On the size of identifying codes in triangle-free graphs

Florent Foucaud\textsuperscript{a,\ast}, Ralf Klasing\textsuperscript{a}, Adrian Kosowski\textsuperscript{a,b}, André Raspaud\textsuperscript{a}

\textsuperscript{a}LaBRI - Université de Bordeaux - CNRS - INRIA, 351 cours de la Libération, 33405 Talence cedex, France.
\textsuperscript{b}Department of Algorithms and System Modeling, Gdańsk University of Technology, Narutowicza 11/12, 80952 Gdańsk, Poland.

Abstract

In an undirected graph $G$, a subset $C \subseteq V(G)$ such that $C$ is a dominating set of $G$, and each vertex in $V(G)$ is dominated by a distinct subset of vertices from $C$, is called an identifying code of $G$. The concept of identifying codes was introduced by Karpovsky, Chakrabarty and Levitin in 1998. For a given identifiable graph $G$, let $\gamma^{ID}(G)$ be the minimum cardinality of an identifying code in $G$. In this paper, we show that for any connected identifiable triangle-free graph $G$ on $n$ vertices having maximum degree $\Delta \geq 3$, $\gamma^{ID}(G) \leq n - \frac{n}{\Delta + o(\Delta)}$. This bound is asymptotically tight up to constants due to various classes of graphs including $(\Delta - 1)$-ary trees, which are known to have their minimum identifying code of size $n - \frac{n}{\Delta - 1 + o(1)}$. We conjecture that the bound $\gamma^{ID}(G) \leq n - \frac{n}{\Delta} + O(1)$ holds for any nontrivial connected identifiable graph $G$.

Keywords: Identifying code, Dominating set, Triangle-free graph, Maximum degree

1. Introduction

Identifying codes, which have been introduced in [19], are dominating sets having the additional property that each vertex of the graph can be uniquely identified using its neighbourhood within the identifying code. They have found numerous applications, such as fault-diagnosis in multiprocessor networks [19], the placement of networked fire detectors in complexes of rooms and corridors [23], compact routing [20], or the analysis of secondary RNA structures [17]. Identifying codes are a variation on the earlier concept of locating-dominating sets (cf. e.g. [9, 25, 26]). Identifying codes have been studied in specific graph classes such as cycles [3, 16], trees [4, 6], grids [19] or hypercubes [18, 22]. Extremal problems regarding the minimum size of an identifying code have been studied in [8, 10, 11, 12, 15, 21].

Herein, we further investigate these extremal questions by giving new upper bounds on the size of minimum identifying codes for triangle-free graphs using their maximum degree.

1.1. Notations and definitions

Let $G = (V, E)$ be a simple undirected graph. We denote the vertex set of $G$ by $V = V(G)$ and its edge set, by $E = E(G)$. We also denote by $n = |V|$ the order of $G$ and by $\Delta = \Delta(G)$, the maximum vertex degree of $G$.

For a vertex $v$ of $G$, the ball $B(v)$ is the set of all vertices of $V$ which are at distance at most 1 from $v$. We denote by $N(v) = B(v) \setminus \{v\}$, the neighbourhood of $v$. For a set $X$ of vertices of $G$, we define $N(X)$ to be the union of the neighbourhoods of all vertices of $X$, that is $N(X) = \cup_{x \in X} N(x)$. Whenever we find it necessary to emphasize on the graph $G$ for which the neighbourhood is considered, we write $B_G(u)$, $N_G(u)$.

\begin{itemize}
\item \textsuperscript{\ast}This research is supported by the ANR Project IDEA - Identifying coDes in Evolving grAphs, ANR-08-EMER-007, 2009-2011 and by the KBN Grant 4 T11C 047 25.
\item \textsuperscript{\ast}Corresponding author. E-mail: foucaud@labri.fr - Telephone: +33540003517 - Fax: +33540006669
\end{itemize}

Preprint submitted to Elsevier June 12, 2011
and $N_G(X)$. Two distinct vertices $u, v$ are called twins if $B(u) = B(v)$ [7]. They are called false twins if $N(u) = N(v)$ but $u$ and $v$ are not adjacent [5].

For a subset $S$ of vertices of $G$, we denote by $G[S]$ the subgraph of $G$ induced by $S$. A matching $M$ of a graph $G$ is a subset of edges of $G$ such that no two edges of $M$ have a common vertex. If within the set of all endpoints of the edges of $M$ no other edges than the ones of $M$ exist, we call $M$ an induced matching.

Given a set $S$ of vertices of $G$, we say that a vertex $x$ of $G$ is $S$-isolated if $x \in S$ and no neighbour of $x$ belongs to $S$. We say that vertex $u$ dominates vertex $v$ if $v \in B(u)$. For two subsets $C,U$ of vertices, $C$ dominates $U$ if each vertex of $U$ is dominated by some vertex of $C$. Set $C \subseteq V$ is called a dominating set of $G$ if $C$ dominates $V$. The vertices of a pair $u,v$ of vertices of $V$ are separated by some vertex $x \in V$ if $x$ dominates exactly one of the vertices $u$ and $v$. We call $C \subseteq V$ an identifying code of $G$ if it is a dominating set of $G$, and for all pairs $u,v$ of vertices of $V$, $u$ and $v$ are separated by some vertex of $C$. The latter condition can be equivalently stated as $B(u) \cap C \neq B(v) \cap C$, or as $(B(u) \oplus B(v)) \cap C \neq \emptyset$ (denoting by $\oplus$ the symmetric difference of sets). In the following, we might simply call an identifying code a code and a vertex of the code, a code vertex. Given a graph $G$ and a subset $S$ of its vertices, we say that a set $C \subseteq S$ is an $S$-identifying code of $G$ if $C$ is an identifying code of $G[S]$.

A graph is said to be identifiable if it admits an identifying code. This is the case if and only if it does not contain any pair of twins [19]. An example of a graph which is not identifiable is the complete graph $K_n$. For an identifiable graph $G$, we denote by $\gamma^{ID}(G)$ the cardinality of a minimum identifying code of $G$. The problem of determining the exact value of $\gamma^{ID}(G)$ is known to be an NP-hard problem, even when $G$ belongs to the class of planar graphs of maximum degree $4$ having arbitrarily large girth [1], or to the class of planar graphs of maximum degree $3$ and girth $9$ [2].

1.2. Main conjecture and motivation

This paper deals with the study of parameter $\gamma^{ID}$ and its relation with the order and the maximum degree of graphs. This work is an extension of earlier results.

For any graph $G$ on $n$ vertices, the lower bound $\gamma^{ID}(G) \geq \lceil \log_2 \{n+1\} \rceil$ was given in [19]. This bound is tight, and all graphs reaching it have been described in [21]. In [19], it was also shown that the improved bound $\gamma^{ID}(G) \geq \frac{\log n}{\log 3}$ holds, and all graphs reaching this bound have been described in [10]. This bound shows that the maximum degree has a strong influence on the minimum possible value of $\gamma^{ID}$.

Considering upper bounds in terms of $n$ and $\Delta$, we conjecture that the following result holds.

**Conjecture 1.** Let $G$ be a nontrivial connected identifiable graph of maximum degree $\Delta$. Then $\gamma^{ID}(G) \leq n - \frac{n}{\Delta-1} + o(1)$.

It is known that there exist examples of specific families of graphs such that $\gamma^{ID}(G) = n - \frac{n}{\Delta}$ (e.g., the complete bipartite graph $K_{\Delta,n-\Delta}$, Sierpiński graphs [14] and other classes of graphs described in the first author’s master thesis [10]). Other classes of graphs with slightly smaller values of parameter $\gamma^{ID}$ are known, including graphs having high girth. For instance, it is shown in [4] that $\gamma^{ID}(T) = \left\lceil n - \frac{n}{\Delta-1+1/\Delta} \right\rceil$ for any complete $(\Delta-1)$-ary tree $T$ on $n$ vertices.

For all identifiable graphs having at least one edge, the upper bound $\gamma^{ID}(G) \leq n - 1$ holds [8, 15]. This bound is tight, in particular for the star $K_{1,n-1}$ and other graphs which have been recently classified in [11]. Hence, for graphs of very high maximum degree (say $\Delta = n - 1$), the conjecture holds since $n - 1 = n - \frac{n}{\Delta} + \frac{1}{n-1}$. Moreover, for any connected graph $G$ of maximum degree $2$ (i.e. when $G$ is either a path or a cycle), the exact value of $\gamma^{ID}(G)$ is known (see [3, 16]). In this case, the bound $\gamma^{ID}(G) \leq \frac{n}{2} + \frac{n}{2} = n - \frac{n}{2} + \frac{n}{2}$ holds and is reached for infinitely many values of $n$ (more precisely, this is the case when $G$ is a cycle of odd order $n \geq 7$). Hence, the conjecture holds for $\Delta = 2$.

