Bounds for identifying codes in terms of degree parameters

Florent Foucaud, Guillem Perarnau

Bounds for identifying codes in terms of degree parameters*

Florent Foucaud† Guillem Perarnau‡

June 17, 2011

Abstract

An identifying code is a subset of vertices of a graph such that each vertex is uniquely determined by its neighbourhood within the identifying code. If $\gamma^{\mathrm{ID}}(G)$ denotes the minimum size of an identifying code of a graph G, it was conjectured by F. Foucaud, R. Klasing, A. Kosowski and A. Raspaud that if a connected graph G has n vertices and maximum degree d and admits an identifying code, then $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{d} + O(1)$. We use probabilistic tools to show that for sufficiently large d, $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{\Theta(d)}$ holds for a large class of graphs containing, among others, all regular graphs and all graphs of bounded clique number. This settles the conjecture (up to constants) for these classes of graphs. In the general case, we prove $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{\Theta(d^3)}$. In a second part, we prove that in any graph G of minimum degree δ and girth at least 5, $\gamma^{\mathrm{ID}}(G) \leq (1 + o_\delta(1))\frac{3\log\delta}{2\delta}n$. Using the former result, we give sharp estimates for the size of the minimum identifying code of random d-regular graphs, which is about $\frac{\log d}{d}n$.

1 Introduction

Given a graph G, an identifying code $\mathcal{C}$ is a dominating set such that for any two vertices, their neighbourhoods within $\mathcal{C}$ are nonempty and distinct. This property can be used to distinguish all vertices of the graph from each other. Identifying codes have found applications to various fields since the introduction of this concept in [18]. These applications include the location of threats in facilities using sensors [23], error-detection schemes [18] and routing [19] in networks, as well as the structural analysis of RNA proteins [17].

In this paper, we address the question of lower and upper bounds on the size of an identifying code, thus extending earlier works on such questions (see e.g. [21, 8, 14, 11, 12]). We focus on degree-related graph parameters such as the minimum and maximum degree, and also study the case of regular graphs. An important part of the paper is devoted to giving the best possible upper bound for the size of an identifying code depending on the order and the maximum degree of the graph, a question raised in [10]. We also give improved bounds for graphs of large girth and study identifying codes in random regular graphs. The main tools used herein are probabilistic.

We begin by giving our notations and defining the concepts used throughout the paper.

As *graphs* and unless specifically mentioned, we understand simple, undirected and finite graphs. The vertex set of a graph G is denoted by $V(G)$ and its edge set $E(G)$. We also denote its order by $n = |V(G)|$. The maximum degree of G will be denoted by $d = d(G)$, its minimum degree, by $\delta = \delta(G)$, and its average degree, by $\overline{d} = \overline{d}(G)$. We denote by $u \sim v$, the adjacency between two vertices u and v, and by $u \not\sim v$, their non-adjacency. The set of neighbours of some vertex v is called its *open neighbourhood* and denoted by $N(v)$, whereas the set of its neighbours and itself is called its *closed neighbourhood* and denoted by $N[v]$. If two distinct vertices u, v are such that $N[u] = N[v]$, they are called *twins*. If $N(u) = N(v)$ but $u \not\sim v$, u and v are called *false twins*. The symmetric difference between two sets A and B is denoted by $A\Delta B$. We use $\log(x)$ to denote the natural logarithm of x. We also make use of the standard asymptotic notations o, O, Θ, Ω and ω. Usually the asymptotics are taken either on variables d, δ or n. We use the

*This research was supported by the ANR Project IDEA - Identifying coDes in Evolving grAphs, ANR-08-EMER-007, 2009-2011. Both authors thank the hospitality and support of the graduate research training network "Methods for Discrete Structures", where part of the research took place. The second author wants to thank the FPU grant from the Ministerio de Educación de España.

†LaBRI - Université de Bordeaux, 351 cours de la Libération, 33405 Talence cedex, France.

‡MA4 - Universitat Politècnica de Catalunya, C/ Jordi Girona 1-3, 08034 Barcelona, Spain.

notations $o_x(1)$ and $O_x(1)$ to stress the fact that the asymptotic is taken on variable x. If we write $o(1)$ or $O(1)$, then by convention the asymptotic is taken only on n, the number of vertices.

Given a graph G and a subset $\mathcal{C}$ of vertices of G, $\mathcal{C}$ is called a *dominating set* if each vertex of $V(G)\setminus\mathcal{C}$ has at least one neighbour in $\mathcal{C}$. Set $\mathcal{C}$ is called a *separating set* of G if for each pair u, v of vertices of G, $N[u]\cap\mathcal{C} \neq N[v]\cap\mathcal{C}$ (equivalently, $(N[u]\Delta N[v])\cap\mathcal{C}\neq\emptyset$). We have the following definition:

Definition. *Given a graph G, a subset of vertices of $V(G)$ which is both a dominating set and a separating set is called an* identifying code *of G.*

First of all it must be stressed that not every graph can have an identifying code. Observe that a graph containing twin vertices does not admit any separating set or identifying code. In fact a graph admits an identifying code if and only if it is *twin-free*, i.e. it has no pair of twins (one can see that if G is twin-free, $V(G)$ is an identifying code of G). Note that if for three distinct vertices u, v, w of a twin-free graph G, $N[u]\Delta N[v] = \{w\}$, then w belongs to any identifying code of G. In this case we say that w is *uv-forced*, or simply *forced*. Observe that any isolated vertex must belong to any identifying code for the reason that it must be dominated. For example, an edgeless graph needs all the vertices in any identifying code. Hence, the bounds of this paper only hold for graphs with few isolated vertices. In order to shorten the statements of our results, we assume that all considered graphs do not have any isolated vertices.

For a given graph, the problem of finding a minimum identifying code is known to be NP-hard, even in graphs having small maximum degree and high girth (to be precise, in planar graphs of maximum degree 4 having arbitrarily large girth [2] and planar graphs of maximum degree 3 and girth at least 9 [3]).

The minimum size of an identifying code in a graph G, denoted $\gamma^{\text{ID}}(G)$, is the *identifying code number* of G. It is known that for a twin-free graph G on n vertices having at least one edge, we have:

$$\lceil \log_2(n+1)\rceil \leq \gamma^{\text{ID}}(G) \leq n-1$$

The lower bound is proved in [18] and the upper bound, in [14]. Both bounds are tight and all graphs reaching these two bounds have been classified (see [21] for the lower bound and [11] for the upper bound).

When considering graphs of given maximum degree d, it was shown in [18] that the lower bound can be improved to $\gamma^{\text{ID}}(G) \geq \frac{2n}{d+2}$. This bound is tight and a classification of all graphs reaching it has been proposed in [10]. For any d, these graphs include some regular graphs and graphs of arbitrarily large girth.

It was conjectured in [12] that the following upper bound holds.

Conjecture 1 ([12])**.** *Let G be a nontrivial connected twin-free graph of maximum degree d. Then $\gamma^{\text{ID}}(G) \leq n - \frac{n}{d} + O(1)$.*

Graphs of maximum degree d such that $\gamma^{\text{ID}}(G) = n - \frac{n}{d}$ are known (e.g. the complete bipartite graph $K_{d,d}$ and richer classes of graphs described in Section 6). Therefore if Conjecture 1 holds, for any graph G on n vertices and of maximum degree d we would have $\frac{2}{d+2}n \leq \gamma^{\text{ID}}(G) \leq n - \frac{n}{d} + O(1)$, with both bounds being tight.

Note that Conjecture 1 holds for graphs of maximum degree 2 (see [15]). It was shown in [11] that $\gamma^{\text{ID}}(G) \leq n - \frac{n}{\Theta(d^5)}$, and $\gamma^{\text{ID}}(G) \leq n - \frac{n}{\Theta(d^3)}$ when G has no forced vertices (in particular, this is true when G is regular). It is also known that the conjecture holds in an asymptotic way if G is triangle-free: then, $\gamma^{\text{ID}}(G) \leq n - \frac{n}{d(1+o_d(1))}$ [12].

Identifying codes have been previously studied in two models of random graphs, that is the classic random graph model [13] and the model of random geometric graphs [22]. To our knowledge random regular graphs have not been studied in the context of identifying codes.

In this paper, we further study Conjecture 1 and prove that it is tight (up to constants) for large enough values of d and for a large class of graphs, including regular graphs and graphs of bounded clique number (Corollaries 8 and 12). In the general case, we prove that $\gamma^{\text{ID}}(G) \leq n - \frac{n}{\Theta(d^3)}$ (Corollary 10). These results improve the known bounds given in [11] and support Conjecture 1. Moreover, we show that the much improved upper bound of $\gamma^{\text{ID}}(G) \leq (1+o_\delta(1))\frac{3\log\delta}{2\delta}n$ holds for graphs having girth at least 5 and minimum degree δ (Theorem 18). This bound is used to give an asymptotically tight bound of about $\frac{\log d}{d}n$ for the identifying code number of almost all random d-regular graphs (Corollary 22).

We summarize our results for the special case of regular graphs in Table 1 and compare them to the bound for the dominating set problem (the table contains references for both the bound and its tightness).

All bounds are asymptotically tight. We note that identifying codes behave far from dominating sets in general, as shown by the first lines of the table: there are regular graphs having much larger identifying code number than domination number. However, for larger girth and for almost all regular graphs, the bounds for the two problems coincide asymptotically, as shown by the last lines of the table.

	Identifying codes	Dominating sets
in general	$n - \frac{n}{85d}$ Thm. 7, Constr. 24	$\sim \frac{\log d}{d} n$ [1], [24]
girth 4	$n - \frac{n}{d(1+o_d(1))}$ [12], Constr. 25	$\sim \frac{\log d}{d} n$ [1], [24]
girth 5	$(1 + o_d(1))\frac{3 \log d}{2d} n$ Thm. 18, Thm. 20	$\sim \frac{\log d}{d} n$ [1], [24]
almost all graphs	$\frac{\log d + \log\log d + O_d(1)}{d} n$ Thm. 19, Thm. 20	$\sim \frac{\log d}{d} n$ [1], [24]

Table 1: Summary of the upper bounds for d-regular graphs

In order to prove our results, we use probabilistic techniques. For some results, we use the weighted version of Lovász' Local Lemma to show the existence of an identifying code, together with the Chernoff bound to show that this code is small enough. We also make use of other probabilistic techniques such as the Alteration Method [1] in order to give better bounds in more restricted cases. Finally, we work with the Configuration Model [7] in order to compute the identifying code number of almost all random regular graphs.

