. Moreover, all pairs of edges containing at least one edge of  $G_{x_j} \setminus \{t_j^1, t_j^2, \bar{t}_j^1\}$  are clearly separated. Now, since for each  $P$ -gadget of the clause gadgets, edge  $a$  is in  $C$ ,  $t_j^1, t_j^2, \bar{t}_j^1$  are separated from all edges in  $G_{\mathcal{Q}}$ . It remains to prove that  $C$  separates the pair  $\{c_1, c_2\}$  of each clause gadget. Since we are considering a satisfying assignment of  $\mathcal{Q}$ , in every clause  $Q_i$  of  $\mathcal{Q}$ , there exists a true literal. Hence, for each clause  $Q_i$ , at least one edge  $a_{i_j}$  with  $1 \leq j \leq 3$ , must be in the code and, therefore, the pair  $\{c_1, c_2\}$  is separated. We conclude that  $C$  is an edge-identifying code of size  $k$ .

For the necessary side, let  $C'$  be an edge-identifying code of  $G_{\mathcal{Q}}$  with  $|C'| \leq k$ . It follows from Claim 1 that three edges of each of the seven  $P$ -gadgets of a clause gadget  $G_{Q_i}$  must belong to  $C'$ . Moreover, by Claim 2, edge  $c_0$  is forced to be in any code. Finally, by Claim 2, for each vertex  $l_{i_k}$  ( $1 \leq k \leq 3$ ) of  $G_{Q_i}$ , one of the edges  $a_{i_k}$  and  $b_{i_k}$  is in  $C'$ .

Note that this is a total of at least twenty-five edges per clause gadget.

Similarly, it follows from Claim 1 that in each variable gadget  $G_{x_j}$ , at least fifteen edges of  $C'$  are contained in the  $P$ -gadgets of  $G_{x_j}$ . Following Claim 2, all edges  $f_i$  ( $1 \leq i \leq 5$ ) belong to  $C'$ . Note that this is a total of at least twenty edges in each variable gadget. We have considered  $25m + 20n$  edges of  $C'$  so far. Hence  $2n$  edges remain to be considered. It follows from Claim 3 that for each variable gadget, at least two additional edges belong to  $C'$  (in order to separate the pairs  $\{d_i, e_i\}$ , for  $1 \leq i \leq 4$ ). Therefore, in each variable gadget, exactly two of these edges belong to  $C'$ . Hence, following the second part of Claim 3, either  $\{t_j^1, t_j^2\}$  or  $\{\bar{t}_j^1, \bar{t}_j^2\}$  is a subset of  $C'$ .

Remark that we have now considered all  $k = 25m + 22n$  edges of  $C'$ . Therefore, in each clause gadget  $G_{Q_i}$ , exactly one of the edges  $a_{i_k}$  and  $b_{i_k}$  of  $G_{Q_i}$  belongs to  $C'$ .

We can now build the following truth assignment: for each variable gadget, if  $\{t_j^1, t_j^2\}$  is a subset of  $C'$ ,  $x_j$  is set to TRUE. Otherwise,  $\{\bar{t}_j^1, \bar{t}_j^2\}$  is a subset of  $C'$  and  $x_j$  is set to FALSE. Let us prove that this assignment satisfies  $\mathcal{Q}$ .

In each clause gadget  $G_{Q_i}$ , note that edges  $c_1$  and  $c_2$  must be separated by  $C'$ ; this means that one edge  $a_{i_k}$  from  $\{a_{i_1}, a_{i_2}, a_{i_3}\}$  belongs to  $C'$ . Hence, as noted in the previous paragraph,  $b_{i_k} \notin C'$  and by Claim 4, in the path formed by edges  $\{a_{i_k}, b_{i_k}, t_j^1\}$ ,  $t_j^1$  belongs to the code (without loss of generality, we suppose that  $l_{i_k} = x_j$  and  $t_j^1$  is the edge of  $G_{x_j}$  incident to vertex  $l_{i_k}$  of  $G_{Q_i}$ ). Therefore, in the constructed truth

assignment, literal  $l_{i_k}$  has value TRUE, and the clause is satisfied. Repeating this argument for each clause shows that the formula is satisfied.