It was shown in [11] that for any connected identifiable graph $G$ of maximum degree $\Delta$, $\gamma^{ID}(G) \leq n - \frac{n}{2(\Delta-1)}$, and if $G$ is $\Delta$-regular, $\gamma^{ID}(G) \leq n - \frac{n}{3(\Delta-1)}$. In this paper, we improve these results by showing that the conjectured bound holds asymptotically when $G$ is triangle-free. More precisely, it is proved in Theorem 13 that $\gamma^{ID}(G) \leq n - \frac{n^{2}}{\Delta+1+1/\Delta}$ when $G$ is a nontrivial connected identifiable triangle-free graph. This result strongly supports Conjecture 1. Moreover, the proof is constructive and can be used to build the corresponding code in polynomial time.
1.3. Organization of the paper

The paper is structured as follows. In Section 2.1, we begin by giving an informal overview of the technique and the construction used to prove our results. In Sections 2.2 to 2.4, we introduce some definitions and preliminary results that are needed in the proof of our main result. This result is proved in Section 2.5. Finally, Section 3 concludes the paper with some remarks.

2. The upper bound

2.1. Proof ideas

The general idea of our proof technique is to construct a sufficiently large independent set of the graph such that some specific conditions hold. Taking the complement of this set and performing some local modifications yields an identifying code. This technique originates from the following proposition, which we give in order to give the reader a first intuition of our technique. The proof is given in Subsection 2.2.

**Proposition 2.** Let $G$ be an identifiable (not necessarily connected) triangle-free graph, and $S$, an independent set of $G$. Then, if the following conditions hold, $V(G) \setminus S$ is an identifying code of $G$.

1. $S$ contains no isolated vertex of $G$.
2. For any pair $u, v$ of vertices of $S$, $N(u) \neq N(v)$ (i.e. $S$ does not contain any pair of false twins).
3. For each vertex $v$ of degree 1 in $G$, some vertex at distance 2 from $v$ does not belong to $S$.
4. The graph $G[V(G) \setminus S]$ has no isolated edges.

In order to prove our main result, we show how to build (large enough) independent sets in triangle-free graphs such that the three first conditions of Proposition 2 hold (see Lemma 10). However, it seems difficult to also ensure that the last condition holds while keeping the size of $S$ reasonably large. Therefore, after building $S$, we compute the set $M$ of isolated edges of $G[V(G) \setminus S]$ and partition $V(G)$ into the end-vertices of $M$ (set $R$) together with their neighbours (set $L$) on the one hand, and the rest of the vertices, $V \setminus (L \cup R)$, on the other hand. We then build a sufficiently small $(L, R)$-quasi-identifying code $C_1$, a variation of an identifying code which will be defined later (see Definition 6). This construction is done in Lemmas 11 and 12. Setting $C_2$ as the complement of $S$ restricted to $V \setminus (L \cup R)$, our final code is $C_1 \cup C_2$. We also combine this method with another technique (Proposition 3) which is suitable for the special case where the graph has a large number of false twins. The whole procedure is sketched in Algorithm 1.

**Algorithm 1** Construction of an identifying code

| Input: a nontrivial connected identifiable triangle-free graph $G = (V, E)$ |
| 1: Compute the set $X$ of vertices having at least one false twin |
| 2: if $X$ is “small” then |
| 3: Use Lemma 10 to compute an independent set $S$ of $G$ fulfilling the three first properties listed in Proposition 2. |
| 4: Compute the set $R \subseteq V$ of vertices such that for each $v \in R$, $v$ has a neighbour $u$ where both $u$ and $v$ are of degree at least 2, and all the vertices of $N(u) \cup N(v) \setminus \{u, v\}$ belong to $S$. |
| 5: $L \leftarrow N(R) \setminus R$ |
| 6: Compute an $(L, R)$-quasi-identifying code $C_1$ of $G$ using the constructions of Lemmas 11 and 12. |
| 7: $C_2 \leftarrow (V \setminus (L \cup R)) \setminus S$ |
| 8: $C \leftarrow C_1 \cup C_2$ |
| else (X is “big”) |
| 10: $C \leftarrow$ an identifying code of $G$ computed using Proposition 3. |
| 11: end if |
| 12: return $C$ |

This process is detailed in Subsection 2.5 (Theorem 13). All auxiliary results needed for this proof are developed in the next subsections.

...
2.2. Preliminary results

The next proposition shows how to build an identifying code of a graph \( G \) which has relatively small size when \( G \) contains a large number of false twins. But first, we let \( \equiv \) denote the false twin relation over \( V(G) \), where \( u \equiv v \) if \( u, v \) are false twins. This relation is an equivalence relation. We call an equivalence class of \( \equiv \) nontrivial if it has at least two elements.

**Proposition 3.** Let \( G \) be a nontrivial connected identifiable triangle-free graph on \( n \) vertices and maximum degree \( \Delta \), non isomorphic to \( C_4 \). Let \( \mathcal{F} = \{F_1, \ldots, F_{|\mathcal{F}|}\} \) be the set of all nontrivial equivalence classes over \( \equiv \) in \( G \). Then \( G \) has an identifying code of size at most \( n - |\mathcal{F}| \).

**Proof.** First, we may suppose that \( G \) is not isomorphic to \( P_3 \) since, in that case, the lemma holds: \( P_3 \) has its minimum identifying code of size 2 and \( |\mathcal{F}| = 1 \).

For each \( F_i \in \mathcal{F}, 1 \leq i \leq |\mathcal{F}| \), let \( x_i \) be an arbitrary vertex of \( F_i \), and let \( X = \cup_{i=1}^{|\mathcal{F}|} x_i \). We claim that if \( G \) is not isomorphic to \( P_3 \) or \( C_4 \), \( C = V(G) \setminus X \) is an identifying code of \( G \). First, observe that \( C \) is a dominating set of \( G \). Now, consider two vertices \( x, y \) and suppose, for the sake of contradiction, that they are not separated by \( C \). This means that \( B(x) \cup B(y) \subseteq X \).

If \( x \) and \( y \) are not adjacent, then \( x \) and \( y \) both belong to \( X \) (otherwise, they are separated). Hence they are not false twins, and one of them (say \( y \)) has a neighbour \( z \) which is not adjacent to \( x \). If \( z \not\in X \) and \( x \) and \( y \) are separated by \( z \) and we are done. Otherwise, \( z \) has a false twin \( z' \) and \( z' \not\in X \). Then \( z' \) is adjacent to \( x \) but not to \( y \), hence \( z' \) separates \( x \) and \( y \), a contradiction.

If \( x \) and \( y \) are adjacent, since \( G \) is identifiable, again one of them (say \( y \)) is adjacent to a vertex \( z \) which is not adjacent to \( x \). Since \( x \) and \( y \) are not separated \( z \in X \). Hence \( z \) has a false twin \( z' \) not adjacent to \( x \), but to \( y \), and \( z' \not\in X \). Hence \( z' \not= x \) and \( y \) are separated by \( z' \). Now suppose \( x = z' \). Since \( G \) is not isomorphic to \( P_3 \), \( G \) contains at least one other vertex. Since \( G \) is connected and triangle-free and \( x, z \) are false twins, some new vertex is adjacent to exactly one of \( x \) and \( y \), and since \( x, y \) are not separated, it belongs to \( X \). Suppose \( y \) has such a neighbour, \( t \). Since \( z \in X \) and \( t \) and \( x, z \) are not false twins. Hence \( t \) has a false twin \( t' \) which is not in \( X \) but is adjacent to \( y \). Hence \( x \) and \( y \) are separated by \( t' \), a contradiction. Now, if \( y \) has no other neighbour, \( x \) has a neighbour, \( u \). Then \( x, y, z, u \) induce a 4-cycle. But since \( G \) is connected and \( G \) is not isomorphic to \( C_4 \), \( G \) contains another vertex, \( v \), and by our assumption \( v \) is not adjacent to \( y \). We have two cases. First, suppose \( v \) is a neighbour of \( x \) and \( z \). Then, \( v \in X \) (otherwise \( v \) separates \( x \) and \( y \)). But since \( u \in X \), \( u \) and \( v \) are not false twins, hence \( v \) must have a false twin \( v' \not\in X \), and \( v' \) separates \( x \) and \( y \), a contradiction. Now, assume \( v \) is a neighbour of \( u \). Since \( v \) is not a neighbour of \( y \), \( u \) and \( y \) are not false twins. But since \( u \in X \), \( u \) must have a false twin \( u' \) and \( u' \not\in X \). Then \( u' \) separates \( x \) and \( y \), a contradiction. \( \square \)

In our main proof, we want to construct an independent set \( S \) having some prescribed properties, consider the set \( V(G) \setminus S \) as a potential code, and modify it in order to identify those vertices which form isolated edges in \( G[V(G) \setminus S] \). The following definition introduces a notion which will help to capture this situation.