The organization of this paper is as follows. In Section 2 we state some preliminary results which will be used throughout the paper. In Section 3, we improve the known upper bounds on the identifying code number of graphs of maximum degree d. This gives new large families of graphs for which Conjecture 1 holds (up to constants). In Section 4, we give an upper bound for graphs having minimum degree δ and girth at least 5. In Section 5, we give sharp bounds for the identifying code number of almost all d-regular graphs. A further section is dedicated to various constructions of families of graphs which show the tightness of some of our results (Section 6).

2 Preliminary results

We first recall a well-known probabilistic tool: the Lovász Local Lemma. We use its weighted version, a particularization of the general version where each event has an assigned weight. The proof can be found in [20].

Lemma 2 (Weighted Local Lemma [20])**.** *Let $\mathcal{E} = \{E_1, \ldots, E_M\}$ be a set of (typically "bad") events such that each E_i is mutually independent of $\mathcal{E} \setminus (\mathcal{D}_i \cup \{E_i\})$ where $\mathcal{D}_i \subseteq \mathcal{E}$. Suppose that there exist some integer weights $t_1, \ldots, t_M \geq 1$ and a real $p \leq \frac{1}{4}$ such that for each $1 \leq i \leq M$:*

- $\Pr(E_i) \leq p^{t_i}$*, and*
- $\sum_{E_j \in \mathcal{D}_i} (2p)^{t_j} \leq \frac{t_i}{2}$

Then $\Pr(\bigcap_{i=1}^{M} \overline{E_i}) \geq \prod_{i=1}^{M} (1 - (2p)^{t_i}) > 0$.

Note that in Lemma 2, since $p \leq \frac{1}{4}$, we have:

$$\Pr(\textstyle\bigcap_{i=1}^{M} \overline{E_i}) \geq \exp\left\{-2 \log 2 \sum_{i=1}^{M} (2p)^{t_i}\right\}. \tag{1}$$

We also use the following version of the well-known Chernoff bound, which is a reformulation of Theorem A.1.13 in [1].

Theorem 3 (Chernoff bound). *Let X be a random variable of n independent trials of probability p, and let $a > 0$ be a real number. Then* $\Pr(X - np \leq -a) \leq e^{-\frac{a^2}{2np}}$.

The following observation gives an equivalent condition for a set to be an identifying code, and follows from the fact that for two vertices u, v at distance at least 3 from each other, $N[u]\Delta N[v] = N[u] \cup N[v]$.

Observation 4. *For a graph G and a set $\mathcal{C} \subseteq V(G)$, if $\mathcal{C}$ is dominating and $N[u] \cap \mathcal{C} \neq N[v] \cap \mathcal{C}$ for each pair of vertices u, v at distance at most two from each other, then $N[u] \cap \mathcal{C} \neq N[v] \cap \mathcal{C}$ for each pair of vertices of the graph.*

The next observation is immediate, but it is worth mentioning here.

Observation 5. *Let G be a twin-free graph and $\mathcal{C}$, an identifying code of G. Any set $\mathcal{C}'$ such that $\mathcal{C} \subseteq \mathcal{C}'$ is also an identifying code of G.*

The next proposition shows an upper bound on the number of false twins in a graph.

Proposition 6. *Let G be a graph on n vertices having maximum degree d and no isolated vertices, then G has at most $\frac{n(d-1)}{2}$ pairs of false twins.*

Proof. Let us build a graph H on $V(G)$, where two vertices u, v are adjacent in H if they are false twins in G. Note that since a vertex can have at most $d-1$ false twins, H has maximum degree $d-1$. Therefore it has at most $\frac{n(d-1)}{2}$ edges and the claim follows. □

Note that the bound of Proposition 6 is tight since in a complete bipartite graph $K_{d,d}$, $n = 2d$ and there are exactly $2\binom{d}{2} = \frac{n(d-1)}{2}$ pairs of false twins.

3 Upper bounds on the identifying code number

3.1 Main theorem

In this section, we improve the known upper bounds of [11] on the identifying code number by using the Weighted Local Lemma, stated in Lemma 2.

In the following, given a graph G on n vertices, we will denote by $f(G)$ the proportion of non-forced vertices of G, i.e. the ratio $\frac{x}{n}$, where x is the number of non-forced vertices of G.

Theorem 7. *There exists an integer d_0 such that for each twin-free graph G on n vertices having maximum degree $d \geq d_0$, $\gamma^{ID}(G) \leq n - \frac{nf(G)^2}{85d}$.*

Proof. Let F be the set of forced vertices of G, and $V' = V(G) \setminus F$. Note that $|V'| = nf(G)$. By the definition of a forced vertex, any identifying code must contain all vertices of F.

In this proof, we first build a set S in a random manner by choosing vertices from V'. Then we exhibit some "bad" configurations - if none of those occur, the set $F \cup (V(G) \setminus S)$ is an identifying code of G. Using the Weighted Local Lemma, we compute a lower bound on the (non-zero) probability that this holds. Finally, we use the Chernoff bound to show that with non-zero probability, the size of S is also large enough for our purposes. This shows that such a "good" large set S exists, and it can be used to build an identifying code that has a sufficiently small size.

Let $p = p(d)$ be a probability which will be determined later. We build the set $S \subseteq V'$ such that each vertex of V' independently belongs to S with probability p. Therefore the random variable $|S|$ follows a binomial distribution Bin($nf(G)$,p) and has expected value $\mathbb{E}(|S|) = pnf(G)$.

Let us now define the set $\mathcal{E}$ of "bad" events. These are of four types. An illustration of these events is given in Figure 1.

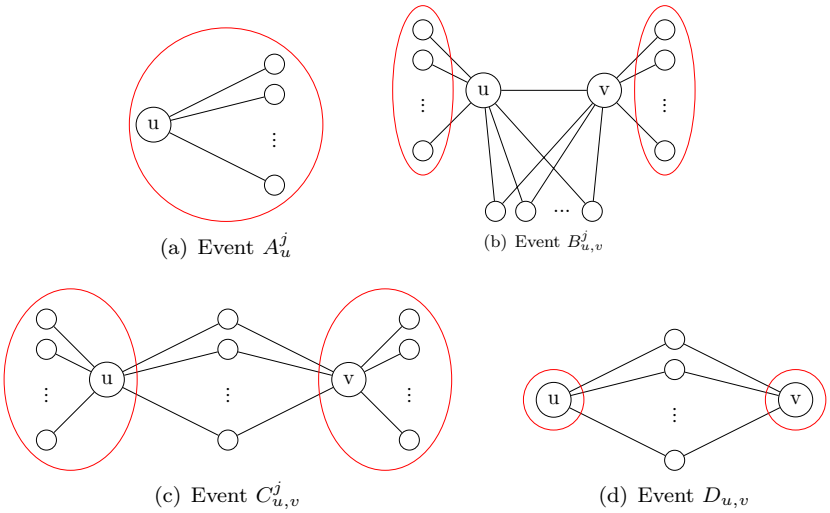
- **Type $\mathbf{A}^j$ ($2 \leq j \leq d+1$):** for each vertex $u \in V'$, let A_u^j be the event that $|N[u] = j|$ and $N[u] \subseteq S$.
- **Type $\mathbf{B}^j$ ($2 \leq j \leq 2d-2$):** for each pair $\{u, v\}$ of adjacent vertices, let $B_{u,v}^j$ be the event that $|(N[u]\Delta N[v])| = j$ and $(N[u]\Delta N[v]) \subseteq S$.
- **Type $\mathbf{C}^j$ ($3 \leq j \leq 2d$):** for each pair $\{u, v\} \in V(G)$ of vertices at distance two from each other, let $C_{u,v}^j$ be the event that $|(N[u]\Delta N[v])| = j$ and $(N[u]\Delta N[v]) \subseteq S$.

- **Type D:** for each pair $\{u, v\} \in V'$ of false twins, let $D_{u,v}$ be the event that $(N[u]\Delta N[v]) = \{u, v\} \subseteq S$.

For the sake of simplicity, we refer to the events of type A^j, B^j and C^j as events of type A, B and C respectively whenever the size of the symmetric difference is not relevant.

Events of type $B^1_{u,v}$ are not defined since then $|N[u]\Delta N[v]| = 1$ and F belongs to the code, so they never happen. Observe that the events $C^j_{u,v}$ and $D_{u,v}$ are just defined over the pairs of vertices in V' because if either u or v belongs to F, the event does not happen.

If no event of type A occurs, $V(G) \setminus S$ is a dominating set of G. If no event of type B occurs, all pairs of adjacent vertices are separated by $V(G) \setminus S$. If no event of type C or D occurs, all pairs of vertices at distance 2 from each other are separated. Thus by Observation 4, $V(G) \setminus S$ is also a separating set of G, and therefore it is an identifying code of G.



(a) Event A^j_u (b) Event $B^j_{u,v}$

(c) Event $C^j_{u,v}$ (d) Event $D_{u,v}$

Figure 1: The "bad" events. The circled vertices belong to set S.

