Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 223 (2023) 157–165

Procedia Computer Science

www.elsevier.com/locate/procedia

XII Latin-American Algorithms, Graphs and Optimization Symposium (LAGOS 2023)

# Identifying codes in bipartite graphs of given maximum degree☆

Dipayan Chakraborty[a,*], Florent Foucaud[a], Tuomo Lehtilä[b]

[a] *Université Clermont Auvergne, CNRS, Mines Saint-Étienne, Clermont Auvergne INP, LIMOS, 63000 Clermont-Ferrand, France*
[b] *University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland*

**Abstract**

An *identifying code* of a closed-twin-free graph $G$ is a set $S$ of vertices of $G$ such that any two vertices in $G$ have a distinct intersection between their closed neighborhoods and $S$. It was conjectured in [F. Foucaud, R. Klasing, A. Kosowski, A. Raspaud. On the size of identifying codes in triangle-free graphs. Discrete Applied Mathematics, 2012] that there exists an absolute constant $c$ such that for every connected graph $G$ of order $n$ and maximum degree $\Delta$, $G$ admits an identifying code of size at most $\frac{\Delta-1}{\Delta}n + c$. We provide significant support for this conjecture by proving it for the class of all bipartite graphs that do not contain any pairs of open-twins of degree at least 2. In particular, this class of bipartite graphs contains all trees and more generally, all bipartite graphs without 4-cycles. Moreover, our proof allows us to precisely determine the constant $c$ for the considered class, and the list of graphs needing $c > 0$. For $\Delta = 2$ (the graph is a path or a cycle), it is long known that $c = 3/2$ suffices. For connected graphs in the considered graph class, for each $\Delta \geq 3$, we show that $c = 1/\Delta \leq 1/3$ suffices and that $c$ is required to be positive only for a finite number of trees. In particular, for $\Delta = 3$, there are 12 trees with diameter at most 6 with a positive constant $c$ and, for each $\Delta \geq 4$, the only tree with positive constant $c$ is the $\Delta$-star. Our proof is based on induction and utilizes recent results from [F. Foucaud, T. Lehtilä. Revisiting and improving upper bounds for identifying codes. SIAM Journal on Discrete Mathematics, 2022].



*Keywords:* identifying codes; maximum degree; trees; bipartite graphs

## 1. Introduction

Let $G = (V(G), E(G))$ be a graph with vertex-set $V(G)$ and edge-set $E(G)$. Any subset $S \subset V(G)$ is called a *vertex subset* of $G$. The *(open) neighborhood* of a vertex $v$ of $G$ is the set $N_G(v)$ of all vertices of $G$ adjacent to $v$. The vertices of $G$ in $N_G(v)$ are also called the *(first) neighbors* of $v$ in $G$. Moreover, the set $N_G[v] = \{v\} \cup N_G(v)$ is called the *closed*

☆ The two first authors were supported by a public grant overseen by the French National Research Agency as part of the "Investissements d'Avenir" through the IMobS3 Laboratory of Excellence (ANR-10-LABX-0016), the IDEX-ISITE initiative CAP 20-25 (ANR-16-IDEX-0001), and the ANR project GRALMECO (ANR-21-CE48-0004). The third author was supported by the Academy of Finland grant 338797.

* Corresponding author. Tel.: +33 07 45 62 74 68.
*E-mail address:* dipayan.chakraborty@uca.fr


10.1016/j.procs.2023.08.225

*neighborhood* of $v$. Vertices $u, v \in V(G)$ are called open (closed) twins in $G$ if and only if they have the same open (closed) neighborhood. Graphs with no open or closed twins are called *twin-free*.

An *identifying code* $C$ of a graph $G$ is a vertex subset of $G$ that (i) *dominates* each vertex $v$ of $G$ (that is, either $v \in C$ or that $v$ has a neighbor in $C$) and (ii) *separates* each pair $u, v$ of distinct vertices of $G$ (that is, there is a vertex of $G$ in $C$ that belongs to exactly one of the two closed neighborhoods $N[u]$, $N[v]$). Note that graph $G$ admits an identifying code only when no two vertices of $G$ are closed twins. Hence, we say that graphs with no closed twins are *identifiable*, that is, they admit an identifying code (for example, the whole vertex set). A vertex subset of a graph $G$ satisfying the property (i) is called a *dominating set* of $G$ and subset satisfying property (ii) is called a *separating set* of $G$. It is natural to ask for a minimum-sized identifying code of an identifiable graph $G$. The size, denoted by $\gamma^{ID}(G)$, of such a minimum-sized identifying code of $G$ is called the *identification number* (or *ID-number* for short) of $G$.

Identifying codes were introduced in 1998 [1], motivated by fault-detection in multiprocessor networks. Since then, numerous other applications of identifying codes have been discovered, such as threat location in facilities using sensor networks [2], logical definability of graphs [3] or canonical labeling of graphs for the graph isomorphism problem [4]. Besides, since the 1960s and long before the introduction of identifying codes of graphs, many related concepts such as *separating systems* or *test covers* have been independently studied. All of them put together form the general area of identification problems in graphs and other discrete structures, see for example [5, 6, 7] and bibliography [8] for over 500 papers on the topic.