**Definition 4.** Given a graph \( G \) together with an induced matching \( M \) of \( G \), we denote by \( R(M) \), the set of end-vertices of the edges of \( M \) and by \( L(M) \), the set of neighbours of the vertices of \( R(M) \): \( L(M) = N(R(M)) \setminus R(M) \). \( M \) is called a strong induced matching if the following holds:

- \( L(M) \) is an independent set in \( G \).
- Each vertex \( x \) of \( R(M) \) has degree at least 2 in \( G \) (i.e. \( N(x) \cap L(M) \neq \emptyset \)).

An illustration of a strong induced matching is given in Figure 1. Note that in some graphs, one cannot necessarily find a strong induced matching. Indeed, if \( G \) is triangle-free, each edge of such a matching must belong to at least some induced path on four vertices.

Note that in any triangle-free graph \( G \) having a strong induced matching \( M \), \( G[L(M) \cup R(M)] \) has no isolated edge (i.e. two adjacent vertices of degree 1). Since in a triangle-free graph, a pair of twins necessarily forms an isolated edge, the following observation is immediate.
Observation 5. Let $G$ be a triangle-free graph $G$ having a strong induced matching $M$. Then $G[L(M) \cup R(M)]$ is identifiable.

In order to construct small identifying codes of a triangle-free graph $G$ having a strong induced matching $M$, we will construct special codes for the subgraph of $G$ induced by set $L(M) \cup R(M)$. These codes are defined as follows.

Definition 6. Let $G$ be a triangle-free identifiable graph having a strong induced matching $M$ with $L = L(M)$ and $R = R(M)$. Let $G' = G[L \cup R]$. We say that $C \subseteq L \cup R$ is an $(L, R)$-quasi-identifying code of $G$ if:

1. Each vertex of $L \cup R$ is dominated by some vertex of $C$.
2. For each pair $u, v$ of vertices in $L \cup R$, $C \cap B_{G'}(u) \neq C \cap B_{G'}(v)$, unless $u$ and $v$ both belong to $L$ and $N_{G'}(u) = N_{G'}(v)$.
3. For each edge $e$ of $M$, at least one of the vertices of $e$ belongs to $C$.

Note that because of condition 2 of Definition 6, an $(L, R)$-quasi-identifying code of $G$ is not necessarily an $(L \cup R)$-identifying code of $G$. Conversely, because of condition 3, an $(L \cup R)$-identifying code of $G$ might not be an $(L, R)$-quasi-identifying code of $G$.

The following proposition shows that we can use an $(L, R)$-quasi-identifying code of $G$ to construct a valid identifying code of $G$.

Proposition 7. Let $G = (V, E)$ be an identifiable triangle-free graph having a strong induced matching $M$, with $L = L(M)$ and $R = R(M)$, and suppose that $L$ does not contain any pair of false twins in $G$. Also suppose that there exists an $(L, R)$-quasi-identifying code $C_1$ of $G$ without $C_1$-isolated vertices and a $(V \setminus (L \cup R))$-identifying code $C_2$ of $G$ where all the neighbours of vertices of $L$ within $V \setminus (L \cup R)$ belong to $C_2$.\footnote{Note that if a $(V \setminus (L \cup R))$-identifying code $C$ exists (i.e. $G[V \setminus (L \cup R)]$ is identifiable), then adding all neighbours of vertices of $L$ to $C$ yields an identifying code. In fact, any superset of an identifying code is still an identifying code.}

Then, $C_1 \cup C_2$ is an identifying code of $G$.

Proof. Let us show that the vertices of all pairs of vertices of $G$ are separated. Since $C_2$ is a $(V \setminus (L \cup R))$-identifying code, all pairs of vertices of $V \setminus (L \cup R)$ are separated. Since $C_1$ is $(L, R)$-quasi-identifying and there are no $C_1$-isolated vertices, each vertex $x$ of $L \cup R$ is dominated by at least one vertex of $R \cap C_1$ (see...
points number 1 and 3 of Definition 6), which we denote \( f_{C_1}(x) \). Moreover, by definition of sets \( L \) and \( R \), no vertex of \( V \setminus (L \cup R) \) is dominated by a vertex of \( R \). Therefore, all pairs of vertices \( u, v \) with \( u \in L \cup R \) and \( v \in \overline{V \setminus (L \cup R)} \) are separated by \( f_{C_1}(x) \). It remains to check the pairs of vertices of \( L \cup R \). By contradiction, suppose there are two vertices \( u, v \) of \( L \cup R \) which are not separated. By point number 2 of Definition 6, \( u \) and \( v \) belong to \( L \) and have the same neighbourhood within \( L \cup R \). But since we assumed that they are not false twins and all their neighbours in \( V \setminus (L \cup R) \) are in \( C_2 \), \( u \) and \( v \) are separated by the neighbours they do not have in common, a contradiction.

We now give a proof of Proposition 2, after recalling its statement.

**Proposition** (Proposition 2). Let \( G \) be an identifiable (not necessarily connected) triangle-free graph, and \( S \), an independent set of \( G \). Then, if the following conditions hold, \( V(G) \setminus S \) is an identifying code of \( G \).

1. \( S \) contains no isolated vertex of \( G \).
2. For any pair \( u, v \) of vertices of \( S \), \( N(u) \neq N(v) \) (i.e. \( S \) does not contain any pair of false twins).
3. For each vertex \( v \) of degree 1 in \( G \), some vertex at distance 2 from \( v \) does not belong to \( S \).
4. The graph \( G[V(G) \setminus S] \) has no isolated edges.

**Proof.** Let \( C = V(G) \setminus S \). Since \( S \) is an independent set and does not contain any isolated vertex, each vertex having at least one neighbour within \( G[V(G) \setminus (L \cup R)] \) is dominated. Thus, \( C \) is a dominating set. Let us now check the separation condition. Let \( u, v \) be an arbitrary pair of vertices of \( V(G) \). We distinguish several cases.

If \( u \) and \( v \) are adjacent and both have degree at least 2, suppose by contradiction that they are not separated. This means that all their neighbours are in \( S \). But then they form an isolated edge in \( G[C] \), a contradiction.

Now, suppose \( u \) and \( v \) are adjacent and one of them, say \( u \), has degree 1. Since \( G \) is identifiable, \( v \) has at least one neighbour and by the first property of \( S \), there is a vertex at distance 2 of \( u \) in \( C \), separating \( u \) and \( v \).

If \( u \) and \( v \) are false twins, they do not both belong to \( S \) and hence they are separated by themselves.

Finally, if \( u \) and \( v \) are not adjacent and are not false twins of each other, if either \( u \) or \( v \) belong to \( C \), they are separated. If both \( u \) and \( v \) belong to \( S \), all their neighbours belong to \( C \) (none of them can belong to \( S \)), and since they have distinct sets of neighbours they are separated.

In order to use Proposition 2, we need to build (large enough) independent sets in triangle-free graphs. The next subsection is devoted to that task.

### 2.3. Building large independent sets in triangle-free graphs

In this subsection, we use the following result of J. Shearer [24] to show that triangle-free graphs have large independent sets which fulfill some useful conditions. Note that this result is constructive.

**Theorem 8** ([24]). Let \( G \) be a triangle-free graph on \( n \) vertices and average degree \( \overline{d} \). Then \( G \) has an independent set of size at least \( \frac{\overline{d}(\ln \overline{d} - 1) + 1}{(d - 1)^2} n \).

The following corollary of Theorem 8 is an approximate bound which is easier to deal with and which is tight enough for our purposes. It follows from the facts that \( \overline{d}(G) \leq \Delta(G) \) and that when \( x > 1 \), the function \( \frac{x(\ln x - 1) + 1}{(x - 1)^2} \) is decreasing. Moreover in that case, \( \frac{x(\ln x - 1) + 1}{(x - 1)^2} \geq \frac{\ln x}{x - 1} \) and for \( x \geq 3 \), \( \frac{\ln x}{x - 1} > \overline{d} = \Delta \) for every connected triangle-free graph.

**Corollary 9.** Let \( G \) be a triangle-free graph on \( n \) vertices and maximum degree \( \Delta \geq 3 \). Then \( G \) has an independent set of size at least \( \frac{\ln \Delta}{\Delta - 1} n \).