We define the weight t_i of each event $E_i \in \mathcal{E}$ as the number of vertices participating in it (i.e. the number of vertices that must belong to set S so that E_i holds). For $j \geq 2$ and for $T \in \{A^j, B^j, C^j, D\}$, let t_T be the weight of an event of type T (for an event $E_i \in \mathcal{E}$ of type T, $t_i = t_T$). We have the following:

$$t_{A^j} = j \qquad t_{B^j} = j \qquad t_{C^j} = j \qquad t_D = 2$$

Some vertex x can intersect at most $d+1$ events of type A since if it intersects some event A^j_u, then $u \in N[x]$. Vertex x can intersect at most $d(d-1)$ events of type B: supposing x intersects with $B^j_{u,v}$ and u is the vertex adjacent to x, there are d ways to choose u, and $d-1$ ways to choose v among $N(u) \setminus x$. Similarly x can intersect at most $d^2(d-1)$ events of type C: for some event $C^j_{u,v}$, there are $d(d-1)$ possibilities if $x = u$ or $x = v$ and $d(d-1)^2$ if u or v is a neighbour of x. Finally, x can intersect at most $d-1$ events $D_{u,v}$ since x can have at most $d-1$ false twins. For each type T of events ($T \in \{A^j, B^j, C^j, D\}$), let us define $int(v, T)$ to be the number of events of type T containing a given vertex v of G. Hence:

$$\begin{aligned} &\sum_{j=2}^{d+1} int(v, A^j) \leq d+1 \qquad \sum_{j=2}^{2d-2} int(v, B^j) \leq d(d-1) \\ &\sum_{j=3}^{2d} int(v, C^j) \leq d^2(d-1) \qquad int(v, D) \leq d-1 \end{aligned} \tag{2}$$

Let us call E_{IC} the event that no event of $\mathcal{E}$ occurs. Using the Weighted Local Lemma, we want to show that $\Pr(E_{\text{IC}}) > 0$. For two events E_i and E_j of $\mathcal{E}$, we note $i \sim j$ if E_i and E_j are not mutually independent. Here, E_i and E_j are mutually independent whenever the two sets of vertices which must

belong to S in order that the events hold (i.e. the vertices circled in red in Figure 1), are disjoint. In order to apply the Weighted Local Lemma (Lemma 2), the following conditions must hold for each event $E_i \in \mathcal{E}$:

$$\sum_{i \sim j} (2p)^{t_j} \leq \frac{t_i}{2}$$

The latter conditions are implied by the following ones (for each event $E_i \in \mathcal{E}$):

$$\sum_{j=2}^{d+1} \sum_{v \in E_i} int(v, A^j)(2p)^{t_{A^j}} + \sum_{j=2}^{2d-2} \sum_{v \in E_i} int(v, B^j)(2p)^{t_{B^j}} + \sum_{j=3}^{2d} \sum_{v \in E_i} int(v, C^j)(2p)^{t_{C^j}} + \sum_{v \in E_i} int(v, D)(2p)^{t_D} \leq \frac{t_i}{2}$$

Which are implied by:

$$t_i \cdot \max_{v \in E_i} \left\{ \sum_{j=2}^{d+1} int(v, A^j)(2p)^{t_{A^j}} \right\} + t_i \cdot \max_{v \in E_i} \left\{ \sum_{j=2}^{2d-2} int(v, B^j)(2p)^{t_{B^j}} \right\} + t_i \cdot \max_{v \in E_i} \left\{ \sum_{j=3}^{2d} int(v, C^j)(2p)^{t_{C^j}} \right\}$$
$$+ t_i \cdot \max_{v \in E_i} \left\{ int(v, D)(2p)^{t_D} \right\} \leq \frac{t_i}{2}$$

Using the bounds of Inequalities (2) and noting that for any j, $(2p)^{t_{A^j}} \leq (2p)^2$, $(2p)^{t_{B^j}} \leq (2p)^2$ and $(2p)^{t_{C^j}} \leq (2p)^3$, for any event E_i this equation is implied by:

$$(d+1)(2p)^2 + d(d-1)(2p)^2 + d^2(d-1)(2p)^3 + (d-1)(2p)^2 \leq \frac{1}{2}$$

Intuitively, we need to set $p = O(d^{-1})$ in order to solve this inequality. In this case, for some sufficiently large d, it is implied by:

$$8d^2p^2 + 16d^3p^3 \leq 1 \tag{3}$$

Hence, we fix $p = \frac{1}{kd}$ where k is constant. Then Equation (3) holds for $k \geq 3.66$. However, in order to optimize our result, we will choose $k \geq 10$. Using this assumption, we have $p \leq \frac{1}{4}$. Under this condition the Weighted Local Lemma can be applied.

Let M_T be the number of events of type T, where $T \in \left\{A^j, B^j, C^j, D\right\}$. By Lemma 2 we have:

$$\Pr(E_{\mathrm{IC}}) \geq \prod_{j=2}^{d+1} \prod_{i=1}^{M_{A^j}} (1 - (2p)^{t_{A^j}}) \prod_{j=2}^{2d-2} \prod_{i=1}^{M_{B^j}} (1 - (2p)^{t_{B^j}}) \prod_{j=3}^{2d} \prod_{i=1}^{M_{C^j}} (1 - (2p)^{t_{C^j}}) \prod_{i=1}^{M_D} (1 - (2p)^{t_D})$$

Note that $\sum_{j=2}^{d+1} M_{A^j} \leq nf(G)$ since by definition there is one event of type A for each vertex of V'. Moreover, $\sum_{j=2}^{2d-2} M_{B^j} \leq \frac{nd}{2}$ since there is at most one event of type B for each edge in G. We also have that $\sum_{j=3}^{2d} M_{C^j}$ is at most the number of pairs of vertices in V' at distance 2 from each other. This is also at most the number of paths of length 2 with both endpoints in V', which is upper-bounded by $\frac{nf(G)d(d-1)}{2}$. Finally, M_D is at most the number of pairs of false twins in V', i.e. $nf(G)\frac{d-1}{2}$ by Proposition 6. Hence, we have:

$$\Pr(E_{\mathrm{IC}}) \geq (1 - (2p)^2)^{nf(G)} (1 - (2p)^2)^{\frac{nd}{2}} (1 - (2p)^3)^{\frac{nf(G)d(d-1)}{2}} (1 - (2p)^2)^{\frac{nf(G)(d-1)}{2}}$$

Using Lemma 2 (more precisely, we use Equation (1)) and the fact that $p = \frac{1}{kd}$, we obtain:

$$\begin{aligned} \Pr(E_{\mathrm{IC}}) &\geq \exp\left\{ -2\log 2 \left(f(G)(2p)^2 + \frac{d}{2}(2p)^2 + \frac{f(G)d(d-1)}{2}(2p)^3 + \frac{f(G)(d-1)}{2}(2p)^2 \right) n \right\} \\ &\geq \exp\left\{ -\frac{4\log 2}{k^2 d} \left(\frac{2f(G)}{d} + 1 + \frac{2f(G)}{k} + f(G) \right) n \right\} \end{aligned}$$

Since $f(G) \leq 1$ and $k \geq 10$ (see Equation (3)), one can check that for sufficiently large d:

$$\Pr(E_{\mathrm{IC}}) \geq \exp\left\{-\frac{17\log 2}{2k^2 d}n\right\}$$

The Weighted Local Lemma shows that S has the desired properties with probability $\Pr(E_{\mathrm{IC}}) > 0$, implying that such a set exists. Note that we have no guarantee on the size of S. In fact, if $S = \emptyset$ then $V(G) \setminus S = V(G)$ is always an identifying code. Therefore we need to estimate the probability that $|S|$ is far below its expected size. In order to do this, we use the Chernoff bound of Theorem 3 by putting $a = \frac{nf(G)}{cd}$ where c is a constant to be determined. Let E_{BIG} be the event that $|S| - np > -\frac{nf(G)}{cd}$. We obtain:

$$\begin{aligned}
\Pr(\overline{E_{\mathrm{BIG}}}) &\leq \exp\left\{-\frac{\left(\frac{nf(G)}{cd}\right)^2}{2pnf(G)}\right\} \\
&= \exp\left\{-\frac{kf(G)}{2c^2 d}n\right\}
\end{aligned}$$

Now we have:

$$\begin{aligned}
\Pr(E_{\mathrm{IC}} \text{ and } E_{\mathrm{BIG}}) &= 1 - \Pr(\overline{E_{\mathrm{IC}}} \text{ or } \overline{E_{\mathrm{BIG}}}) \\
&\geq 1 - \Pr(\overline{E_{\mathrm{IC}}}) - \Pr(\overline{E_{\mathrm{BIG}}}) \\
&= 1 - (1 - \Pr(E_{\mathrm{IC}})) - \Pr(\overline{E_{\mathrm{BIG}}}) \\
&= \Pr(E_{\mathrm{IC}}) - \Pr(\overline{E_{\mathrm{BIG}}}) \\
&\geq \exp\left\{-\frac{17\log 2}{2k^2 d}n\right\} - \exp\left\{-\frac{kf(G)}{2c^2 d}n\right\}
\end{aligned}$$

Thus, $\Pr(E_{\mathrm{IC}} \text{ and } E_{\mathrm{BIG}}) > 0$ if $c < \frac{k^{3/2} f(G)^{1/2}}{\sqrt{17\log 2}}$. We arbitrarily set $c = \frac{k^{3/2} f(G)^{1/2}}{3\sqrt{2\log 2}}$.

Now we have to check that E_{BIG} implies that S is still large enough.

$$\begin{aligned}
|S| &\geq \mathbb{E}(|S|) - \frac{nf(G)}{cd} \\
&= \frac{nf(G)}{kd} - \frac{nf(G)}{cd} \\
&= \left(\frac{1}{k} - \frac{3\sqrt{2\log 2}}{k^{3/2} f(G)^{1/2}}\right)\frac{nf(G)}{d} \qquad (4)
\end{aligned}$$

Since $|S|$ must be positive, from Equation (4) we need $k^{3/2} f(G)^{1/2} > 3\sqrt{2\log 2}\, k$, which leads to $k = \frac{a_0}{f(G)}$ for $a_0 > 18\log 2$. Using this fact and derivating the expression of $|S|$, one can check that $|S|$ is optimized when $a_0 = \frac{81\log 2}{2}$.

Remark that under this condition, our assumption following Equation (3) that $k \geq 10$, is fulfilled.