Now, it remains to show that similar arguments can be used to prove the final statement of the theorem for larger girth. Consider some integers  $\lambda \geq 1$  and  $\mu \geq 2$ . We build the graph  $G_{\mathcal{Q}}(\lambda, \mu)$  using modified variable gadgets  $G_{x_j}(\mu)$  and modified clause gadgets  $G_{Q_i}(\lambda)$ , which are depicted in Figure 9. The construction is the same as in the previous proof and  $G_{\mathcal{Q}}(\lambda, \mu)$  has  $(36\lambda + 9)m + (30\mu - 18)n$  vertices. We claim that the girth of  $G_{\mathcal{Q}}(\lambda, \mu)$  is now at least  $\min\{4\mu, 8(\lambda + 1)\}$ . Indeed,  $G_{x_j}(\mu)$  has a cycle of size exactly  $4\mu$  and since the girth of  $B(\mathcal{Q})$  is at least 4, it follows that the minimum length of a cycle between some clause gadgets (at least two) and some variable gadgets (at least two) is at least  $4(2\lambda + 1) + 2 + 2 = 8(\lambda + 1)$ .

Now, using a similar proof as the proof for girth 8, it can be shown that  $\mathcal{Q}$  is satisfiable if and only if  $G_{\mathcal{Q}}(\lambda, \mu)$  has an identifying code of size at most  $k = (21\lambda + 4)m + (17\mu - 12)n$ .  $\square$

It is known that a line graph  $\mathcal{L}(G)$  is perfect if and only if  $G$  has no odd cycles of length more than 3, see [20]. Moreover, one can check that the line graphs of the graphs constructed in the previous proof are planar, have maximum degree 4 and clique number 3. Therefore, the following corollary follows:

**Corollary 21.** *IDCODE is NP-complete even when restricted to perfect 3-colorable planar line graphs of maximum degree 4.*

In the following, by slightly restricting the class of graphs considered in Theorem 20, we show that EDGE-IDCODE becomes linear-time solvable in this restricted class.

Let us first introduce some necessary concepts.

A graph property  $\mathcal{P}$  is expressible in *counting monadic second-order logic*, CMSOL for short (see [7] for further reference), if  $\mathcal{P}$  can be defined using:

- vertices, edges, sets of vertices and sets of edges of a graph
- the binary adjacency relation  $adj$  where  $adj(u, v)$  holds if and only if  $u, v$  are two adjacent vertices
- the binary incidence relation  $inc$ , where  $inc(v, e)$  holds if and only if edge  $e$  is incident to vertex  $v$
- the equality operator  $=$  for vertices and edges
- the membership relation  $\in$ , to check whether an element belongs to a set
- the unary cardinality operator  $card$  for sets of vertices
- the logical operators OR, AND, NOT (denoted by  $\vee, \wedge, \neg$ )
- the logical quantifiers  $\exists$  and  $\forall$  over vertices, edges, sets of vertices or sets of edges

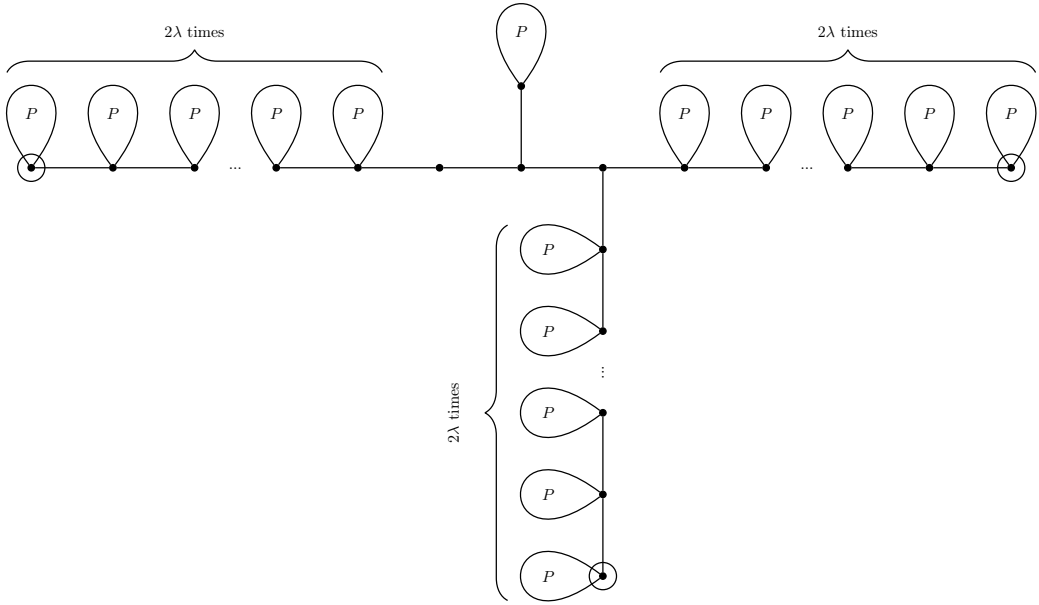
It has been shown that CMSOL is particularly useful when combined with the concept of the graph parameter *tree-width* (we refer the reader to [7] for a definition). Some important classes of graphs have bounded tree-width. For example, trees have tree-width 1, series-parallel graphs have tree-width 2 and outerplanar graphs have tree-width 3.