We get the following lemma as a corollary, which we will use in the proof of our main result.

**Lemma 10.** Let \( G \) be an identifiable triangle-free graph on \( n \) vertices and maximum degree \( \Delta \geq 3 \), and let \( Y \) be the set of all vertices of \( G \) having no false twin. Then \( G[Y] \) has an independent set \( S \) with the following properties:
1. For each vertex $u$ of degree 1 in $G$, there exists a vertex of $G$ at distance 2 of $u$ which does not belong to $S$.
2. $|S| \geq \frac{\ln \Delta - 1}{\Delta} |Y|$.

Proof. Let $S_1 \subseteq Y$ be the set of vertices of $Y$ having degree 1 in $G$. Note that since $G$ is identifiable, it has no isolated edges and therefore $S_1$ is an independent set in $G$ and $G[Y]$. Moreover since $Y$ has no vertices having a false twin, all vertices of $S_1$ are at distance at least 3 from each other. Let $T_1$ be the set of vertices constructed as follows. All the vertices of $S_1$ belong to $T_1$. For each element $s$ of $S_1$, its unique neighbour in $G$ belongs to $T_1$, and some arbitrary neighbour at distance 2 of $s$ belongs to $T_1$. Since all the vertices of $S_1$ are at distance at least 3 from each other, for each vertex $s$ of $S_1$ there is a vertex at distance 2 of $s$ belonging to $T_1 \setminus S_1$. We now set $Y_1 = T_1 \cap Y$. Note that we have $|S_1| \geq \frac{|Y_1|}{3} \geq |Y_2|$ since for each vertex of $S_1$, at most three vertices of $G$ have been inserted into $T_1$.

We also set $Y_2 = Y \setminus Y_1$. By the previous construction, $Y_2$ does not contain any vertex of degree 1 in $G$. By Corollary 9, $G[Y_2]$ has an independent set $S_2$ of size at least $\frac{\ln \Delta - 1}{\Delta} |Y_2|$.

Taking $S = S_1 \cup S_2$, we get an independent set of $G[Y]$ fulfilling the first property of the claim. Moreover, $Y_1$ and $Y_2$ form a partition of $Y$, $S_1 \subseteq Y_1$ and $S_2 \subseteq Y_2$. Since for strictly positive $x, \frac{\ln x - 1}{x} > \frac{\ln x - 1}{x}$, we have:

$$|S| \geq \min\{\frac{|Y_1|}{3}, \frac{\ln \Delta - 1}{\Delta} |Y_1|\} \geq \frac{\ln \Delta - 1}{\Delta} |Y|$$

The next subsection is devoted to the construction of small enough quasi-identifying codes.

2.4. Quasi-identifying the vertices in and around a strong induced matching

Recall that in order to prove our main result, given a nontrivial identifiable connected triangle-free graph $G$, we will construct an independent set $S$ and consider the (possibly empty) strong induced matching $M$ such that $R(M)$ forms the set of isolated edges of $V(G) \setminus S$. In order to ensure that there are no isolated edges $uv$ in $G[V(G) \setminus S]$, it would suffice to remove an arbitrary neighbour of either $u$ or $v$ from $S$. However, this could lead to a very large identifying code. Indeed, consider the example of a complete graph $K_n$ where each edge is subdivided twice, $K_n^*$. The original vertices of $K_n$ form a (maximal) independent set $S$ and each original edge of $K_n$ corresponds to an isolated edge in the subgraph of $K_n^*$ induced by the complement of $S$, $K_n[V(K_n^*) \setminus S]$. Now, in $K_n^*$, getting rid of all isolated edges of $K_n[V(K_n^*) \setminus S]$ by removing vertices from $S$ requires a vertex cover of $K_n$, that is, $n - 1$ vertices. This would yield an identifying code of size $|V(K_n^*)| - 1$, which is not interesting.

Hence, in order to overcome this problem, we show in this subsection how to build an $(L(M), R(M))$-quasi-identifying code of bounded size. We first deal with the special case where all vertices of $R(M)$ have degree exactly 2 (note that by Definition 4 they must have degree at least 2).

Lemma 11. Let $G$ be an identifiable (not necessarily connected) triangle-free graph having a strong induced matching $M$ where $L = L(M)$, $R = R(M)$, and all vertices of $R$ have degree exactly 2. Then, there is an $(L, R)$-quasi-identifying code $C$ of $G$ having the following properties:

1. $|C| \leq |L| + \frac{|R|}{2}$.
2. No vertex of $R$ is $C$-isolated.
3. At least half of the vertices of $L$ belong to $C$.

Proof. In order to simplify its construction, let us first define the multigraph $G_{L,R} = (L, E)$ with vertex set $L$ and in which there is an edge between two vertices $l_1$ and $l_2$ if and only if there exist two vertices $r_1, r_2$ of $R$, such that $l_1, r_1, r_2, l_2$ is a 3-path in $G$. In other words, we contract every path of length 3 of $G[L \cup R]$ having both endpoints in $L$, into one edge. There can be multiple edges in $G_{L,R}$ (but no loops), since several disjoint 3-paths may join $l_1$ to $l_2$.

From $G_{L,R}$ we will build an oriented multigraph $\overrightarrow{G}_{L,R}$. Given an orientation of $\overrightarrow{G}_{L,R}$, we define the subset $S(\overrightarrow{G}_{L,R})$ of vertices of $L \cup R$ in the following way: all the vertices of $L$ belong to $S(\overrightarrow{G}_{L,R})$, and...
for each arc $l_1l_2$ of $\overrightarrow{G}_{L,R}$ corresponding to the path $l_1r_1r_2l_2$ in $G$, $r_2$ belongs to $S(\overrightarrow{G}_{L,R})$. Note that $|S(\overrightarrow{G}_{L,R})| = |L| + \frac{|R|}{2}$. An illustration is given in Figure 2, where the gray vertices belong to $S(\overrightarrow{G}_{L,R})$. Our aim is to construct an orientation of $\overrightarrow{G}_{L,R}$ for which $S(\overrightarrow{G}_{L,R})$ is the desired $(L, R)$-quasi-identifying code of $G$.

![Figure 2: Correspondance between a special subset of $L \cup R$ and $\overrightarrow{G}_{L,R}$](image)

We start by orienting the arcs of $\overrightarrow{G}_{L,R}$ in an arbitrary way. Note that $S(\overrightarrow{G}_{L,R})$ fulfills all three required properties of the statement of the lemma. Hence, if $S(\overrightarrow{G}_{L,R})$ is an $(L, R)$-quasi-identifying code of $G$, we are done. So suppose this is not the case. Note that $S(\overrightarrow{G}_{L,R})$ also fulfills conditions number 1 and 3 of Definition 6. Hence, there are pairs of vertices of $L \cup R$ which are not separated by $S(\overrightarrow{G}_{L,R})$. The only case where a pair $l, r$ is not separated by $S(\overrightarrow{G}_{L,R})$, is when $l \in L$, $r \in R$, and both belong to $S(\overrightarrow{G}_{L,R})$, but they are only dominated by each other and themselves. This is equivalent to the case where $l$ is of in-degree 1 in $\overrightarrow{G}_{L,R}$ (see Figure 3 for an illustration). In this case, in order to fix this problem, we modify the orientation of $\overrightarrow{G}_{L,R}$ in possibly two steps.

![Figure 3: Vertices $l$ and $r$ are not separated](image)

Let us describe the first step. Consider a connected component $\overrightarrow{G}_1$ of $\overrightarrow{G}_{L,R}$, and construct an arbitrary spanning tree $\overrightarrow{T}_1$ of $\overrightarrow{G}_1$, rooted in some vertex $l$. Now, go through all vertices of $\overrightarrow{T}_1$, level by level in a bottom-up order from the leaves up to the root. Whenever the in-degree of the current vertex, $v$, is equal to 1, swap the orientation of the arc joining $v$ to its parent in $\overrightarrow{T}_1$. Doing so, the in-degree of $v$ in $\overrightarrow{G}_1$ becomes different from 1, and the in-degree of its parent is either incremented or decremented by 1. Note that except for the root $l$, all vertices of $\overrightarrow{G}_1$ have now an in-degree different from 1. This process is repeated for all connected components of $\overrightarrow{G}_{L,R}$.