Now we can see that:

$$|S| \geq \left(\frac{1}{k} - \frac{1}{c}\right)\frac{nf(G)}{d} = \frac{a_0^{1/2} - 3\sqrt{2\log 2}}{a_0^{3/2}}\frac{f(G)^2}{d}n = \frac{2}{243\log 2}\frac{f(G)^2}{d}n \geq \frac{f(G)^2}{85d}n$$

Hence finally:

$$|\mathcal{C}| \leq n - \frac{nf(G)^2}{85d}$$

□

Note that for regular graphs, $f(G) = 1$ because a forced vertex implies the existence of two vertices with distinct degrees. We obtain the following result:

Corollary 8 (Graphs with constant proportion of non-forced vertices)**.** *There exists an integer d_0 such that for each twin-free graph G on n vertices having maximum degree $d \geq d_0$ and $f(G) = \frac{1}{\alpha}$ for some constant $\alpha \geq 1$, $\gamma^{ID}(G) \leq n - \frac{n}{85\alpha^2 d}$. In particular if G is d-regular, $\gamma^{ID}(G) \leq n - \frac{n}{85d}$.*

The next proposition will be proved in the next subsection.

Proposition 9. *Let G be a graph on n vertices and of maximum degree d. Then $f(G) \geq \frac{1}{d+1}$.*

We obtain the following general result:

Corollary 10 (General case)**.** *There exists an integer d_0 such that for each twin-free graph G on n vertices having maximum degree $d \geq d_0$, $\gamma^{ID}(G) \leq n - \frac{n}{85d(d+1)^2} = n - \frac{n}{\Theta(d^3)}$.*

The next proposition will be proved in the next subsection as well.

Proposition 11. *Let G be a graph having no k-clique. Then there exists a constant $\gamma(k)$ depending only on k, such that $f(G) \geq \frac{1}{\gamma(k)}$.*

This leads to the following extension of Corollary 8, where $c(k) \leq 85\gamma(k)^2$:

Corollary 12 (Graphs with bounded clique number)**.** *There exists an integer d_0 such that for each twin-free graph G on n vertices having maximum degree $d \geq d_0$ and clique number smaller than k, $\gamma^{ID}(G) \leq n - \frac{n}{c(k)d}$ for some constant $c(k)$ depending only on k. In particular this applies to triangle-free graphs, planar graphs, or more generally, graphs of bounded genus.*

We remark here that the previous corollaries support Conjecture 1. They also lead us to think that the difficulty of the problem lies in forced vertices.

3.2 Bounding the number of non-forced vertices: proofs

In this section, we prove the lower bounds for function $f(G)$ of the statement of Theorem 7.

The following lemma was first proved in [4], and a proof can be found in [11] (as [4] is not accessible).

Lemma 13 ([4])**.** *If G is a finite twin-free graph without isolated vertices, then for every vertex u of G, there is a vertex $v \in N[u]$ such that $V(G) \setminus \{v\}$ is an identifying code of G.*

We recall the statement of Proposition 9:

Proposition. *Let G be a graph on n vertices and of maximum degree d. Then $f(G) \geq \frac{1}{d+1}$.*

Proof. Observe that a vertex v of G is not forced only if $V(G) \setminus \{v\}$ is an identifying code of G. Hence, by Lemma 13, the set S of non-forced vertices is a dominating set of G, and thus $|S| \geq \frac{n}{d+1}$. □

Note that Proposition 9 is tight. Indeed, consider the graph A_k on $2k$ vertices defined in [11] as follows: $V(A_k) = \{x_1, \ldots, x_{2k}\}$ and $E(A_k) = \{x_i x_j, |i-j| \leq k-1\}$. A_k can be seen as the $(k-1)$-th power of the path P_{2k}. In the graph A_k with an additional universal vertex x (i.e. x is adjacent to all vertices of A_k), one can check that all vertices but x are forced. This graph has $n = 2k+1$ vertices, maximum degree $2k$ and exactly $1 = \frac{n}{d+1}$ non-forced vertex. Taking all forced vertices gives a minimum identifying code of this graph.

However, note that since for a fixed value of d, we know only one such graph, it is not enough to give a counterexample to Conjecture 1. Indeed in this case $n - \frac{n}{d+1} = n - \frac{n}{d} + O(1)$. So we ask the following question:

Question 14. *For a fixed value of d and arbitrarily large values of n, do there exist graphs of maximum degree d on n vertices having exactly $\frac{n}{d+1}$ non-forced vertices?*

Answering this question in positive would provide counterexamples to Conjecture 1. Note that for the similar question where we replace $d+1$ by d, the answer is positive by Construction 23 of Section 6. For any d, this construction provides arbitrarily large graphs having $\frac{n}{d}$ non-forced vertices.

Observe that graph A_k contains two cliques of k vertices. In fact, we can improve the bound of Proposition 9 for graphs having no large cliques. Let us first introduce an auxiliary structure that will be needed in order to prove this result.

Let G be a twin-free graph. We define a partial order $\preceq$ over the set of vertices of G such that $u \preceq v$ if $N[u] \subseteq N[v]$. We construct an oriented graph $\mathcal{H}(G)$ on $V(G)$ as a subgraph of the Hasse diagram of poset $(V(G), \preceq)$. The arc set of $\mathcal{H}(G)$ is the set of all arcs $\overrightarrow{uv}$ where there exists some vertex x such that $N[v] = N[u] \cup \{x\}$. Then x is uv-forced, and we note $x = f(\overrightarrow{uv})$. For a vertex v of $V(G)$, we define the set $F(v)$ as the union of v itself and the set of all predecessors and successors of v in $\mathcal{H}(G)$.

Lemma 15. *Let G be a graph having no k-clique. Then for each vertex u, $|F(u)| \leq \beta(k)$, where $\beta(k)$ is a function depending only on k.*

Proof. First of all, we prove that the maximum in-degree of $\mathcal{H}(G)$ is at most $2k-3$, and its out-degree is at most $k-2$.

Let u be a vertex of G. Suppose u has $2k-2$ in-neighbours in $\mathcal{H}(G)$. Since for each in-neighbour v of u, $|N[u] \Delta N[v]| = 1$ in G, each of them is non-adjacent to at most one of the other in-neighbours (in the worst case the in-neighbours of u induce a clique of $2k-2$ vertices minus the edges of a perfect matching). Hence they induce a clique of size at least $k-1$. Together with vertex u, they form a k-clique in G, a contradiction.

Now suppose u has $k-1$ out-neighbours in $\mathcal{H}(G)$. Since for each out-neighbour v of u, $N[u] \subseteq N[v]$, u and its out-neighbours form a k-clique, a contradiction.

Now, consider the subgraph of $\mathcal{H}(G)$ induced by $F(u)$. We claim that the longest directed chain in this subgraph has at most $k-1$ vertices. Indeed, all the vertices of such a chain are pairwise adjacent, hence otherwise they would induce a k-clique.

Finally, we obtain that $F(u)$ has size at most $\beta(k) = \sum_{i=0}^{k-2}(2k-3)^i$ and the claim of the lemma follows. □

We now need to prove a few additional claims regarding the structure of $\mathcal{H}(G)$. In the following claims, we suppose that G is a twin-free graph.

Claim A. *Let s be a forced vertex in G with $s = f(\overrightarrow{uv})$ for some vertices u and v. If t is an in-neighbour of s in $\mathcal{H}(G)$, then $v = f(\overrightarrow{ts})$. Moreover if v is forced with $v = f(\overrightarrow{xy})$, then necessarily $y = s$.*

Proof. For the first implication, suppose s has an in-neighbour t in $\mathcal{H}(G)$. An illustration is provided in Figure 2. Since $u \not\sim s$, $u \not\sim t$. Moreover $v \not\sim t$ since $s = f(\overrightarrow{uv})$. Since $s \sim v$ the claim follows. For the other implication, suppose there exist two vertices x, y such that $v = f(\overrightarrow{xy})$. Hence $y \sim v$ but $x \not\sim v$. Therefore $u \not\sim x$ (otherwise v would be adjacent to x too) and hence $u \not\sim y$. Now the only vertex adjacent to v but not to u is s, so $y = s$. □

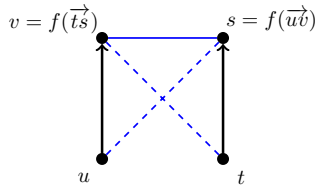


Figure 2: The situation of Claim A. Black arcs belong to $\mathcal{H}(G)$. Blue full edges belong to G only, blue dashed edges are non-edges in G.

Claim B. *Let s be a forced vertex in G with $s = f(\overrightarrow{uv})$ for some vertices u and v. Then s has a unique in-neighbour in $\mathcal{H}(G)$.*

Proof. Suppose s has two distinct in-neighbours t and t' in $\mathcal{H}(G)$ (see Figure 3 for an illustration). By Claim A, v is both ts-forced and $t's$-forced. But then $N[t] = N[s] \setminus \{v\} = N[t']$. Then t and t' are twins, a contradiction since G is twin-free. □

Claim C. *Let s be a forced vertex in G with $s = f(\overrightarrow{uv})$, and let t be a forced in-neighbour of s in $\mathcal{H}(G)$ with $t = f(\overrightarrow{xy})$ for some vertices u, v, x, y. Then $x = v$.*

Proof. Since $t \sim y$, $s \sim y$ too. But since $t = f(\overrightarrow{xy})$, $x \sim s$ and $x \not\sim t$. Now by Claim A, $v = f(\overrightarrow{ts})$, that is, v is the unique vertex such that v is adjacent to s, but not to t. Therefore $x = v$. □

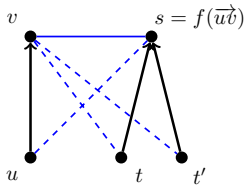


Figure 3: The situation of Claim B. Black arcs belong to $\mathcal{H}(G)$. Blue full edges belong to G only, blue dashed edges are non-edges in G.

We now obtain the following lemma using the previous claims.

Lemma 16. *Let s be a non-isolated sink in $\mathcal{H}(G)$ which is forced in G with $s = f(\overrightarrow{uv})$ for some vertices u and v. Then either s has a non-forced predecessor t in $\mathcal{H}(G)$ such that $F(s) \subseteq F(t)$, or there exists a non-forced vertex $w(s)$ such that $F(s) \subseteq N_G[w(s)]$. Moreover, if there are ℓ additional sinks $\{s_1, \ldots, s_\ell\}$ such that $w(s) = w(s_1) = \ldots = w(s_\ell)$, then there exists a set of $\ell + 1$ distinct vertices inducing a clique together with $w(s)$.*

Proof. Suppose s has a non-forced predecessor in $\mathcal{H}(G)$ and let t be the one having shortest distance to s in $\mathcal{H}(G)$. By Claim B, s, t and the predecessors of s which are successors of t, induce a directed path. Hence $F(s) \subseteq F(t)$, which proves the first part of the statement.