A natural question that arises in the study of identifying codes is the one of extremal values for the identification number: how large can it be, with respect to some relevant graph parameters? When only the order $n$ of the graph is considered, it is known that the identification number of an identifiable graph with at least one edge lies between $\log_2(n+1)$ [1] and $n-1$ [9]; both values are tight for graphs with edges and the extremal examples have been characterized in [10] and [11], respectively.

Nevertheless, it was observed in [11] that when the maximum degree $\Delta$ of the graph $G$ is small enough with respect to the order $n$ of the graph, the $(n-1)$-upper bound can be significantly improved (for connected graphs) to $n - \frac{n}{\Theta(\Delta^5)}$. The latter was thereafter subsequently reduced to $n - \frac{n}{\Theta(\Delta^3)}$ in [12]. This raises the question of what is the largest possible identification number of a connected identifiable graph of order $n$ and maximum degree $\Delta$. Towards this problem, the following conjecture was posed, which is the main topic of this paper.

**Conjecture 1 ([13, Conjecture 1]).** *There exists a constant $c$ such that for every connected identifiable graph on $n \geq 2$ vertices and of maximum degree $\Delta \geq 2$,*

$$\gamma^{ID}(G) \leq \frac{\Delta - 1}{\Delta} n + c.$$

Note that for $\Delta \leq 1$, the only connected identifiable graph is the one-vertex graph. From the known results in the literature [14] and [15], the above conjecture holds for $\Delta = 2$ (that is, for paths and cycles) with $c = 3/2$. The following lemma recapitulates the results for the case that $\Delta = 2$.

**Lemma 2 ([14, 15]).** *Let $G$ be an identifiable graph on $n$ vertices and of maximum degree $\Delta = 2$. Then:*

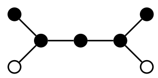
(1) *For all paths $G$: $\gamma^{ID}(G) = \lfloor \frac{n}{2} \rfloor + 1 \leq \frac{\Delta-1}{\Delta} n + 1$;*
(2) *For $G$ isomorphic to either a 4-cycle or a 5-cycle: $\gamma^{ID}(G) = \lfloor \frac{n}{2} \rfloor + 1 \leq \frac{\Delta-1}{\Delta} n + 1$;*
(3) *For all even cycles $G$ of order $n \geq 6$: $\gamma^{ID}(G) = \frac{n}{2} = \frac{\Delta-1}{\Delta} n$;*
(4) *For all odd cycles $G$ of order $n \geq 7$: $\gamma^{ID}(G) = \frac{n}{2} + \frac{3}{2} = \frac{\Delta-1}{\Delta} n + \frac{3}{2}$.*

Hence, for the rest of this extended abstract, we assume that $\Delta \geq 3$. If true, Conjecture 1 would be tight: for any $\Delta \geq 3$, there are arbitrarily large graphs of order $n$ and maximum degree $\Delta$ with identification number $\frac{\Delta-1}{\Delta} n$ [12]. A bound of the form $n - \frac{n}{103(\Delta+1)^3}$ [12] proved using probabilistic arguments is the best known general result towards Conjecture 1 (for the sake of comparison, the conjectured bound can be rewritten as $n - \frac{n}{\Delta} + c$). For triangle-free graphs, this was improved to $n - \frac{n}{\Delta + o(\Delta)}$ in [13], and to smaller bounds for subclasses of triangle-free graphs, such as $n - \frac{n}{\Delta+9}$ for bipartite graphs and $n - \frac{n}{3\Delta/(\ln \Delta - 1)}$ for triangle-free graphs without (open) twins. The latter result implies that Conjecture 1 holds for triangle-free graphs without any open twins, whenever $\Delta \geq 55$ (because then, $3\Delta/(\ln\Delta - 1) \leq \Delta$). It holds for bipartite graphs without (open) twins by [16]. Until this work, the conjecture remained open even for trees, and one of the challenges for proving it on trees was to allow open-twins of degree 1 (note that for any set of mutual open-twins, one needs all of them but one in any identifying code).

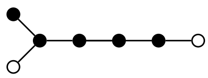
**Our work.** We investigate Conjecture 1 for bipartite graphs of maximum degree $\Delta \geq 3$ and we prove it (with $c = \frac{1}{\Delta} \leq \frac{1}{3}$) for a large subclass of connected bipartite graphs with no twins of degree 2 or greater. In particular, this subclass contains connected bipartite graphs with no 4-cycles and hence, all trees as well. Moreover, for each $\Delta \geq 3$, we characterize the graphs of this class with maximum degree $\Delta$ for which $c > 0$. It is given by the collection $\mathcal{F}_\Delta$, whereby

$$\mathcal{F}_\Delta = \{K_{1,3}\} \cup \mathcal{T}_{\text{tree}} \text{ for } \Delta = 3; \text{ and } \mathcal{F}_\Delta = \{K_{1,\Delta}\} \text{ for } \Delta \geq 4,$$

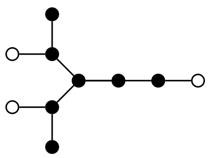
where $\mathcal{T}_{\text{tree}}$ is a set of 11 trees of maximum degree 3 and diameter at most 6. See Figure 1 for the full list of the 11 trees in $\mathcal{T}_{\text{tree}}$. Here, $K_{1,\Delta}$ is a star on $\Delta + 1$ vertices. Note that, for maximum degree at least 3, all bipartite graphs are identifiable. Hence, throughout the rest of the paper, we tacitly assume all our graphs (of the considered graph class) to be identifiable. Our main results are stated as follows.