Let $C = S(\overrightarrow{G}_{L,R})$ be the new set corresponding to the new orientation. If $C$ is an $(L, R)$-quasi-identifying code of $G$, we are done. Otherwise, as observed earlier, it means that some roots of the spanning trees we built, have in-degree 1 in $\overrightarrow{G}_{L,R}$. Let $l$ be such a root with in-degree 1. Observe that $l$ has a unique neighbour in $C \cap R$, say $r$. Let $r_2$ be the neighbour of $r$ in $R$. It is sufficient to take out $l$ from $C$ and to replace
it by \( r_2 \) in order to separate \( l \) from \( r \) in \( G[L \cup R] \) (see Figure 4 for an illustration), without changing the cardinality of \( C \). Moreover, all neighbours of \( l \) are still separated from the other vertices because they are all in \( R \setminus C \) and therefore have a neighbour in \( R \cap C \), which itself has at least one neighbour in \( L \cap C \). Hence \( C \) is now an \((L, R)\)-quasi-identifying code of \( G \). Since the process did not change the cardinality of \( C \), we get property number 1 of the claim of the lemma.

![Figure 4: Local modification of the constructed code](image)

Notice that there are at most \( \frac{|L|}{2} \) connected components in \( G[L \cup R] \) since each of them contains at least two vertices of \( L \). Thus property number 3 of the claim of the lemma follows.

Property number 2 is fulfilled by the construction of \( C \) since in each pair of adjacent vertices of \( R \), either it has a code vertex in \( L \) as a neighbour if there was no modification done, or in \( R \) if a switch of two elements of \( L \) and \( R \) was necessary. Moreover, for each such pair, at least one of its elements belongs to the code. This shows that \( C \) is an \((L, R)\)-quasi-identifying code and completes the proof.

We now deal with the general case, where the vertices of \( R(M) \) have degree at least 2 as required in Definition 4.

**Lemma 12.** Let \( G \) be an identifiable (not necessarily connected) triangle-free graph having a strong induced matching \( M \), with \( L = L(M) \) and \( R = R(M) \). There exists a set \( L' \) of vertices of \( L \cup R \) such that \( |L'| \geq \frac{|L|}{4} \), and \( C = (L \cup R) \setminus L' \) is an \((L, R)\)-quasi-identifying code of \( G \) having no \( C \)-isolated vertices.

**Proof.** Let us first divide sets \( L \) and \( R \) into the following subsets: let \( R_1 \subseteq R \) be such that \( r \in R_1 \) if both \( r \) and its unique neighbour in \( R \) are of degree 2. Let \( L_1 \subseteq L \) be the set of all neighbours of vertices of \( R_1 \), let \( R_2 = R \setminus R_1 \), and let \( L_2 = L \setminus L_1 \) (see Figure 5 for an illustration).

We can use Lemma 11 to construct an \((L_1, R_1)\)-quasi-identifying code \( C_1 \) of \( G \) such that the three properties described in the statement of Lemma 11 are fulfilled. Let \( C_1 \) be such a code, in particular we have \( |C_1| \leq |L_1| + \frac{|R_1|}{2} \). Let us now describe the construction of two distinct \((L, R)\)-quasi-identifying codes \( C_a \) and \( C_b \).

- **Construction of code \( C_a \).**
  We construct \( C_a \) such that \( |C_a| \leq |L_1| + \frac{|R_1|}{2} + |L_2| + \frac{|R_2|}{2} + \min \left\{ \frac{|L_1|}{2}, \frac{|R_2|}{2} \right\} \), as follows.
  1. Put \( C_1 \) into \( C_a \).
  2. Put \( L_2 \) into \( C_a \).
  3. For each pair \( r, r' \) of adjacent vertices of \( R_2 \), let \( r^* \) be one of them having at least two neighbours in \( L \) (by definition of \( R_2 \) either \( r \) or \( r' \) has this property). Put \( r^* \) into \( C_a \).
  4. For each pair \( r, r' \) of adjacent vertices of \( R_2 \), let \( r^* \) be the one which was put into \( C_a \) in the previous step. Check if \( r^* \) has less than two neighbours within \( C_a \cap L \) (this may happen if some of its neighbours are in \( L_1 \), and they do not belong to \( C_1 \)). If this is the case, pick an additional neighbour of \( r^* \) — which exists since \( r \) has at least two neighbours in \( L \) — and put it into \( C_a \). Note that this is done at most \( \frac{|R_2|}{2} \) times. Moreover, at most \( \frac{|L_1|}{2} \) new vertices from \( L_1 \) are put into \( C_a \) in such a way since by property number 3 of Lemma 11, there are at most \( \frac{|L_1|}{2} \) vertices of \( L_1 \) not in \( C_1 \).
5. Finally, consider each $C_a$-isolated vertex $l$ of $L$, take it out of $C_a$ and put an arbitrary neighbour of $l$ into $C_a$ (this operation does not affect the size of $C_a$).

- **Construction of code $C_b$.**
  We construct $C_b$ such that $|C_b| \leq |L_1| + \frac{|R_1|}{2} + 3\frac{|R_2|}{2}$, as follows.

  1. Put $C_1$ into $C_b$.
  2. Put $R_2$ into $C_b$.
  3. For each pair $r, r'$ of adjacent vertices of $R_2$, one arbitrary neighbour in $L$ of either $r$ or $r'$ is put into $C_b$.
  4. Finally, in the same way as for the construction of $C_a$, we get rid of each $C_b$-isolated vertex $l$ of $L$ by taking $l$ out of $C_b$ and putting an arbitrary neighbour of $l$ into $C_b$ instead.

Let us now prove that $C_a$ and $C_b$ are $(L, R)$-quasi-identifying codes without $C_a$-isolated or $C_b$-isolated vertices. First note that in both constructions, the final step consists in replacing some $C_a$-isolated vertices from $C_a$ (resp. $C_b$). In order to simplify the proof, let $C_a^*$ (resp. $C_b^*$) be the code as it is before this last step. We first prove that $C_a^*$ (resp. $C_b^*$) have all desired properties except that there remain $C_a^*$-isolated (resp. $C_b^*$-isolated) vertices in $L$. We then prove that performing the last step transforms it into an $(L, R)$-quasi-identifying code with all required properties.

It can first be noticed that both $C_a^*$ and $C_b^*$ are dominating sets, so point number 1 of Definition 6 holds.

Let us now show point number 2 of Definition 6 (the separation condition). In both codes, the vertices of all pairs $u, v$ of vertices of $L_1 \cup R_1$ are separated from each other, since $C_1$, which is an $(L_1, R_1)$-quasi-identifying code, is a subset of both $C_a^*$ and $C_b^*$.

Now, suppose that $u \in R_1$ and $v \in L_2 \cup R_2$. By definition of $R_1$, no vertex of $R_1$ is adjacent to any vertex of $L_2 \cup R_2$. Therefore, by condition number 3 of Definition 6, either $u$ or its neighbour in $R_1$ belong to $C_1$, hence $u$ and $v$ are separated.

Thus, it remains to check if $u$ and $v$ are separated when $u \in L_1$ and $v \in L_2 \cup R_2$, and when both $u$ and $v$ belong to $L_2 \cup R_2$. We deal with $C_a^*$ and $C_b^*$ separately.
Suppose \( u \in L_1 \) and \( v \in L_2 \cup R_2 \). Note that \( u \) is dominated by some vertex \( x \) within \( L_1 \cup R_1 \) since \( C_1 \subseteq C'_a \). If \( v \in L_2 \) and \( u \) and \( v \) are separated by \( x \) since no vertex of \( L_2 \) is adjacent to any vertex of \( L_1 \cup R_1 \). If \( v \in R_2 \) and \( v \notin C'_a \), then \( u \) and \( v \) are separated by the neighbour of \( v \) in \( R_2 \), which belongs to \( C'_a \). Similarly, if \( u \) has a neighbour in \( R_1 \) belonging to \( C_1 \), we are done. Otherwise, it means that \( v \in C'_a \) and \( u \in C_1 \) (otherwise \( u \) would not be dominated by \( C_1 \)). Hence \( v \) has another neighbour in \( L \), \( u' \), belonging to \( C'_a \), and \( u' \) separates \( u \) from \( v \). Indeed, at step 4 of the construction of \( C_a \), either \( v \) already had at least two neighbours in \( L \cap C'_a \), or an additional one has been added.

Now, suppose both \( u \) and \( v \) belong to \( L_2 \cup R_2 \).
If both \( u \) and \( v \) belong to \( L_2 \), they are separated since the whole set \( L_2 \) belongs to \( C'_a \) and \( L_2 \) is an independent set.
If both \( u \) and \( v \) belong to \( R_2 \) and they are not adjacent, they are separated since either themselves or their respective neighbours in \( R_2 \) belong to \( C'_a \) by step 3 of its construction. Otherwise, for the same reason one of them (say \( u \)) belongs to the code. It is ensured in step 4 that at least one neighbour of \( u \) in \( L \) belongs to \( C'_a \), therefore \( u \) and \( v \) are separated by this neighbour.
If \( u \in L_2 \) and \( v \in R_2 \) and they are not adjacent, they are separated by \( u \) since the whole set \( L_2 \) belongs to \( C'_a \). Otherwise, if \( v \notin C'_a \), they are separated by the neighbour of \( v \) in \( R_2 \). Otherwise, again by step 4 of the construction \( v \) has a second neighbour in \( L \cap C'_a \), separating them.