Now, suppose all predecessors of s are forced. By Claim B, s and its predecessors form a directed path $\{t_0, \ldots, t_m, s\}$ in $\mathcal{H}(G)$ (for an illustration, see Figure 4(a)). Note that by Claim A, there exists a forced vertex $v = f(\overrightarrow{t_m s})$. By our assumption we know that t_m is forced, say $t_m = f(\overrightarrow{xv_m})$ for some vertices x and v_m. But now by Claim C, $x = v$ and $t_m = f(\overrightarrow{vv_m})$. Now, repeating these arguments for each other predecessor of s shows that there is a directed path $\{u, v, v_m, \ldots, v_0\}$ with $t_m = f(\overrightarrow{vv_m})$ and for all i, $0 \leq i \leq m-1$, $t_i = f(\overrightarrow{v_{i+1}v_i})$. In particular, $t_0 = f(\overrightarrow{v_1 v_0})$. By applying Claim C on vertices v_1, v_0 and t_0, if v_0 is forced then t_0 has an in-neighbour in $\mathcal{H}(G)$, a contradiction — hence v_0 is non-forced. Moreover note that since $v_0 \sim t_0$, v_0 is adjacent to all successors of t_0 in $\mathcal{H}(G)$, that is, to all elements of $F(s)$. Therefore, putting $w(s) = v_0$, we obtain the second part of the statement.

For the last part, suppose there exists a set of forced sinks $\{s_1, \ldots, s_\ell\}$ such that all their predecessors in $\mathcal{H}(G)$ are forced and $w(s_i) = v_0$ for $1 \leq i \leq \ell$ (for an illustration, see Figure 4(b)). For each such sink s_i, by the previous paragraph, the vertices of $F(s_i)$ induce a directed path $\{t_0^i, \ldots, t_{m_i}^i, s_i\}$ in $\mathcal{H}(G)$. Moreover we know that there is a vertex x_i such that t_0^i is $x_i v_0$-forced. We claim that the set of vertices $X = \{x_1, \ldots, x_\ell\}$ forms a clique of $\ell + 2$ vertices together with v_1 and v_0.

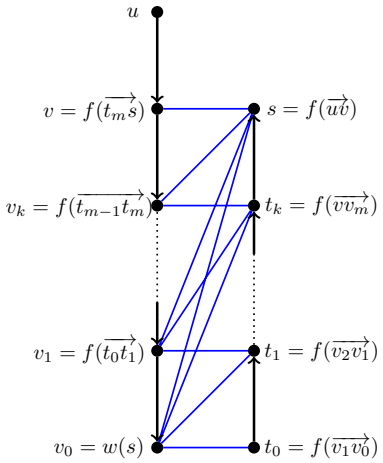
Now, observe that the vertices of X must all be pairwise adjacent since they are all adjacent to v_0, and for each x_i, $N[v_0] = N[x_i] \cup \{t_0^i\}$. For the same reason they are all adjacent to v_1.

Finally, let us show that all the vertices of X are distinct: by contradiction, suppose that $x_i = x_j$ for some $i \neq j$, $1 \leq i, j \leq \ell$. Since t_0^i is $x_i v_0$-forced and t_0^j is $x_j v_0$-forced, we have $t_0^i = t_0^j$. Since s_i and s_j are distinct, this means that s_i and s_j have one predecessor in common. Hence their common predecessor which is nearest to s_i and s_j, say t, has two out-neighbours. Let t_i (respectively t_j) be the out-neighbour of t which is a predecessor of s_i (respectively s_j) — see Figure 4(c) for an illustration. We know that there are two vertices y_i, y_j such that $y_i = f(\overrightarrow{tt_i})$ and $y_j = f(\overrightarrow{tt_j})$. First note that y_i and y_j are distinct: otherwise, we would have $N[t_i] = N[t] \cup \{y_i\} = N[t] \cup \{y_j\} = N[t_j]$ and then t_i, t_j would be twins in G. Observe that since $t \not\sim y_i$ and $y_i \neq f(\overrightarrow{tt_j})$, we have $t_j \not\sim y_i$. We know that t is forced, in fact by the first part of this proof, we also know that $t = f(\overrightarrow{y_i z_i})$ for some vertex z_i. Hence $z_i \sim t$, and since $N[t] \subseteq N[t_j]$, $z_i \sim t_j$. But since $t_j \neq f(\overrightarrow{y_i z_i})$, $t_j \sim y_i$, a contradiction. Hence x_i and x_j are distinct, which completes the proof. □

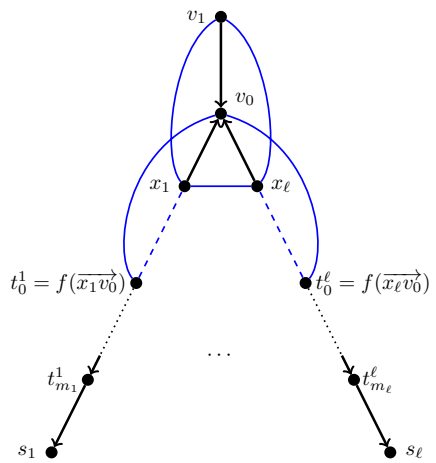
Finally, let us recall and prove Proposition 11.

Proposition. *Let G be a graph having no k-clique. Then there exists a constant $\gamma(k)$ depending only on k, such that $f(G) \geq \frac{1}{\gamma(k)}$.*

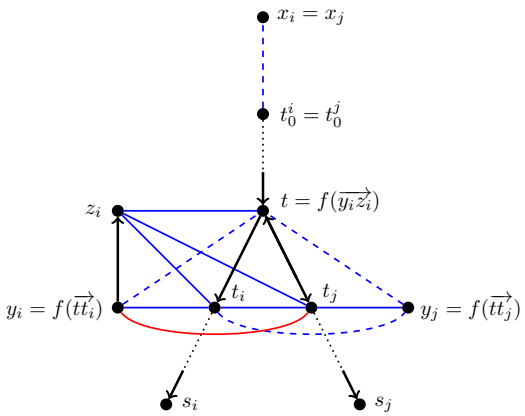
Proof. To prove the result, we use $\mathcal{H}(G)$ to construct a set $X = \{x_1, \ldots, x_\ell\}$ of non-forced vertices such that $\bigcup_{i=1}^{\ell} A(x_i) = V(G)$, where $A(x_i)$ is a set of at most $\gamma(k)$ vertices. Then we have $\ell \geq \frac{n}{\gamma(k)}$ vertices in X and the claim of the proposition follows.



(a) Vertex s and all its predecessors in $\mathcal{H}(G)$ are forced



(b) Vertices $v_0, v_1, x_1, \ldots, x_\ell$ induce a clique in G



(c) If $x_i = x_j$, the red edge is both an edge and a non-edge of G

Figure 4: Three situations in the proof of Lemma 16. Black arcs belong to $\mathcal{H}(G)$. Blue edges belong to G only, blue dashed edges are non-edges in G.

We now describe a procedure to build set X while considering each non-isolated sink of $\mathcal{H}(G)$. We denote by s the currently considered sink.

Case 1: Sink s is non-forced. Then we set $A(s)$ to be $F(s)$ together with all the vertices which are forced by a pair u, v of vertices of $F(s)$. Note that by Lemma 15, $|F(s)| \leq \beta(k)$, where $\beta(k)$ only depends on k. Hence, $|A(s)| \leq \beta(k) + \binom{\beta(k)}{2}$.

Case 2: Sink s is forced. By Lemma 16, either s has a non-forced predecessor t such that $F(s) \subseteq F(t)$, or there exists a non-forced vertex $w(s)$ such that $F(s) \subseteq N_G[w]$.

In the first case, we choose t as our non-forced vertex, and we set $A(t)$ to be $F(t)$ together with all the vertices which are forced by a pair u, v of vertices of $F(t)$. Again we have $|A(t)| \leq \beta(k) + \binom{\beta(k)}{2}$.

In the second case, we choose $w = w(s)$ as our non-forced vertex. Now, let $S = \{s, s_1, \ldots, s_\ell\}$ be the set of forced sinks having no non-forced predecessor and such that $w(s) = w(s_1) = \ldots w(s_\ell)$. By Lemma 16 we know that there are $\ell+1$ distinct vertices inducing a clique together with w, hence $\ell+2 < k$. We set $A(w)$ to be $F(w) \cup F(s) \cup F(s_1) \cup \ldots \cup F(s_\ell)$ together with all the vertices which are forced by a pair u, v of vertices of this set. We have $|A(w)| \leq k\beta(k) + k\binom{\beta(k)}{2}$.

We have now covered all the vertices which are not isolated in $\mathcal{H}(G)$, since for each non-isolated sink s of $\mathcal{H}(G)$, $F(s)$ is a subset of $A(x)$ for some $x \in X$. Moreover all isolated vertices of $\mathcal{H}(G)$ which are forced, have also been put into some set $A(x)$. Hence only non-forced isolated vertices of $\mathcal{H}(G)$ need to be covered. For each such vertex v, we add v to X and set $A(v) = \{v\}$.

Finally, all vertices belong to some set $A(x)$, $x \in X$, and the size of each set $A(x)$ is at most $\gamma(k) = k\beta(k) + k\binom{\beta(k)}{2}$, which completes the proof. □

4 Upper bounds for graphs with girth at least 5

This section is devoted to the study of graphs that have girth at least 5. We will use these results in Section 5, which deals with random regular graphs.

Despite being different than our previous proofs, the ones of this section have also a probabilistic flavour. One can check that for graphs of girth 5, applying the Local Lemma does not lead to a satisfying result. However, by using the Alteration method, a better bound can be given.

We start by defining an auxiliary notion that will be used in this section. A subset $D \subseteq V(G)$ is called a 2-*dominating set* if for each vertex v of $V(G) \setminus D$, $|N(v) \cap D| \geq 2$ [9]. The next lemma shows that we can use a 2-dominating set to construct an identifying code.