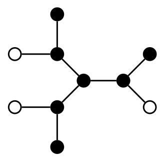
(a) $T_1$: $\gamma^{ID}(T_1) = \frac{14}{3} + \frac{1}{3} = 5$.



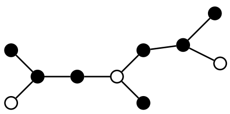
(b) $T_2$: $\gamma^{ID}(T_2) = \frac{14}{3} + \frac{1}{3} = 5$.



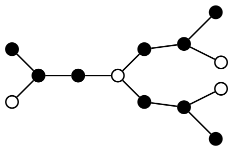
(c) $T_3$: $\gamma^{ID}(T_3) = \frac{20}{3} + \frac{1}{3} = 7$.



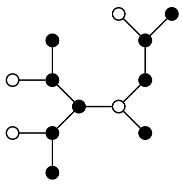
(d) $T_4$: $\gamma^{ID}(T_4) = \frac{20}{3} + \frac{1}{3} = 7$.



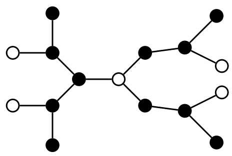
(e) $T_5$: $\gamma^{ID}(T_5) = \frac{20}{3} + \frac{1}{3} = 7$.



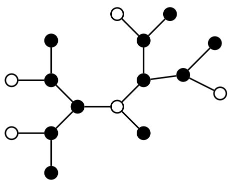
(f) $T_6$: $\gamma^{ID}(T_6) = \frac{26}{3} + \frac{1}{3} = 9$.



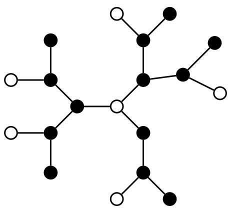
(g) $T_7$: $\gamma^{ID}(T_7) = \frac{26}{3} + \frac{1}{3} = 9$.



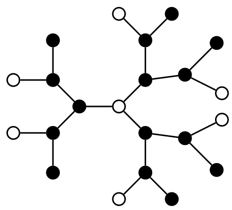
(h) $T_8$: $\gamma^{ID}(T_8) = \frac{32}{3} + \frac{1}{3} = 11$.



(i) $T_9$: $\gamma^{ID}(T_9) = \frac{32}{3} + \frac{1}{3} = 11$.



(j) $T_{10}$: $\gamma^{ID}(T_{10}) = \frac{38}{3} + \frac{1}{3} = 13$.



(k) $T_{11}$: $\gamma^{ID}(T_{11}) = \frac{44}{3} + \frac{1}{3} = 15$.

Fig. 1. The family $\mathcal{T}_{\text{tree}}$ of trees. The set of black vertices in each figure constitutes an identifying code of the tree.

In the following proposition we present the exact ID-number for every connected identifiable bipartite graph $G$ on at least $n \geq 3$ without twins of degree at least two which have $\gamma^{ID}(G) > \frac{\Delta-1}{\Delta} n$.

**Proposition 3.** *Let $G$ be a graph of order $n$, maximum degree $\Delta \geq 3$ and isomorphic to a graph in $\mathcal{F}_\Delta$. Then, we have*

$$\gamma^{ID}(G) = \frac{\Delta - 1}{\Delta} n + \frac{1}{\Delta}.$$

Our next (main) result proves the conjecture for all other connected bipartite graphs of maximum degree $\Delta \geq 3$ and not in the collection $\mathcal{F}_\Delta$.

**Theorem 4.** *Let $G$ be a connected bipartite graph of order $n$, of maximum degree $\Delta \geq 3$, with no twins of degree 2 or greater, and not isomorphic to any graph in the collection $\mathcal{F}_\Delta$. Then, we have*

$$\gamma^{ID}(G) \leq \frac{\Delta - 1}{\Delta} n.$$

We will see that Theorem 4 is tight for many connected bipartite graphs without twins of degree at least 2. When $\Delta = 3$ there are infinitely many examples for such graphs. Furthermore, we give in Proposition 17 an infinite family of graphs for any $\Delta \geq 4$ which has identification number quite close to the conjectured bound; as $\Delta$ increases, our construction gets closer and closer to the conjectured bound.

**Definitions and notations.** For any vertex $v$ of $G$, the symbol $\deg_G(v)$ denoting the *degree* of the vertex $v$ *in* $G$ is the total number of neighbors of $v$ in $G$. A *leaf* of a graph $G$ is a vertex of degree 1 in $G$. The (only) neighbor of a leaf $v$ in a graph $G$ is called the *support vertex* of $v$ in $G$. Any vertex of a graph $G$ that is not a leaf of $G$ is usually referred to as a *non-leaf* vertex of $G$. The length (or the number of edges) of a longest induced path in a graph $G$ is called the *diameter* of $G$. On many occasions throughout this article, we shall have the need to look at a subgraph of a graph $G$ formed by deleting away some vertices or edges from $G$. To that end, given a graph $G$ and a set $S$ containing some vertices and edges of $G$, we define $G - S$ as the subgraph of $G$ obtained by deleting from $G$ all vertices (and edges incident with them) and edges of $G$ in $S$.