The following result shows that many graph properties can be checked in linear time for graphs of bounded tree-width.

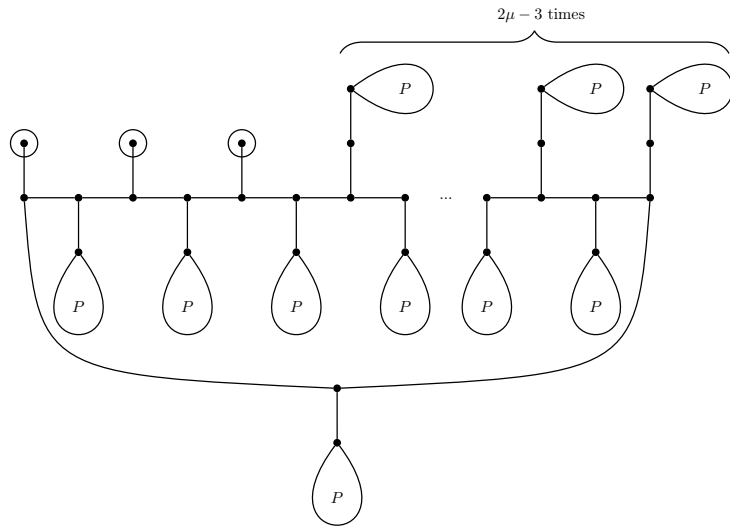
**Theorem 22** (Courcelle [7]). *Let  $\mathcal{P}$  be a graph property expressible in CMSOL and let  $c$  be a constant. Then, for any graph  $G$  of tree-width at most  $c$ , it can be checked in linear time whether  $G$  has property  $\mathcal{P}$ .*

We now show that CMSOL can be used in the context of edge-identifying codes:

**Proposition 23.** *Given a graph  $G$  and an integer  $k$ , let  $\mathcal{EID}(G, k)$  be the property that  $\gamma^{\text{EID}}(G) \leq k$ . Property  $\mathcal{EID}(G, k)$  can be expressed in CMSOL.*



(a) Clause gadget  $G_{Q_i}(\lambda)$



(b) Variable gadget  $G_{x_j}(\mu)$

Figure 9: Reduction gadgets for clause  $Q_i$  and variable  $x_j$  for arbitrarily large girth.

*Proof.* Let  $V = V(G)$  and  $E = E(G)$ . We define the CMSOL relation  $dom(e, f)$  which holds if and only if  $e, f$  are edges of  $E$  and  $e, f$  dominate each other, i.e.  $e$  and  $f$  are incident to the same vertex. We have  $dom(e, f) := \exists x \in V, (inc(x, e) \wedge inc(x, f))$ .



Now we define  $\mathcal{EID}(G, k)$  as follows:

$$\mathcal{EID}(G, k) := \exists C, C \subseteq E, \text{card}(C) \leq k, (\forall e \in E, \exists f \in C, \text{dom}(e, f)) \wedge \\ \left( \forall e \in E, \forall f \in E, e \neq f, \exists g \in C, ((\text{dom}(e, g) \wedge \neg \text{dom}(f, g)) \vee (\text{dom}(f, g) \wedge \neg \text{dom}(e, g))) \right)$$

□

This together with Theorem 22 implies the following corollary.