Lemma 17. *Let G be a twin-free graph on n vertices having girth at least 5. Let D be a 2-dominating set of G. If the subgraph induced by D, $G[D]$, has no isolated edge, D is an identifying code of G.*

Proof. First observe that D is dominating since it is 2-dominating. Let us check that D is also separating.

Note that all the vertices that do not belong to D are separated because they are dominated at least twice each and $g(G) > 4$.

Similarly, a vertex $x \in D$ and a vertex $y \in V(G) \setminus D$ are separated since y has two vertices which dominate it, but they cannot both dominate x (otherwise there would be a triangle or a 4-cycle in G).

Finally, consider two vertices of D. If they are not adjacent they are separated by themselves. Otherwise, by the assumption that $G[D]$ has no isolated edge and that G has no triangles, we know that at least one of them has a neighbour in D, which separates them since it is not a neighbour of the other. □

The following theorem makes use of Lemma 17. The idea of the proof is inspired by a classic proof of a result on dominating sets which can be found in the first chapter of [1].

Theorem 18. *Let G be a graph on n vertices with minimum degree δ and girth at least 5. Then $\gamma^{ID}(G) \leq (1+o_\delta(1))\frac{3\log\delta}{2\delta}n$. Moreover if G has average degree $\overline{d} = O_\delta(\delta(\log\delta)^2)$ then $\gamma^{ID}(G) \leq \frac{\log\delta+\log\log\delta+O_\delta(1)}{\delta}n$.*

Proof. Let $S \subseteq V(G)$ be a random subset of vertices, where each vertex $v \in V(G)$ is added to S uniformly at random with probability p (where p will be determined later). For every vertex $v \in V(G)$, we define the random variable X_v as follows:

$$X_v = \begin{cases} 0 & \text{if } |N[v] \cap S| \geq 2 \\ 1 & \text{otherwise} \end{cases}$$

Let T be the set of vertices which are not 2-dominated, i.e. $T = \{v \mid X_v = 1\}$. Note that $|T| = \sum X_v$. Let us estimate the size of T. Observing that $|N[v] \cap S| \sim \text{Bin}(\delta + 1, p)$ we obtain:

$$\begin{aligned}
\mathbb{E}(|T|) &= \sum_{v \in V(G)} \mathbb{E}(X_v) \\
&= n\left((1-p)^{\delta+1} + (\delta+1)p(1-p)^{\delta}\right) \\
&= n(1-p)^{\delta}((1-p) + (\delta+1)p) \\
&\leq n(1+\delta p)e^{-\delta p}.
\end{aligned}$$

where we have used the fact that $1 - x \leq e^{-x}$. Now, note that the set $D = S \cup T$ is a 2-dominating set of G. We have $|D| \leq |S| + |T|$. Hence

$$\begin{aligned}
\mathbb{E}(|D|) &\leq \mathbb{E}(|S|) + \mathbb{E}(|T|) \\
&\leq np + n(1+\delta p)e^{-\delta p} \qquad (5)
\end{aligned}$$

Let us set $p = \frac{\log\delta + \log\log\delta}{\delta}$. Plugging this into Equation (5), we obtain:

$$\mathbb{E}(|D|) \leq \frac{\log\delta + \log\log\delta}{\delta} n + \frac{1 + \log\delta + \log\log\delta}{\delta \log\delta} n = \frac{\log\delta + \log\log\delta + O_\delta(1)}{\delta} n$$

This shows that there exists at least one 2-dominating set D having this size.

Case 1: (general case) Note that we can use Lemma 17 by considering all pairs u, v of vertices of D forming an isolated edge in $G[D]$, and add an arbitrary neighbour of either one of them to D. Observe that such a vertex exists, otherwise u and v would be twins in G. Since there are at most $\frac{|D|}{2}$ such pairs, we obtain a 2-dominating set of size at most $|D| + \frac{|D|}{2} = (1 + o_\delta(1))\frac{3\log\delta}{2\delta} n$ having the desired property. Now applying Lemma 17 completes Case 1.

Case 2: (sparse case) Whenever $\overline{d} = O_\delta(\delta(\log\delta)^2)$, we can get a better bound by estimating the number of isolated edges of $G[D]$. For each edge uv of G, let Y_{uv} denote the random variable defined as follows:

$$Y_{uv} = \begin{cases} 1 & \text{if } uv \text{ is isolated in } G[D] \\ 0 & \text{otherwise} \end{cases}$$

Note that $Y_{uv} = 1$ when $N[u] \Delta N[v] \subseteq V(G) \setminus S$. Using the facts that $p = \frac{\log\delta + \log\log\delta}{\delta}$ and $1 - x \leq e^{-x}$, let us calculate the expected value of $Y = \sum_{uv \in E} Y_{uv}$.

$$\mathbb{E}(Y) = \sum_{uv \in E(G)} \mathbb{E}(Y_{uv}) \leq \frac{n\overline{d}}{2}(1-p)^{2\delta-2}p^2 \leq \frac{n\overline{d}}{2} e^{-(2\delta-2)p} = \frac{n\overline{d}e^{-2(\log\delta + \log\log\delta)}}{2} = \frac{n\overline{d}}{2\delta^2(\log\delta)^2}$$

We construct U by picking an arbitrary neighbour of either u or v for each edge uv such that $Y_{uv} = 1$. We have $|U| \leq Y$. The final set $\mathcal{C} = S \cup T \cup U$ is an identifying code. Now we have:

$$\mathbb{E}(|\mathcal{C}|) \leq \mathbb{E}(|S|) + \mathbb{E}(|T|) + \mathbb{E}(|U|) \leq \frac{\log\delta + \log\log\delta + O_\delta(1)}{\delta} n + \frac{\overline{d}}{2\delta^2(\log\delta)^2} n$$

Using that $\overline{d} = O_\delta(\delta(\log\delta)^2)$,

$$\mathbb{E}(|\mathcal{C}|) \leq \frac{\log\delta + \log\log\delta + O_\delta(1)}{\delta} n$$

Then there exists some choice of S such that $|\mathcal{C}|$ has the desired size, and completes the proof. □

In fact, it is shown in the next section (Corollary 22) that Theorem 18 is asymptotically tight.

Moreover, note that Theorem 18 cannot be extended much in the sense that if we drop the condition on girth 5, we know arbitrarily large d-regular triangle-free graphs having large minimum identifying codes. For instance, Construction 25 of Section 6 provides a graph G which satisfies $\gamma^{\text{ID}}(G) = n - \frac{n}{d}$. Similarly, we cannot drop the minimum degree condition. Indeed it is known that any $(d-1)$-ary complete tree $T_{d,h}$ of height h, which is of maximum degree d, minimum degree 1 and has infinite girth, also has a large identifying code number (i.e. $\gamma^{\text{ID}}(T_{d,h}) = n - \frac{n}{d-1+o_d(1)}$ [5]).

5 Identifying codes of random regular graphs

From the study of regular graphs arises the question of the value of the identifying code number for most regular graphs. We know some lower and upper bounds for this parameter, but is it concentrated around some value? A good way to study this question is to look at random regular graphs.

Consider the Configuration Model, where a d-regular multigraph on n vertices is obtained by selecting some perfect matching of K_{nd} at random (see [7] for further reference). We will denote by $\mathcal{G}^*(n,d)$ the former probability space and by $\mathcal{G}(n,d)$ the same probability space conditioned on the event that G is simple. It is well-known (see e.g. [7]) that the following holds:

$$\Pr\left(G \in \mathcal{G}(n,d) \mid G \in \mathcal{G}^*(n,d)\right) = e^{\frac{1-d^2}{4}}$$

Then any property that holds with probability tending to 1 for $\mathcal{G}^*(n,d)$ as $n \to \infty$, will also hold with probability tending to 1 for $\mathcal{G}(n,d)$. In this case we will say that the property holds *with high probability* (w.h.p.). In fact our bounds include asymptotic terms in d, which means they are meaningful for sufficiently large d.

Theorem 19. *Let $G \in \mathcal{G}(n,d)$ then w.h.p. $\gamma^{ID}(G) \leq \frac{\log d + \log\log d + O_d(1)}{d} n$.*

Proof. First of all we have to show that almost all random regular graphs are twin-free.

Observe that the number of perfect matchings of K_{2m} is $(2m-1)!! = (2m-1)(2m-3)(2m-5)\ldots 1$. Fix a vertex u of G and let $N(u) = \{v_1, \ldots, v_d\}$. We compute the probability that u and v_1 are twins, i.e. $N[u] = N[v_1]$. The number of perfect matchings of K_{nd} such that in the resulting graph G of $\mathcal{G}(n,d)$, v_1 and v_2 are adjacent, is at most $(d-1)(d-1)(nd-2d-3)!!$. Indeed, there must be an edge between v_1 and v_2, which gives $(d-1)(d-1)$ possibilities. Since u has d neighbours, the number of possibilities for the remaining graph is the number of perfect matchings of $K_{nd-2d-2}$.

Analogously the number of perfect matchings with $v_2, v_3 \in N(v_1)$ is at most $(d-1)(d-1)(d-2)(d-1)(nd-2d-5)!!$. Thus we have:

$$\begin{aligned}
\Pr(N[u] = N[v_1]) &= \frac{(d-1)(d-1)(d-2)(d-1)\ldots 2(d-1)1(d-1)(nd-4d+1)!!}{(nd-2d-1)!!} \\
&\leq \frac{d^{d-1}(d-1)!}{(nd-2d-1)\ldots(nd-4d+3)} \\
&\leq \left(\frac{d}{n}\right)^{d-1}
\end{aligned}$$

As we have at most $\frac{nd}{2}$ possible pairs of twins (one for each edge), by the union bound and since $d \geq 3$, we obtain:

$$\Pr(G \text{ has twins}) \leq \frac{nd}{2}\left(\frac{d}{n}\right)^{d-1} \longrightarrow 0$$

Therefore, random regular graphs are twin-free w.h.p.

By the proof of Theorem 18, for any $G \in \mathcal{G}(n,d)$, we have a set $\mathcal{C}$ with

$$|\mathcal{C}| \leq \frac{\log d + \log\log d + O_d(1)}{d} n$$

that separates any pair of vertices except from the ones where both vertices belong to a triangle or a 4-cycle. We have to add some vertices to $\mathcal{C}$ in order to separate the vertices of these small cycles.