**Structure of the paper.** Following the current Introduction, Section 2 is entirely dedicated to the proof of Theorem 4. Section 3 deals with the tightness of the conjectured bound and we conclude in Section 4.

## 2. Proof of main result

Throughout this section, unless otherwise mentioned, all (connected bipartite) graphs we consider are identifiable. We prove Proposition 3 and Theorem 4 here. Toward proving Proposition 3, we next look at the *star* graphs. For any $\Delta \geq 3$, the complete bipartite graph $K_{1,\Delta}$ is called a $\Delta$-*star*, or simply a *star*. Noting that for any $\Delta$-star $S$ the set of all its leaves constitutes a minimum identifying code of $S$, it can therefore be readily verified that Proposition 3 is true for all $\Delta$-stars.

**Lemma 5.** *For a $\Delta$-star $S$ with $\Delta \geq 3$ and on $n$ $(= \Delta + 1)$ vertices, we have $\gamma^{ID}(S) = \frac{\Delta-1}{\Delta} n + \frac{1}{\Delta}$.*

In particular, Lemma 5 shows that $\Delta$-stars satisfy the conjectured bound with $c = \frac{1}{\Delta}$. To fully establish Proposition 3 now, one needs to only show the veracity of the result for the rest of the graphs in $\mathcal{F}_3$, that is, the trees in $\mathcal{T}_{\text{tree}}$. To describe the trees in $\mathcal{T}_{\text{tree}}$ and other graphs later in a more unified manner, we start by defining a particular "join" of graphs with stars. Let $G'$ be a graph and $S$ be any star. Then, let $G' \rhd_v S$ denote the graph obtained by identifying a vertex $v$ of $G'$ with a leaf $l$ of $S$ (for example, if $S$ and $P$ are a 3-star and 4-path, respectively, each with a leaf $v$, then the graphs in Figures 1(a) and 1(b) are $S \rhd_v S$ and $P \rhd_v S$, respectively). We call the $G' \rhd_v S$ the *graph $G'$ appended with a star* and it is said to be obtained by *appending $S$ (by its leaf $l$) onto (the vertex $u$ of) $G'$*. In the case that the vertex $v$ of $G'$ is inconsequential to the context or is (up to isomorphism) immaterial to the graph $G$, we may simply drop the suffix $v$ in the notation $G' \rhd_v S$ and denote it as $G' \rhd S$ (for example, if $P$ is a 2-path and $S$ is a star, then

$P \rhd_v S$ is (up to isomorphism) the only graph irrespective of which vertex of the 2-path $v$ is). As a convention, we continue to call the vertices of $G' \rhd S$ by the same names as they were called in the graphs $G'$ and $S$. In other words, the graph $G' \rhd S$ is said to *inherit* its vertices from $G'$ and $S$. In particular, if $G' \rhd S$ is obtained by identifying the vertex $v$ of $G'$ and a leaf $l$ of $S$, both the names $v$ and $l$ (as and when convenient) also refer to the identified vertex in $G' \rhd S$.

For any positive integer $p$, let $[p]$ denote the set $\{1, 2, \ldots, p\}$. Let $G_0$ be a fixed graph, $p \geq 1$ be an integer and for each $i \in [p]$, let $S_i$ be a $\Delta_i$-star for $\Delta_i \geq 3$. Now, we may carry out the process of inductively appending stars by defining $G_i = G_{i-1} \rhd_{v_{i-1}} S_i$ for all $i \in [p]$, where $v_{i-1}$ is a vertex of $G_{i-1}$. Then the graph $G_p$ is called the *graph $G_0$ appended with $p$ stars*. In the case that each $S_i$ is isomorphic to a $\Delta$-star $S$ for $\Delta \geq 3$, we call the graph $G_p$ the *graph $G_0$ appended with $p$ $\Delta$-stars*. In the particular case that $G_0 = S_0$ is itself a $\Delta_0$-star for $\Delta_0 \geq 3$, we simply call $G_p$ an *appended star*. Further, if $\Delta = \Delta_0 = \Delta_1 = \cdots = \Delta_p$, then we call the graph $G_p$ an *appended $\Delta$-star*. We next furnish some general results for any identifiable graph appended with a star.

**Lemma 6.** *Let $G'$ be an identifiable graph and let $G = G' \rhd_v S$, where $v$ is a vertex of $G'$ and $S$ is a $\Delta$-star for $\Delta \geq 3$. Then $G$ is also identifiable and $\gamma^{ID}(G) \leq \gamma^{ID}(G') + \Delta - 1$.*

The next lemma shows that if $G$ is a graph obtained by starting from an identifiable "base" graph $G_0$ and iteratively appending stars thereon, then the graphs $G$ and $G_0$ share the same constant $c$ in Conjecture 1. It is worth mentioning here that Lemma 7 is central to our inductive proof arguments later whereby, if a graph $G$ is structurally a "smaller" identifiable graph $G'$ appended with a star, then by using Lemma 7 and an inductive hypothesis that the "smaller" graph $G' = G - S$ satisfies the conjectured bound, one can show that so does the "bigger" graph $G$.