**Corollary 24.** *EDGE-IDCODE can be solved in linear time for all classes of graphs having their tree-width bounded by a constant.*

This result implies, in particular, that one can find the edge-identifying code number of a tree in linear time. Note that a similar approach has been used for the case of vertex-identifying codes in [17]. The proof of Theorem 22 is constructive and gives a linear-time algorithm, but it is very technical. Therefore, it would be interesting to give a simpler linear-time algorithm for EDGE-IDCODE in trees. Observe that this has been done in [1] for the case of vertex-identifying codes.

## References

- [1] D. Auger. Minimal identifying codes in trees and planar graphs with large girth, *European Journal of Combinatorics* 31(5):1372–1384, 2010.
- [2] L. W. Beineke. Characterizations of derived graphs, *Journal of Combinatorial Theory* 9(2):129–135, 1970.
- [3] N. Bertrand, I. Charon, O. Hudry and A. Lobstein. Identifying and locating-dominating codes on chains and cycles, *European Journal of Combinatorics* 25(7), 969–987, 2004.
- [4] I. Charon, G. Cohen, O. Hudry and A. Lobstein. New identifying codes in the binary Hamming space, *European Journal of Combinatorics* 31(2):491–501, 2010.
- [5] I. Charon, O. Hudry and A. Lobstein. Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard. *Theoretical Computer Science* 290(3):2109–2120, 2003.
- [6] I. Charon, O. Hudry and A. Lobstein. Extremal cardinalities for identifying and locating-dominating codes in graphs. *Discrete Mathematics* 307(3-5):356–366, 2007.
- [7] B. Courcelle. The monadic second-order logic of graphs. I. Recognizable sets of finite graphs. *Information and Computation* 85(1):12–75, 1990.
- [8] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour and M. Yannakakis. The complexity of multiterminal cuts. *SIAM Journal of Computing* 23(4):864–894, 1994.
- [9] F. Foucaud, E. Guerrini, M. Kovše, R. Naserasr, A. Parreau and P. Valicov. Extremal graphs for the identifying code problem. *European Journal of Combinatorics* 32(4):628–638, 2011.
- [10] F. Foucaud, R. Klasing, A. Kosowski and A. Raspaud. On the size of identifying codes in triangle-free graphs. Submitted, 2010.
- [11] F. Foucaud and G. Perarnau. Bounds on identifying codes in terms of degree parameters. Submitted, 2011.
- [12] S. Gravier, R. Klasing and J. Moncel. Hardness results and approximation algorithms for identifying codes and locating-dominating codes in graphs. *Algorithmic Operations Research* 3(1):43–50, 2008.
- [13] S. Gravier and J. Moncel. On graphs having a  $V \setminus \{x\}$  set as an identifying code. *Discrete Mathematics* 307(3-5):432–434, 2007.
- [14] T. W. Haynes, D. J. Knisley, E. Seier and Y. Zou. A quantitative analysis of secondary RNA structure using domination based parameters on trees. *BMC Bioinformatics* 7:108, 2006.
- [15] M. G. Karpovsky, K. Chakrabarty and L. B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Transactions on Information Theory* 44:599–611, 1998.
- [16] M. Laifenfeld, A. Trachtenberg and T. Y. Berger-Wolf. Identifying codes and the set cover problem. Proceedings of the 44th Annual Allerton Conference on Communication, Control and Computing, Monticello, USA, September 2006.
- [17] J. Moncel. *Codes Identifiants dans les Graphes*. PhD Thesis, Université Joseph-Fourier - Grenoble I, France, June 2005. Available online at <http://tel.archives-ouvertes.fr/tel-00010293>.
- [18] J. Moncel. On graphs on  $n$  vertices having an identifying code of cardinality  $\log_2(n + 1)$ . *Discrete Applied Mathematics* 154(14):2032–2039, 2006.
- [19] J. Suomela. Approximability of identifying codes and locating-dominating codes. *Information Processing Letters* 103(1):28–33, 2007.
- [20] M. E. Trotter. Line perfect graphs. *Mathematical Programming* 12(1):255–259, 1977.