Let us now check point number 3 of Definition 6, i.e. that for each pair of adjacent vertices in \( R \), at least one of them belongs to the code. This is true for vertices of \( R_1 \) since \( C_1 \) is an \((L_1, R_1)\)-quasi-identifying code and therefore fulfills this condition. This is also ensured for vertices of \( R_2 \) at step 3 of the construction of \( C_a \) and at step 2 of the construction of \( C_b \).
Hence, we have shown that both \( C'_a \) and \( C'_b \) are \((L, R)\)-quasi-identifying codes.
Moreover, there are no \( C'_a \)-isolated (resp. \( C'_b \)-isolated) vertices in \( R \). There are no such vertices in \( R_1 \) by Lemma 11, and no such vertices in \( R_2 \) for \( C'_a \) by step 4 of its construction, and for \( C'_b \) as well since \( R_2 \subseteq C'_a \).
As announced previously, we now have to deal with the last step of the constructions of both \( C_a \) and \( C_b \). Note that this step does not affect the domination property of both codes. Indeed, on the one hand the isolated vertices themselves are now dominated by some neighbour. On the other hand, each neighbour of a \( C_a \)-isolated or \( C_b \)-isolated vertex belongs to \( R \), and since \( C_a \) and \( C_b \) are \((L, R)\)-quasi-identifying the neighbour of \( u \) in \( R \) belongs to the code.
Let us prove that the separation condition is still satisfied by \( C_a \) and \( C_b \). Let \( C_x \ (x \in \{a, b\}) \) be the considered code and let \( l \in L \) be a \( C_x \)-isolated vertex which gets replaced in \( C_x \) by one of its neighbours. The only vertices which might be affected by the modification, are vertices which were previously dominated by \( l \), i.e. vertices of \( B(l) \).
First, by contradiction suppose some neighbour \( r \in R \) of \( l \) is not separated anymore from some other vertex, \( s \). Since \( r \) does not belong to the code, its neighbour, \( r' \), does. If \( s = r' \), observe that for \( C_x = C_a \) and \( C_x = C_b \), \( C_x \) has no \( C_x \)-isolated vertices in \( R \). Hence \( r' \) has a neighbour in \( C_x \cap L \) and \( r \) and \( r' \) are separated, a contradiction. If \( s \) is a neighbour of \( r \) in \( L \), \( r \) and \( s \) are separated by \( r' \), a contradiction. Finally, suppose \( r \) and \( s \) are not adjacent. Then \( s \) must be a neighbour of \( r' \) in \( L \), and both \( r \) and \( s \) are only dominated by \( r' \). Hence \( r \notin R_1 \), since otherwise \( r' \) would have degree 2 and hence, would be \( C_x \)-isolated, a contradiction. Hence \( r \in R_2 \). Moreover, then \( s \notin L_1 \), because otherwise \( s \) would be dominated within \( L_1 \cup R_1 \) since \( C_1 \subseteq C_x \). Hence \( s \in L_2 \). Now, if \( C_x = C_a \), by construction of \( C_a \), \( s \in C_a \), a contradiction. Hence, \( C_x = C_b \). But then by construction of \( C_b \), \( r \in C_b \), a contradiction again.

Now, suppose that \( l \) itself is not separated from some other vertex anymore. By the previous case, vertex \( l \) is separated from all its neighbours, as well as from all other vertices of \( R \) since they either belong to the code or their neighbour does, and \( l \) is adjacent to only one vertex of \( C_x \). Thus, suppose that \( l' \) is a vertex of \( L \) and \( l' \) are no longer separated by \( C_x \). Since \( l \) was \( C_x \)-isolated, \( r \) did not belong to \( C_x \). But since \( C_x \)-was dominating, \( l' \) was dominated. Moreover \( l' \) was not dominated by a neighbour, since otherwise \( l \) and \( l' \) would be separated by this neighbour in \( C_x \). Hence, \( l' \) was also \( C_x \)-isolated. But then, in the last step of the construction of \( C_x \), one of them, say \( l \), has been considered first and replaced by \( r \). If \( r \) is not a neighbour of \( l' \), they are separated and we are done. Otherwise, after the modification, \( l' \) was no longer \( C_x \)-isolated and \( l' \) still belongs to \( C_x \), and \( l, l' \) are separated by \( l' \). This completes the proof of the separation property.

Moreover, note that point number 3 of Definition 6 remains verified as no vertex of \( R \) is removed from neither \( C_a \) or \( C_b \) in the last step of their construction.

Finally, observe that thanks to the last step of the constructions, there are no \( C_x \)-isolated \((x \in \{a, b\})\) vertices in \( L \) anymore. Moreover, this step has not created any \( C_x \)-isolated vertices in \( R \). Indeed, the vertices which are added, did not belong to \( C_x \)-, and hence their neighbour in \( R \) did.

This completes the proof of the validity of both constructions \( C_a \) and \( C_b \).

Let us now determine a lower bound on the cardinality of \((L \cup R) \setminus C_x \), for \( x \in \{a, b\} \). Taking into account that \(|L| \leq |R_1| \), we obtain:

\[
|L \cup R \setminus C_a| \geq |L_1| + |L_2| + |R_1| + |R_2| - |C_a| \\
\geq \frac{|R_1|}{2} + \frac{|R_2|}{2} - \min \left\{ \frac{|L_1|}{2}, \frac{|R_2|}{2} \right\}
\]

Thus, both following equations hold:

\[
|L \cup R \setminus C_a| \geq \frac{|R_1|}{2} + \frac{|R_2|}{2} - \frac{|L_1|}{2} \geq \frac{|R_2|}{2} \tag{1}
\]

\[
|L \cup R \setminus C_a| \geq \frac{|R_1|}{2} + \frac{|R_2|}{2} - \frac{|R_2|}{2} = \frac{|R_1|}{2} \geq \frac{|L_1|}{2} \tag{2}
\]

Similarly,

\[
|L \cup R \setminus C_b| \geq |L_1| + |L_2| + |R_1| + |R_2| - |C_b| \\
\geq |L_2| + \frac{|R_1|}{2} + \frac{|R_2|}{2} \\
\geq |L_2| + \frac{|L_1|}{2} - \frac{|R_2|}{2} \tag{3}
\]

\[
= |L| - \frac{|L_1|}{2} - \frac{|R_2|}{2}
\]

Hence intuitively, the previous equations show that our two codes fit to two different situations: \( C_a \) is useful when either \(|L_1| \) or \(|R_2| \) is large enough compared to \(|L| \), whereas \( C_b \) is useful when \(|L_1| + |R_2| \) is small enough compared to \(|L| \). Let \( C \in \{C_a, C_b\} \) be the code having the minimum cardinality. Then, using inequalities (1), (2) and (3) and denoting \( b = \max \left\{ \frac{|L_1||R_1|}{|L|} \right\} \) we get:
Figure 6: Vertices $u, v$ with $(N(u) \cup N(v)) \setminus \{u, v\} \subseteq S$

$$|(L \cup R) \setminus C| \geq \max \left\{ \frac{|L_1|}{2}, \frac{|R_2|}{2}, |L| - \frac{|L_1|}{2} - \frac{|R_2|}{2} \right\}$$

$$= \frac{|L|}{2} \cdot \max \left\{ \frac{|L_1|}{|L|}, \frac{|R_2|}{|L|}, 2 - \frac{|L_1| + |R_2|}{|L|} \right\}$$

$$\geq \frac{|L|}{2} \cdot \max \left\{ \max \{|L_1|, |R_2|\}, 2 - 2 \cdot \max \{|L_1|, |R_2|\} \right\}$$

$$= \frac{|L|}{2} \cdot \max \{b, 2 - 2b\}$$

$$\geq \frac{|L|}{2} \cdot \min_{b \geq 0} \{\max \{b, 2 - 2b\}\}$$

Note that $\min_{b \geq 0} \{\max \{b, 2 - 2b\}\} = \frac{2}{3}$. Hence, we get:

$$|(L \cup R) \setminus C| \geq \frac{|L|}{2} \cdot \frac{2}{3} = \frac{|L|}{3}$$

Note that equality in the previous inequality is achieved when $|L_1| = |R_1| = |R_2| = 2|L_2|$.