**Lemma 7.** *Let $c$ be a constant, $G_0$ be an identifiable graph on $n_0$ vertices, of maximum degree $\Delta_0$ and be such that $\gamma^{ID}(G_0) \leq \frac{\Delta_0 - 1}{\Delta_0} n_0 + c$. For an integer $p \geq 1$ and for all $i \in [p]$, let $S_i$ be a $\Delta_i$-star for $\Delta_i \geq 3$. Also, for all $i \in [p]$, let $G_i = G_{i-1} \rhd_{v_{i-1}} S_i$, where $v_{i-1}$ is a vertex of $G_{i-1}$, and let $G = G_p$ be the graph $G_0$ appended with $p$ stars on $n$ vertices and of maximum degree $\Delta$. Moreover, assume that $\Delta_{max} = \max\{\Delta_i : 0 \leq i \leq p\}$. Then, we have*

$$\gamma^{ID}(G) \leq \frac{\Delta_{max} - 1}{\Delta_{max}} n + c \leq \frac{\Delta - 1}{\Delta} n + c.$$

**Corollary 8.** *For an integer $p \geq 1$ and for all $i \in [p]$, let $S_i$ be a $\Delta_i$-star for $\Delta_i \geq 3$. For each $i \in [p]$, let $G_i = G_{i-1} \rhd_{v_{i-1}} S_i$, where $v_{i-1}$ is a vertex of $G_{i-1}$, and let $G = G_p$ be an appended star on $n$ vertices and of maximum degree $\Delta$. Moreover, assume that $\Delta_{max} = \max\{\Delta_i : 0 \leq i \leq p\}$. Then, we have*

$$\gamma^{ID}(G) \leq \frac{\Delta_{max} - 1}{\Delta_{max}} n + \frac{1}{\Delta_0} \leq \frac{\Delta - 1}{\Delta} n + \frac{1}{3}.$$

We next look at the identification numbers of the trees $T_2$ and $T_3$ in the collection $\mathcal{T}_{\text{tree}}$.

**Lemma 9.** *Let $T_2$ and $T_3$ be the trees as in $\mathcal{T}_{\text{tree}}$ on $n$ vertices and of maximum degree $\Delta = 3$. Then, $\gamma^{ID}(T_2) = 5 = \frac{2}{3} \times 7 + \frac{1}{3} = \frac{\Delta - 1}{\Delta} n + \frac{1}{3}$ and $\gamma^{ID}(T_3) = 7 = \frac{2}{3} \times 10 + \frac{1}{3} = \frac{\Delta - 1}{\Delta} n + \frac{1}{3}$.*

Lemma 9 therefore establishes the result in Proposition 3 for the trees $T_2$ and $T_3$ in $\mathcal{T}_{\text{tree}}$. Next, we look at other paths appended with stars which are *not* isomorphic to either the graph $T_2$ or $T_3$. In particular, we show that all paths appended with stars other than $T_2$ and $T_3$ satisfy Conjecture 1 with constant $c = 0$.

**Lemma 10.** *Let $G = P \rhd_v S$ be a graph on $n$ vertices, where $S$ is a $\Delta$-star with $\Delta \geq 3$, $P$ is a path and $v$ is a vertex of $P$. Moreover, let $G$ be of any of the following types: either (1) $P$ is not a 4-path; or (2) $P$ is a 4-path and $v$ is a non-leaf vertex of $P$; or (3) $P$ is a 4-path, $v$ is a leaf of $P$ and $\Delta \geq 4$. Then, in each of the cases, we have $\gamma^{ID}(G) \leq \frac{\Delta - 1}{\Delta} n$.*

Lemma 10 shows that all trees of the form $P \rhd S$, where $P$ is a 4-path and $S$ is a star, but not isomorphic to $T_2$ satisfy the bound in Conjecture 1 with $c = 0$. Our next lemma shows that all trees of the form $T_i \rhd S$, for $i \in \{2, 3\}$, where $S$ is a star, but not isomorphic to $T_3$ also satisfy the bound in Conjecture 1 with $c = 0$.

**Lemma 11.** *Let $G = T_i \rhd_v S$, for $i \in \{2, 3\}$, be a graph on $n$ vertices and of maximum degree $\Delta \geq 3$ such that $G \not\cong T_3$, where, $v$ is a vertex of $T_i$ and $S$ is a $\Delta_S$-star with $\Delta_S \geq 3$. Then, we have $\gamma^{ID}(G) \leq \frac{\Delta-1}{\Delta}n$.*

We next turn to the trees in the collection $\mathcal{T}_{\text{tree}}$ other than $T_2$ and $T_3$ and show that they satisfy Proposition 3. These are precisely the appended 3-stars of maximum degree 3 and of diameter at most 6. One characteristic of appended $\Delta$-stars is that they all have an even diameter. Using that, the following proposition unifies the identification number for all such appended 3-stars in $\mathcal{T}_{\text{tree}}$.