Classical results on random regular graphs (independently, [6] and [25]) state that the random variables that count the number of cycles of length k, X_k, tend in distribution to independent Poisson variables with parameter $\lambda_k = \frac{1}{2k}(d-1)^k$.

Observe that:

$$\mathbb{E}(X_3) = \frac{(d-1)^3}{6} \quad \mathbb{E}(X_4) = \frac{(d-1)^4}{8}$$

i.e. a constant number of triangles and 4-cycles are expected.

Using Markov's inequality we can bound the probability of having too many small cycles:

$$\Pr(X_3 > t) \leq \frac{(d-1)^3}{6t} \qquad \Pr(X_4 > t) \leq \frac{(d-1)^4}{8t}$$

Setting $t = \vartheta(n)$, where $\vartheta(n) \to \infty$, the previous probabilities are $o(1)$. Then w.h.p., we have at most $\vartheta(n)$ cycles of length 3 and $\vartheta(n)$ cycles of length 4.

Let $T = \{u_1, u_2, u_3\}$ be a triangle in G. As $d \geq 3$ there exists at least one vertex v_i outside the triangle. Since our graph is twin-free, for each ordered pair (u_i, u_j) there exists some vertex v_{ij}, such that $v_{ij} \in N(u_i) \backslash N(u_j)$. Observe that we can add v_{12}, v_{23} and v_{31} to $\mathcal{C}$ and then any pair of vertices from T will be separated.

If $T = \{u_1, u_2, u_3, u_4\}$ induces a K_4, each pair of vertices of T is contained in some triangle and is separated by the last step. If T induces a 4-cycle, adding T to $\mathcal{C}$ separates all the elements in T. Otherwise, T induces two triangles and adding T to $\mathcal{C}$ separates the two vertices which have not been separated in the last step.

After these two steps, we have added at most $7\vartheta(n)$ vertices to $\mathcal{C}$. Hence, for any $G \in \mathcal{G}(n,d)$ w.h.p. we obtain:

$$\gamma^{\text{ID}}(G) \leq \frac{\log d + \log\log d + O_d(1)}{d} n + 7\vartheta(n) = \frac{\log d + \log\log d + O_d(1)}{d} n$$

Observe that the $\frac{O_d(1)}{d} n$ term contains the $7\vartheta(n)$ term. □

Theorem 19 shows that despite the fact that for any d, we know infinitely many d-regular graphs having a very large identifying code number (e.g. $n - \frac{n}{d}$ for the graphs of Construction 24 of Section 6), almost all d-regular graphs have a very small identifying code.

Moreover, $\gamma^{\text{ID}}(G)$ is concentrated, as the following theorem and its corollary show. In fact the following result might be already known, since a similar result is stated for independent dominating sets in [16]. However we could not find it in the literature and decided to give a proof for the sake of completeness.

Theorem 20. *Let $G \in \mathcal{G}(n,d)$, then w.h.p. all the dominating sets of G have size at least $\frac{\log d - 2\log\log d}{d} n$.*

Proof. We will proceed by contradiction. Given a set of vertices D of size m, we will compute the probability that D dominates $Y = V(G) \setminus D$. Recall that G has been obtained from the configuration model by selecting a random perfect matching of K_{nd}. Let $y \in Y$ fixed, then let $A_y = \{N(D) \cap y \neq \emptyset\}$ be the event that y is dominated by D. Its complementary event corresponds to the situation where none of the edges of the perfect matching of K_{nd} connects the points corresponding to y to the ones corresponding to any vertex of D. Hence:

$$\begin{aligned}
\Pr(\overline{A_y}) &= \left(1 - \frac{d}{nd-1}\right)\left(1 - \frac{d}{nd-3}\right) \cdots \left(1 - \frac{d}{nd-(2md-1)}\right) \\
&= \prod_{i=1}^{md} \left(1 - \frac{d}{nd-(2i-1)}\right) \\
&\geq \prod_{i=1}^{md} \left(1 - \frac{1}{n-2m}\right)
\end{aligned}$$

Since $1 - x = e^{-x+(\log(1-x)+x)}$ (here we take $x = \frac{1}{n-2m}$) and $\log(1-x) + x = O(x^2)$ (by the Taylor expansion of the logarithm in $x = 0$), we obtain:

$$\begin{aligned}
\Pr(\overline{A_y}) &\geq \exp\left\{-\sum_{i=1}^{md} \frac{1}{n-2m} + O\left(\frac{1}{(n-2m)^2}\right)\right\} \\
&= \exp\left\{-(1+o(1))\frac{md}{n-2m}\right\}
\end{aligned}$$

The probability that D is dominating all vertices of $Y = \{y_1, \ldots, y_{n-m}\}$ is:

$$\Pr\left(\cap_{y \in Y} A_y\right) = \Pr\left(A_{y_1}\right) \Pr\left(A_{y_2} \mid A_{y_1}\right) \ldots \Pr\left(A_{y_{n-m}} \mid \cap_{j=1}^{n-m-1} A_{y_j}\right)$$

We claim that $\Pr\left(A_{y_i} \mid \cap_{j=1}^{i-1} A_{y_j}\right) \leq \Pr\left(A_{y_i}\right)$. Suppose that $y_1, \ldots, y_{i-1}$ are dominated. This means that the corresponding perfect matching of K_{nd} has an edge between one of the points corresponding to y_j $(1 \leq j \leq i-1)$ and one of the points corresponding to the vertices of D. The probability that y_i is not dominated by D is now the probability that none of the remaining edges of the perfect matching connect any vertex of D with y_i. Hence:

$$\begin{aligned}
\Pr\left(\overline{A_{y_i}} \mid \cap_{j=1}^{i-1} A_{y_j}\right) &= \left(1 - \frac{d}{nd - 2(i-1) - 1}\right)\left(1 - \frac{d}{nd - 2(i-1) - 3}\right) \cdots \left(1 - \frac{d}{nd - 2md + 1}\right) \\
&\geq \left(1 - \frac{d}{nd-1}\right)\left(1 - \frac{d}{nd-3}\right) \cdots \left(1 - \frac{d}{nd - 2md + 1}\right) \\
&= \Pr(\overline{A_{y_i}})
\end{aligned}$$

By considering the complementary events, $\Pr\left(A_{y_i} \mid \cap_{j=0}^{i-1} A_{y_j}\right) \leq \Pr\left(A_{y_i}\right)$. Hence these events are negatively correlated, and:

$$\Pr\left(\cap_{y \in Y} A_y\right) \leq \prod_{i=1}^{n-m} \Pr(A_{y_i}) \leq \left(1 - e^{-\frac{md}{n-2m}}\right)^{n-m} \leq \exp\left\{-(n-m)e^{-\frac{md}{n-2m}}\right\}$$

For the sake of contradiction, let $m \leq \frac{\log d - c \log\log d}{d} n$ for some $c > 2$. Then:

$$\begin{aligned}
\Pr\left(\cap_{y \in Y} A_y\right) &\leq \exp\left\{-\left(1 - \frac{\log d - c\log\log d}{d}\right) n \exp\left\{-\frac{\log d - c\log\log d}{1 - 2\frac{\log d - c\log\log d}{d}}\right\}\right\} \\
&= \exp\left\{-(1 + o_d(1))\, n \exp\left\{-\frac{\log d - c\log\log d}{1 + o_d(1)}\right\}\right\} \\
&= (1 + o_d(1)) e^{-\frac{(\log d)^c}{d} n}
\end{aligned}$$

Note that if no set of size m dominates Y, neither will do a smaller one. So we have to look just at the sets of size m. The number of these sets can be bounded by

$$\binom{n}{m} \leq \frac{n^m}{m!} \leq \left(\frac{en}{m}\right)^m = \left(\frac{de}{\log d - c\log\log d}\right)^{\frac{\log d - c\log\log d}{d} n} = (1 + o_d(1)) \left(\frac{de}{\log d}\right)^{\frac{\log d - c\log\log d}{d} n}$$

where we have used $m! \geq \left(\frac{m}{e}\right)^m$.

Let E_{DS} be the event that G has a dominating set of size m. Applying the union bound, we obtain:

$$\begin{aligned}
\Pr(E_{DS}) &\leq (1 + o_d(1)) \left(\frac{de}{\log d}\right)^{\frac{\log d - c\log\log d}{d} n} e^{-\frac{(\log d)^c}{d} n} \\
&= (1 + o_d(1)) \exp\left\{\frac{\log d - c\log\log d}{d} (\log d + 1 - \log\log d) n - \frac{(\log d)^c}{d} n\right\} \\
&= (1 + o_d(1)) \exp\left\{\left(\frac{(\log d)^2}{d} - \frac{(\log d)^c}{d} + o_d\left(\frac{(\log d)^2}{d}\right)\right) n\right\} \longrightarrow 0
\end{aligned}$$

since $c > 2$. This shows that w.h.p. no set of size less than $\frac{\log d - 2\log\log d}{d} n$ can dominate the whole graph and completes the proof. □

Since any identifying code is also a dominating set, we obtain the following immediate corollary.

Corollary 21. *Let $G \in \mathcal{G}(n,d)$, then w.h.p. $\gamma^{ID}(G) \geq \frac{\log d - 2\log\log d}{d} n$.*

Plugging together Theorems 19 and 20, we obtain the following result.

Corollary 22. *Let $G \in \mathcal{G}(n,d)$, then w.h.p.*

$$\frac{\log d - 2\log\log d}{d} n \leq \gamma^{ID}(G) \leq \frac{\log d + \log\log d + O_d(1)}{d} n$$

6 Extremal constructions

This section gathers some constructions which show the tightness of some of our upper bounds. Some of these constructions can be found in [10].