Putting $L' = (L \cup R) \setminus C$, we obtain the claim of the lemma.

2.5. The main result

We are now ready to prove the main theorem of this paper. The proof has been sketched in Algorithm 1, we now provide all the details.

**Theorem 13.** Let $G$ be a connected identifiable triangle-free graph on $n$ vertices with maximum degree $\Delta \geq 3$. Then $\gamma'(G) \leq n - \frac{n}{\Delta + m(\Delta + 1)} = n - \frac{n}{\Delta + \Delta - 1}$.

**Proof.** Let $\mathcal{F} = \{F_1, \ldots, F_{|\mathcal{F}|}\}$ be the set of all nontrivial equivalence classes over the false twin relation $\equiv$ over $V(G)$. Let $X = \cup_{i=1}^{|\mathcal{F}|} F_i$ and $Y = V(G) \setminus X$. We distinguish two cases.

**Case 1:** $|Y| \geq \frac{3n}{\Delta + 2}$.

In this case, let $S$ be an independent set of $G[Y]$ given by Lemma 10: we have $|S| \geq \frac{\ln \Delta - 1}{\Delta} |Y| \geq \frac{3n(\ln \Delta - 1)}{\Delta(\ln \Delta + 2)}$. Consider all pairs $u, v$ of vertices of $G$ such that $u$ and $v$ are adjacent, both $u$ and $v$ have degree at least 2, and all the vertices of $N(u) \cup N(v) \setminus \{u, v\}$ belong to $S$ (see Figure 6 for an illustration). Since all neighbours of $u$ and $v$ (except $u$ and $v$ themselves) are in $S$, these neighbours form an independent set. Let $M$ be the set of all edges $uv$ such that $u$ and $v$ form such a pair. By the previous remark, $M$ is a strong induced matching of $G$. Let us denote $L = L(M)$ and $R = R(M)$. Note that we have $L(M) \subseteq S$. Also note that $M$ might be empty if there are no such pairs of vertices.

Let us now partition $V(G)$ into two subsets of vertices: $L \cup R$ on the one hand, and $V(G) \setminus (L \cup R)$ on the other hand. Such a partition is illustrated in Figure 7. Note that $G[L \cup R]$ is identifiable by Observation 5. Let us show that $G[V(G) \setminus (L \cup R)]$ is also identifiable. By contradiction, suppose it is not the case and let $u, v$ be a pair of vertices such that $B_{G[V(G) \setminus (L \cup R)]}(u) = B_{G[V(G) \setminus (L \cup R)]}(v)$. Vertices $u$ and $v$ are therefore
adjacent, and since \( G \) is triangle-free, neither \( u \) nor \( v \) has other neighbours within \( G[V(G) \setminus (L \cup R)] \). Since \( G \) is identifiable, at least one of them has a neighbour in \( L \). Suppose they both have a neighbour in \( L \). Then by construction of \( S \), \( u \) and \( v \) both do not belong to \( S \). But then \( u \) and \( v \) should belong to \( R \), a contradiction. Thus, one of them, say \( u \), has degree 1 in \( G \), and all neighbours of \( v \) belong to \( L \subseteq S \). But by the first property of \( S \) in Lemma 10, at least one vertex at distance 2 of \( u \) does not belong to \( S \), a contradiction.

We will now build two subsets \( C_1 \subseteq L \cup R \) and \( C_2 \subseteq V(G) \setminus (L \cup R) \) such that \( C = C_1 \cup C_2 \) is an identifying code of \( G \).

- **Building** \( C_1 \subseteq L \cup R \).
  If \( L \cup R = \emptyset \) we take \( C_1 = \emptyset \). Otherwise, we build \( C_1 \) using Lemma 12: applying it to \( G \) and \( M \), we know that there exists an \((L, R)\)-quasi-identifying code \( C_1 \) of \( G \) without \( C_1 \)-isolated vertices. From Lemma 12 we also know that \(|L'| \geq \frac{|L|}{3}\), where \( L' = (L \cup R) \setminus \{C_1\} \).

- **Building** \( C_2 \subseteq V(G) \setminus (L \cup R) \).
  Again if \( V(G) \setminus (L \cup R) = \emptyset \) we take \( C_2 = \emptyset \).

Otherwise, we take \( C_2 \) to be the complement of \( S \) in \( V(G) \setminus (L \cup R) \): \( C_2 = (V(G) \setminus (L \cup R)) \setminus S \). Let us show that \( C_2 \) is a \((V(G) \setminus (L \cup R))\)-identifying code of \( G \).

First, recall that \( G' = G[V(G) \setminus (L \cup R)] \) is identifiable. Note that \( S \) does not contain any vertex \( v \) which is isolated in \( G' \). Indeed, \( G \) does not contain any isolated vertex, hence if \( v \) is isolated in \( G' \), \( v \) has a neighbour in \( L \). But since \( L \subseteq S \), \( S \) would not be independent, a contradiction. We also claim that for each vertex \( v \) of degree 1 in \( G' \), there is a vertex at distance 2 of \( v \) in \( G' \) not belonging to \( S \). Let \( w \) be the unique neighbour of \( v \) in \( G' \). If \( v \) is also of degree 1 in \( G \), since \( G' \) has no pair of twins, by the first property of \( S \) in Lemma 10, \( w \) must have a neighbour \( x \) not in \( S \). Vertex \( x \) cannot belong to \( L \), hence it belongs to \( G' \) and we are done. Now, if \( v \) is not of degree 1 in \( G \), all its neighbours in \( G \) other than \( w \) belong to \( L \). But since \( G' \) is identifiable, \( w \) has at least one neighbour other than \( v \), belonging to \( G' \) but not to \( S \), since otherwise \( v \) and \( w \) would belong to set \( R \). Finally, by construction of \( G' \), there are no isolated edges in \( G[V(G') \setminus S] \).

Under these conditions we can apply Proposition 2 on \( G' \), and set \( S \) restricted to \( V(G') \), which shows that \( C_2 \) is a \((V(G) \setminus (L \cup R))\)-identifying code of \( G \).

We now have an \((L, R)\)-quasi-identifying code \( C_1 \) of \( G \) without \( C_1 \)-isolated vertices, and showed that \( C_2 \) is a \((V(G) \setminus (L \cup R))\)-identifying code of \( G \). Moreover, \( S \) does not contain any pair of false twins.
Furthermore, since $C_2$ is the complement of $S$ in $G[V(G) \setminus (L \cup R)]$, all neighbours of $L$ in $G[V(G) \setminus (L \cup R)]$ belong to $C_2$. Therefore, we can apply Proposition 7 and $C = C_1 \cup C_2$ is an identifying code of $G$.

Let us now upper-bound the size of $C$. To this end, we lower-bound the size of its complement. From the construction of $C_1$ and $C_2$, we have $V(G) \setminus C = (S \setminus L) \cup L'$.

Since $L \subseteq S$ and $|L'| \geq \frac{|L|}{3}$, we have $|(S \setminus L) \cup L'| \geq \frac{|S|}{3}$.

Hence, we get: 

$$|V(G) \setminus C| \geq \frac{|S|}{3} \geq \frac{\ln \Delta}{\Delta(\ln \Delta + 2)} n^2 = \frac{n}{\Delta + \ln \Delta - 1}.$$ 

Hence, $|C| \leq n - \frac{n}{\Delta + \ln \Delta - 1}$.

**Case 2:** $|Y| \leq \frac{3n}{\ln \Delta + 2}$. 

Then, $|X| \geq n - \frac{3n}{\ln \Delta + 2}$. Since each set of $\mathcal{F}$ has size at most $\Delta$, we have:

$$|\mathcal{F}| \geq \frac{|X|}{\Delta} \geq \frac{\ln \Delta}{\Delta(\ln \Delta + 2)} n = \frac{n}{\Delta + \ln \Delta - 1}.$$ 

Since $\Delta \geq 3$, $G$ is not isomorphic to $C_4$ and we can apply Proposition 3: $G$ has an identifying code of size at most $n - |\mathcal{F}| \leq n - \frac{n}{\Delta + \ln \Delta - 1}$.