**Proposition 12.** *Let $G$ be an appended 3-star on $n$ vertices, of maximum degree $\Delta = 3$ and of diameter at most 6. Then, we have $\gamma^{ID}(G) = \frac{2}{3}n + \frac{1}{3}$.*

By Lemmas 5, 9 and Proposition 12 therefore, we have the *proof of Proposition 3*. In the next two lemmas, however, we show that all other appended stars not in the collection $\mathcal{T}_{\text{tree}}$ satisfy the conjectured bound with $c = 0$.

**Lemma 13.** *Let $G$ be an appended 3-star on $n$ vertices, of maximum degree $\Delta = 3$ and of diameter at least 8. Then, we have $\gamma^{ID}(G) \leq \frac{2}{3}n = \frac{\Delta-1}{\Delta}n$.*

**Lemma 14.** *Let $G$ be an appended star on $n$ vertices and of maximum degree $\Delta \geq 4$. Then, we have $\gamma^{ID}(G) \leq \frac{\Delta-1}{\Delta}n$.*

We next focus on the proof of Theorem 4. Before we start the proof sketch of Theorem 4, we cite two lemmas from [16]. These two lemmas allow us to show that in order to prove Theorem 4, one needs to consider the connected bipartite graphs $G$ only of the form $G' \rhd_v S$, whereby $G'$ is necessarily connected and bipartite as well. Thus, as we see in the proof sketch, induction plays a central role in proving Theorem 4.

**Lemma 15 ([16, Lemma 4]).** *Let $G$ be a connected bipartite graph on $n \geq 4$ vertices, with $s$ support vertices and not isomorphic to a 4-path. Then, we have $\gamma^{ID}(G) \leq n - s$.*

**Lemma 16 ([16, Theorem 6]).** *Let $G$ be a connected bipartite graph on $n \geq 3$ vertices, with $\ell$ leaves and with no twins of degree 2 or greater. Then, we have $\gamma^{ID}(G) \leq \frac{n+\ell}{2}$.*

Our restrictions to the graph class in Theorem 4 are due to the restrictions in Lemma 16. With that, we are now ready to provide the proof sketch of our main theorem.

*Sketch of Proof (Theorem 4).* The proof is by induction on the 2-tuple $(n, m)$ ordered by dictionary order denoted by $<_d$, say, where $m$ is the number of edges of the graph $G$. Since we have $\Delta \geq 3$, this implies that $n \geq 4$. However, $n = 4$ implies that $G$ is a 3-star and thus is isomorphic to a graph in $\mathcal{F}_3$. Therefore, we take the base case of the induction hypothesis to be when $(n, m) = (5, 4)$ (note that for $n = 5$, by the connectivity of $G$, the latter has at least 4 edges). In the base case now, one can check that $G \cong P \rhd S$, where $P$ is a 2-path and $S$ is a 3-star. Therefore, by Lemma 10(1), the result is true in the base case. Let us assume that the induction hypothesis is true for all connected bipartite graphs $G'$ on $n'$ vertices and $m'$ edges such that $(5, 4) \leq_d (n', m') <_d (n, m)$, of maximum degree $\Delta' \geq 3$, not isomorphic to a graph in $\mathcal{F}_{\Delta'}$ and with no twins of degree 2 or greater.

Let $\ell$ and $s$ be the number of leaves and support vertices, respectively, in $G$. If $s \geq \frac{n}{\Delta}$, then, by Lemma 15, we have $\gamma^{ID}(G) \leq n - s \leq n - \frac{1}{\Delta}n = \frac{\Delta-1}{\Delta}n$ and, hence, we are done. Moreover, if $\ell \leq \frac{\Delta-2}{\Delta}n$, then, again, we have $\gamma^{ID}(G) \leq \frac{n+\ell}{2} \leq \left(1 + \frac{\Delta-2}{\Delta}\right)\frac{n}{2} = \frac{\Delta-1}{\Delta}n$ and we are done in this case too. We therefore assume that both $s < \frac{n}{\Delta}$ and that $\ell > \frac{\Delta-2}{\Delta}n$. The latter inequality implies that there is at least one leaf and, hence, at least one support vertex as well in $G$. In this case, we have $\frac{\ell}{s} > \Delta - 2$. Moreover, as $G$ is not a star, the maximum number of leaves adjacent to a support vertex is $\Delta - 1$. Hence, there exists a support vertex which is adjacent to exactly $\Delta - 1$ leaves. So, let $u$ be one such support vertex in $G$ with exactly $\Delta - 1$ leaves adjacent to it. This implies that $G = G' \rhd_x S$, where $S$ is a $\Delta$-star with $u$ as its universal vertex and $G'$ is necessarily a connected bipartite graph. Let $n' = |V(G')|$. To begin with, by the use of Lemmas 10 and 14, one can check that, for $n' \leq 4$, the desired result holds for $G$. We therefore assume that $n' \geq 5$.

Moreover, since $G'$ is connected, it must have at least 4 edges and, thus, we must have $(5,4) \leq_d (n', m')$. Again, by Lemmas 10, 11, 13 and 14, it can be verified that if $G'$ is isomorphic to a graph in $\mathcal{F}_{\Delta'}$, then the desired result holds. We therefore assume that $G'$ is not isomorphic to any graph in $\mathcal{F}_{\Delta'}$. Now, if $G'$ has no twins of degree 2 or greater, then by the induction hypothesis, $\gamma^{ID}(G') \leq \frac{\Delta-1}{\Delta}n' = \frac{\Delta-1}{\Delta}(n-\Delta)$. Thus, by Lemma 7, the result holds. Hence, for the rest of this proof, we assume that $G'$ has vertices $x$ and $y$, say, which are twins of degree 2 or greater in $G'$. Recall that $x$ is adjacent to $u$ in $G$. Moreover, $y$ is not. We also notice that $x, y$ is a unique pair of vertices which are twins of degree at least 2 in $G'$. Observe that $x$ and $y$ must have a common neighbor $v$ of degree at least 3 in $G$ (or else the vertices in the common neighborhood of $x$ and $y$ in $G'$ are twins of degree at least 2 in $G$, a contradiction). We then look at the graph $G''$ obtained from $G$ by removing all edges of the type $xw$, where $w$ is a neighbor of $x$ and $w \neq u, v$. Let $n''$, $m''$ and $\Delta''$ be the number of vertices, number of edges and the maximum degree, respectively, of $G''$. Then, we have $(5,4) \leq_d (n'', m'') <_d (n, m)$. By some structural analysis of the graph $G''$, we are able to show next that neither $G''$ contains any twins of degree 2 or greater nor $G''$ is isomorphic to any graph in $\mathcal{F}_{\Delta''}$. Therefore, by the induction hypothesis, we have $\gamma^{ID}(G'') \leq \frac{\Delta-1}{\Delta}n$ (notice that $n = n''$).

Let $C''$ be a minimum identifying code of $G''$. If $C''$ is an identifying code of $G$ too, then we are done. So, in what follows, let us assume that $C''$ is not an identifying code of $G$. Therefore, there exists a pair $p, q$ of vertices of $G$ that are not separated by $C''$ in $G$. This must be because adding back an edge of type $xw$, where $w \neq u, v$ is a neighbor of $x$ in $G$, causes the vertices $p$ and $q$ to have the same neighborhoods (in $G$) in $C''$. This is possible if and only if at least one of the following three cases occurs.

(1) $p = v$ and $q = w$, without loss of generality, and $x \in C''$; or
(2) $p = x$, without loss of generality, $q \neq w$, $qw \in E(G'')$ and $w \in C''$; or
(3) $p = x$ and $q = w$, without loss of generality.

However, one can verify that Case (3) does not occur. Hence, we only analyze the first two possibilities in the above list (it can be verified that they cannot occur simultaneously). In Case (1), we are able to show by some case analysis that it is always possible to construct from $C''$ an identifying code $C$ of $G$ such that $|C| = |C''|$. Thus, we are done in Case (1). In analyzing Case (2), we conclude that it is enough to look at the graph for maximum degree $\Delta = 3$. In this case, if $w$ is of degree 2 in $G$, then again we are able to construct from $C''$ an identifying code $C$ of $G$ such that $|C| = |C''|$. So, assuming that $\deg_G(w) = 3$, we look at another graph $G^\star = G - \{a, b, y\}$, where $a$ and $b$ are the leaves $G$ inherits from $S$ with the common support vertex $u$ of $G$. It is clear that $G^\star$ has no twins of degree 2 or greater. We also show further that $G^\star$ is not isomorphic to any graph in $\mathcal{F}_3$ (3 being the maximum degree of $G^\star$). Thus, by the induction hypothesis, we have $\gamma^{ID}(G^\star) \leq \frac{2}{3}(n-3)$. So, let $C^\star$ be a minimum size identifying code of $G^\star$. Then, again, by a case analysis, we are able to construct from $C^\star$ an identifying code $C$ of $G$ such that $|C| = |C^\star| + 2$ and thus, $\gamma^{ID}(G) \leq |C| \leq \frac{2}{3}n$ and the result holds. □

## 3. Extremal examples

We now consider some classes of bipartite graphs (without twins of degree 2 or greater) for which Conjecture 1 is tight. With tightness we mean that graph $G$ on $n$ vertices has $\gamma^{ID}(G) \geq \frac{\Delta-1}{\Delta}n$. Clearly, Conjecture 1 is tight for every graph in $\mathcal{F}_\Delta$. Moreover, it is tight for *double stars* (that is, $S_1 \rhd_u S_2$, where $S_1$ is a $(\Delta-1)$-star, $S_2$ is a $\Delta$-star and $u$ is the universal vertex of $S_1$) with $2\Delta - 2$ leaves. Another class of graphs of maximum degree $\Delta = 3$ for which the conjecture is tight is the 3-*corona* of a path $P$: we obtain the 3-corona of a graph $G$ by identifying each vertex $v$ of $G$ with a leaf of a 3-path $P_v$ (see [16]). Moreover, when both $n$ and $\Delta$ are large, there are graphs which almost attain the conjectured bound. This is shown next, by noticing that for large $\Delta$, the value $\frac{\Delta-1+\frac{1}{\Delta-2}}{\Delta+\frac{2}{\Delta-2}}$ is close to $\frac{\Delta-1}{\Delta}$.