It can be noted that the value of the bound of Theorem 13 heavily relies on the corollary of J. Shearer’s bound (Corollary 9). For large values of $\Delta$, this bound is nearly optimal [24]. However, directly using the slightly stronger bound of J. Shearer (Theorem 8) or a stronger bound holding for some particular class of graphs, one could obtain a strengthened result as follows. Let $G$ be a nontrivial connected identifiable triangle-free graph on $n$ vertices having maximum degree $\Delta$. Suppose each subgraph $H$ of $G$ has an independent set of size at least $f(\Delta)|V(H)|$, where $f(\Delta) \leq \frac{1}{3}$. Then, the value $\frac{n\Delta - 1}{\Delta}$ in Lemma 10 can be replaced by $f(\Delta)$, and the condition for applying Case 1 in the proof of Theorem 13 can be replaced by $|Y| \geq \frac{3n}{f(\Delta) + 2}$. We then get the following theorem:

**Theorem 14.** Let $G$ be a connected identifiable triangle-free graph on $n$ vertices with maximum degree $\Delta$ such that each subgraph $H$ of $G$ has an independent set of size at least $f(\Delta)|V(H)|$, where $f(\Delta) \leq \frac{1}{3}$. Then $\gamma^{ID}(G) \leq n - \frac{n}{\Delta + f(\Delta)}$.

### 3. Concluding remarks

We conclude this paper by the following remarks.

**Remark 1.** Let $G$ be a triangle-free graph without any pair of false twins. By considering Case 1 of the proof of Theorem 13, we have $Y = V(G)$ and hence $G$ has an identifying code of size at most $n - \frac{n}{\ln \Delta} = n - \frac{n}{o(\Delta)}$.

Hence any class of connected triangle-free graphs of maximum degree $\Delta$ having its minimum identifying code of size at least $n - \frac{n}{o(\Delta)}$ should contain false twins. Note that this is the case of the complete $(\Delta - 1)$-ary tree already mentioned in the introduction (all its leaves are false twins), and of the classes of graphs described in [10] (which are built using copies of small complete bipartite graphs $K_{\Delta,\Delta}$ joined to each other, and therefore contain many false twins).
Remark 2. We note that our proofs provide a polynomial-time algorithm to compute the identifying codes of Theorem 13. Indeed, their constructions are based on the codes computed in Lemmas 11, and 12, and the independent set of Lemma 10 for the first code, and on the construction of Proposition 3 for the second code. All these constructions are described in the corresponding proofs and can be done in polynomial time. Let us give an explicit complexity bound.

We observe that the running time of the constructions is at most of the order $O(n^2 \ln n)$. Indeed, the most difficult step is to compute and compare the neighbourhoods of the vertices in order to build the false twin equivalence classes in the proof of Theorem 13. To do this one can represent each neighbourhood as a binary word of length $n$. Bitwise comparing two of them requires $O(n)$ operations, hence a classical sorting algorithm can sort them all in time $O(n^2 \ln n)$. Comparing them takes $O(n^2)$ time. Moreover, the construction of the independent set of Lemma 10 is based on Theorem 8 given in [24]. There, the author gives a randomized linear-time algorithm for computing the independent set. Note that the random (constant-time) step of this algorithm can be turned into a deterministic linear-time computation, which leads to an $O(n^2)$ algorithm. All other steps and constructions can also be done in time $O(n^2)$. Hence, we have the following theorem.

Theorem 15. Let $G$ be a connected identifiable triangle-free graph on $n$ vertices with maximum degree $\Delta \geq 3$. Then, an identifying code of $G$ having cardinality at most $n - \frac{n}{\Delta + \frac{\delta}{\delta + 1}}$ can be computed in time $O(n^2 \ln n)$.

Remark 3. In this paper, we have considered triangle-free graphs, that is, graphs of girth at least 4. It is natural to ask whether much stronger bounds on parameter $\gamma_{ID}$ hold for graphs of larger girth. However note that the answer to this question is negative because of the complete $(\Delta - 1)$-ary tree on $n$ vertices $T$, which has infinite girth and was already mentioned in the introduction. This graph has infinite girth and $\gamma_{ID}(T) = \lfloor n - \frac{n}{\Delta-1+1/\Delta} \rfloor$ [4].

However, with an additional condition on the minimum degree of the graph, the question was answered in the positive in [10] and very recently in [13], where the following bounds are given.

Theorem 16 ([10]). Let $G$ be a connected identifiable graph on $n$ vertices having minimum degree at least 2 and girth at least 5. Then $\gamma_{ID}(G) \leq \frac{\Delta}{\delta} + 1$.

Theorem 17 ([13]). Let $G$ be an identifiable graph on $n$ vertices having minimum degree $\delta \geq 1$ and girth at least 5. Then $\gamma_{ID}(G) \leq \left(\frac{\Delta}{\delta} + o(1)\right)\frac{\delta + 1}{\delta}n$, where $o(1)$ is a function of $\delta$ tending to 0 when $\delta$ tends to infinity.

Note that these two bounds are much stronger than any bound of the form $n - \frac{n}{\Theta(\Delta)}$, such as the one of Conjecture 1. They are best possible in the sense that relaxing either the condition on girth 5 or minimum degree 2, there are graphs which have much larger identifying codes. If one drops the minimum degree 2 condition, such a graph is the complete $(\Delta - 1)$-ary tree. If one drops the girth 5 condition, there are $\Delta$-regular graphs $(\Delta \geq 2)$ having girth 4 and their minimum identifying code of size $n - \frac{n}{\Theta(\Delta)}$ [10]. We would like to refer the interested reader to [13], where this question is studied in more detail.

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] D. Auger, I. Charon, O. Hudry and A. Lobstein. Complexity results for identifying codes in planar graphs, International Transactions in Operational Research 17(6):691–710, 2010.
[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] N. Bertrand, I. Charon, O. Hudry and A. Lobstein. 1-identifying codes on trees. Australasian Journal of Combinatorics 31:21–35, 2005.
[5] M. Burley and J. P. Ullry. Parity graphs, Annals of Discrete Mathematics 16:1–26, 1982.
[6] I. Charon, S. Gravier, O. Hudry, A. Lobstein, M. Mollard and J. Moncel. A linear algorithm for minimum 1-identifying codes in oriented trees. Discrete Applied Mathematics 154(8):1246–1253, 2006.
[7] I. Charon, I. Honkala, O. Hudry and A. Lobstein. Structural properties of twin-free graphs. *Electronic Journal of Combinatorics* 14(1):R16, 2007. http://www.combinatorics.org.

[8] I. Charon, O. Hudry and A. Lobstein. Extremal cardinalities for identifying and locating-dominating codes in graphs. *Discrete Mathematics* 307(5-5):356–366, 2007.

[9] C. Colbourn, P. J. Slater and L. K. Stewart. Locating–dominating sets in series-parallel networks. *Congressus Numerantium* 56:135–162, 1987.

[10] F. Foucaud. Identifying codes in special graph classes. Master thesis, Université Bordeaux 1, France, June 2009. Available online at http://www.labri.fr/perso/foucaud/Research/MastersThesis/.

[11] 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.

[12] F. Foucaud, R. Naserasr and A. Parreau. Extremal problems on identifying codes in digraphs and Bondy’s theorem on induced subsets. Submitted for publication, 2010. Available online at http://arxiv.org/abs/1010.2985.

[13] F. Foucaud and G. Perarnau. Bounds for identifying codes in terms of degree parameters. Submitted for publication, 2011. Available online at http://arxiv.org/abs/1103.3756.

[14] S. Gravier, M. Kovše, M. Mollard and J. Moncel. New results on variants of covering codes in Sierpiński graphs. Manuscript, 2011.

[15] 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.

[16] S. Gravier, J. Moncel and A. Semri. Identifying codes of cycles, *European Journal of Combinatorics* 27(5):767–776, 2006.

[17] 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.

[18] S. Janson and T. Laihonen. Note: On the size of identifying codes in binary hypercubes. *Journal of Combinatorial Theory Series A* 116(5):1087–1096, 2009.

[19] 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.

[20] M. Laifenfeld, A. Trachtenberg, R. Cohen and D. Starobinski. Joint monitoring and routing in wireless sensor networks using robust identifying codes. *Proceedings of IEEE Broadnets 2007*, pages 197–206, September 2007.

[21] 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.

[22] S. Ranto, I. Honkala and T. Laihonen. Two families of optimal identifying codes in binary hamming spaces. *IEEE Transactions on Information Theory* 48(5):1200–1203, 2002.

[23] S. Ray, R. Ungrangsi, F. De Pellegrini, A. Trachtenberg and D. Starobinski. Robust location detection in emergency sensor networks. *Proceedings of IEEE INFOCOM 2003*, pages 1044–1053, April 2003.

[24] J. Shearer. A note on the independence number of triangle-free graphs. *Discrete Mathematics* 46(1):83–87, 1983.

[25] P. J. Slater. Domination and location in acyclic graphs. *Networks* 17(1):55–64, 1987.

[26] P. J. Slater and D. F. Rall. On location-domination numbers for certain classes of graphs. *Congressus Numerantium* 45:97–106, 1984.