**Proposition 17.** *Let $C_{2t}$ be a cycle on $2t \geq 6$ vertices. Let an $n$-vertex intermediate graph be formed by appending onto every vertex of the cycle $\Delta - 2 \geq 2$ copies of $\Delta$-stars. Thereafter, for each vertex $c_i$ of the cycle, subdivide a single edge between $c_i$ and an adjacent support vertex of the intermediate graph. Let the final graph be called $G_{t,\Delta}$*

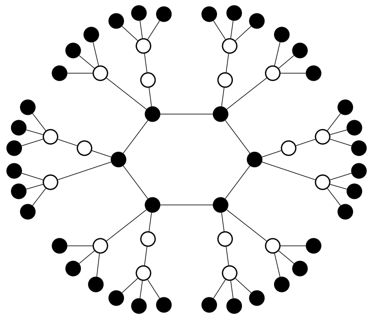


Fig. 2. Graph $G_{t,\Delta}$ as in Proposition 17 with $t = 3$ and $\Delta = 4$. The set of black vertices constitutes an identifying code of $G_{t,\Delta}$.

*(see Figure 2 for an example with $t = 3$ and $\Delta = 4$). Then, we have*

$$\gamma^{ID}(G_{t,\Delta}) = \frac{\Delta - 1 + \frac{1}{\Delta-2}}{\Delta + \frac{2}{\Delta-2}} n.$$

## 4. Conclusion

We have made significant progress towards Conjecture 1 by proving it for a large class of bipartite graphs: those with no twins of degree at least 2, including all trees and all bipartite graphs with no 4-cycles.

A possible next step could be to prove the conjecture for all bipartite graphs, or perhaps even, for all triangle-free graphs. The much weaker bound of $n - \frac{n}{3\Delta/(\ln \Delta - 1)}$ was proved for triangle-free graphs in [13]. Even progress on the (sub)cubic case would be interesting. The currently best known upper bounds for triangle-free subcubic graphs and cubic graphs are $\frac{8n}{9}$ and $\frac{5n}{6}$, respectively (see [17, Corollary 4.46]; the proof uses the technique developed in [13]).

## References

[1] M. G. Karpovsky, K. Chakrabarty, L. B. Levitin, On a new class of codes for identifying vertices in graphs, IEEE Transactions on Information Theory 44 (2) (1998) 599–611.

[2] R. Ungrangsi, A. Trachtenberg, D. Starobinski, An implementation of indoor location detection systems based on identifying codes, in: F. A. Aagesen, C. Anutariya, V. Wuwongse (Eds.), Intelligence in Communication Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 175–189.

[3] O. Pikhurko, H. Veith, O. Verbitsky, The first order definability of graphs: Upper bounds for quantifier depth, Discrete Applied Mathematics 154 (17) (2006) 2511–2529.

[4] L. Babai, On the complexity of canonical labeling of strongly regular graphs, SIAM Journal on Computing 9 (1) (1980) 212–216.

[5] J. A. Bondy, Induced subsets, Journal of Combinatorial Theory, Series B 12 (2) (1972) 201–202.

[6] B. M. E. Moret, H. D. Shapiro, On minimizing a set of tests, SIAM Journal on Scientific and Statistical Computing 6 (4) (1985) 983–1003.

[7] A. Rényi, On random generating elements of a finite boolean algebra, Acta Scientiarum Mathematicarum Szeged 22 (1961) 75–81.

[8] D. Jean, A. Lobstein, Watching systems, identifying, locating-dominating and discriminating codes in graphs, `https://dragazo.github.io/bibdom/main.pdf`.

[9] S. Gravier, J. Moncel, On graphs having a V\{x} set as an identifying code, Discrete Mathematics 307 (3-5) (2007) 432–434.

[10] J. Moncel, On graphs on n vertices having an identifying code of cardinality $\lceil \log_2(n+1) \rceil$, Discrete Applied Mathematics 154 (14) (2006) 2032–2039.

[11] F. Foucaud, E. Guerrini, M. Kovše, R. Naserasr, A. Parreau, P. Valicov, Extremal graphs for the identifying code problem, European Journal of Combinatorics 32 (4) (2011) 628–638.

[12] F. Foucaud, G. Perarnau, Bounds for identifying codes in terms of degree parameters, The Electronic Journal of Combinatorics 19 (1) (2012) 32.

[13] F. Foucaud, R. Klasing, A. Kosowski, A. Raspaud, On the size of identifying codes in triangle-free graphs, Discrete Applied Mathematics 160 (10-11) (2012) 1532–1546.
[14] N. Bertrand, I. Charon, O. Hudry, A. Lobstein, Identifying and locating-dominating codes on chains and cycles, European Journal of Combinatorics 25 (7) (2004) 969–987.
[15] S. Gravier, J. Moncel, A. Semri, Identifying codes of cycles, European Journal of Combinatorics 27 (5) (2006) 767–776.
[16] F. Foucaud, T. Lehtilä, Revisiting and improving upper bounds for identifying codes, SIAM Journal on Discrete Mathematics 36 (4) (2022) 2619 – 2634.
[17] F. Foucaud, Algorithmic and combinatorial aspects of identifying codes in graphs, Ph.D. thesis, Université Bordeaux 1 (2012).