Construction 23. *Given any d_H-regular multigraph H (without loops) on n_H vertices, let $\mathcal{C}_1(H)$ be the graph on $n = n_H(d_H + 1)$ and maximum degree $d = d_H + 1$ constructed as follows:*

1. *Replace each vertex v of H by a clique $K(v)$ of $d_H + 1$ vertices*
2. *For each vertex v of H, let $N(v) = \{v_1, \dots, v_{d_H}\}$ and $K(v) = \{k_0(v), \dots, k_{d_H}(v)\}$. For each $k_i(v)$ but one ($1 \leq i \leq d_H$), connect it with an edge in $\mathcal{C}_1(H)$, to a unique vertex of $K(v_i)$, denoted $f(k_i(v))$.*

One can see that the graphs $\mathcal{C}_1(H)$ given by Construction 23 are twin-free. Moreover, for each vertex v of H and for each $1 \leq i \leq d_H$, note that $f(k_i(v))$ is $k_0(v)k_i(v)$-forced. Therefore $\mathcal{C}_1(H)$ has $d_H n_H = n - \frac{n}{d}$ forced vertices. In fact these forced vertices form an identifying code, therefore $\gamma^{\text{ID}}(\mathcal{C}_1(H)) = n - \frac{n}{d}$. An example of this construction is given in Figure 5, where H is the hypercube of dimension 3, H_3, and the black vertices are those which belong to a minimum identifying code of $\mathcal{C}_1(H_3)$.

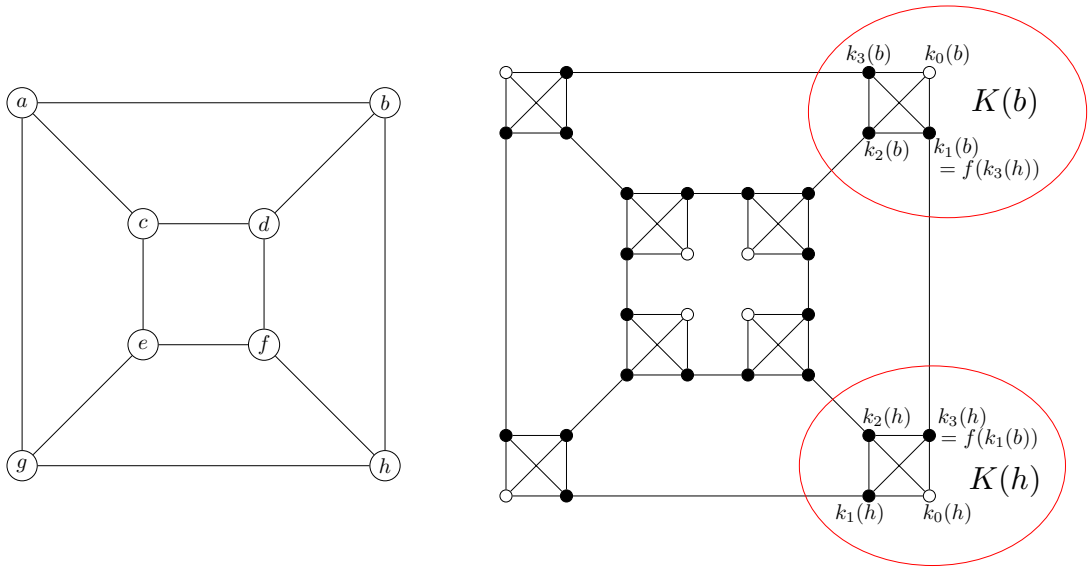


Figure 5: The graphs H_3 and $\mathcal{C}_1(H_3)$

The following construction is very similar, but yields regular graphs.

Construction 24. *[10] Given any d_H-regular multigraph H (without loops) on n_H vertices, let $\mathcal{C}_2(H)$ be the d-regular graph on $n = n_H d_H$ vertices (where $d = d_H$) constructed as follows:*

1. *Replace each vertex v of H by a clique $K(v)$ of d_H vertices.*
2. *For each vertex v of H, let $N(v) = \{v_1, \dots, v_{d_H}\}$ and $K(v) = \{k_1(v), \dots, k_{d_H}(v)\}$. For each $k_i(v)$ ($1 \leq i \leq d_H$), connect it with an edge in $\mathcal{C}_2(H)$, to a unique vertex of $K(v_i)$, denoted $f(k_i(v))$.*

Note that for some vertex v of H, in order to separate each pair of vertices $k_i(v), k_j(v)$ of $K(v)$ in $\mathcal{C}_2(H)$, either $f(k_i(v))$ or $f(k_j(v))$ must belong to any identifying code. Repeating this argument for each pair shows that at least $d - 1$ such vertices are needed in the code. Since for any two cliques $K(u)$ and $K(v)$, the set of these neighbours are disjoint, this shows that at least $n_H(d-1)$ vertices are needed in an identifying code of $\mathcal{C}_2(H)$. In fact it is easy to construct an identifying code of this size. This shows that despite the fact that $\mathcal{C}_2(H)$ has no forced vertices, $\gamma^{\text{ID}}(\mathcal{C}_2(H)) = n - \frac{n}{d}$. An example of this construction is given in Figure 6, where H is the complete graph K_5, and the black vertices form a minimum identifying code of $\mathcal{C}_2(K_5)$.

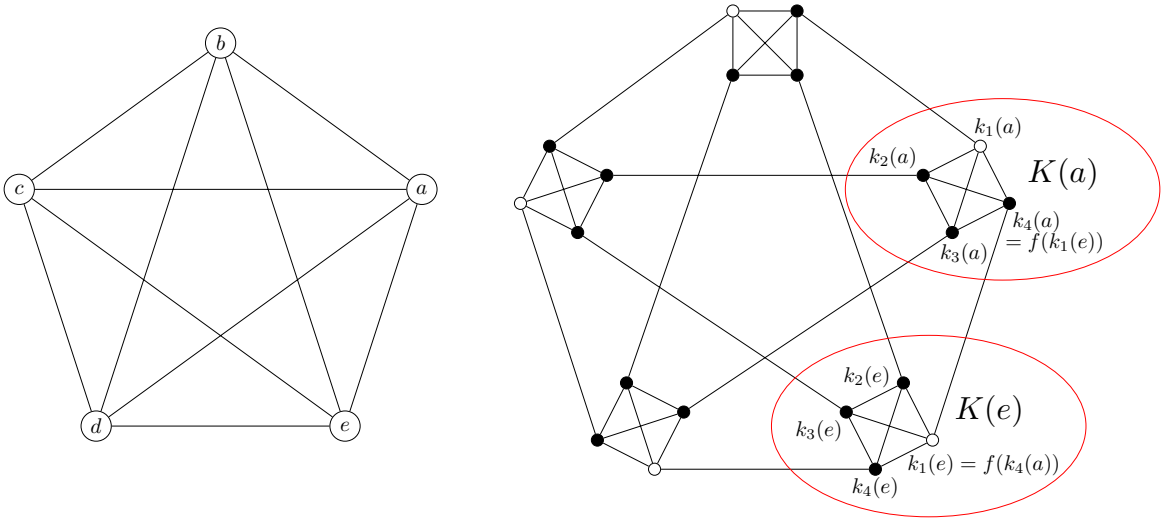


Figure 6: The graphs K_5 and $\mathcal{C}_2(K_5)$

Construction 25. *[10] Given an even number $2k$ and an integer $d \geq 3$, we construct a twin-free d-regular triangle-free graph $\mathcal{C}_3(2k, d)$ on $n = 2kd$ vertices as follows.*

1. *Let $\{c_0, \ldots, c_{2k-1}\}$ be a set of $2k$ vertices and add the edges of the perfect matching $\{c_i c_{i+1 \bmod 2k} \mid i \text{ is odd}\}$.*
2. *For each even i $(0 \leq i \leq 2k-2)$, build a copy $K(i)$ of the complete bipartite graph $K_{d-1,d-1}$. Join vertex c_i to all vertices of one part of the bipartition of $K(i)$, and join vertex c_{i+1} to all other vertices of $K(i)$.*

Consider an identifying code of $\mathcal{C}_3(2k, d)$. Note that in each copy $K(i)$ of $K_{d-1,d-1}$, at least $2d-4$ vertices belong to the code in order to separate the vertices being in the same part of the bipartition of $K(i)$. Now if exactly $2d-4$ vertices of $K(i)$ belong to the code, in order to separate the two remaining vertices, either c_i or c_{i+1} belongs to the code. Hence for each odd i, at most three vertices from $\{c_i, c_{i+1}\} \cup V(K(i))$ do not belong to a code of $\mathcal{C}_3(2k, d)$. On the other hand, taking all vertices c_i such that i is even together with $d-2$ vertices of each part of the bipartition of each copy of $K_{d-1,d-1}$ yields an identifying code of this size. Hence $\gamma^{\text{ID}}(\mathcal{C}_3(2k, d)) = k + 2k(d-2) = n - \frac{n}{2d/3}$. An example of this construction is given in Figure 7, where $2k = 8$, $d = 3$, and the black vertices form a minimum identifying code of $\mathcal{C}_3(8, 3)$.

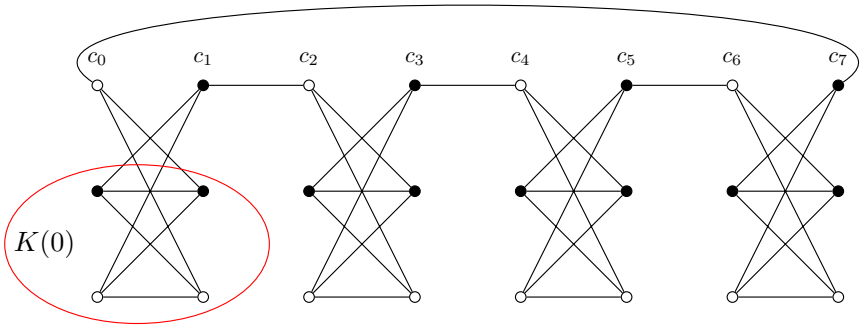


Figure 7: The graph $\mathcal{C}_3(8, 3)$

References

[1] N. Alon and J. H. Spencer. *The probabilistic method*. 2nd edition, Wiley-Interscience, New York, 2000.

[2] D. Auger. Minimal identifying codes in trees and planar graphs with large girth, *European Journal of Combinatorics* 31(5):1372–1384, 2010.

[3] 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.

[4] N. Bertrand. Codes identifiants et codes localisateurs-dominateurs sur certains graphes, Master thesis, ENST, Paris, France, June 2001.

[5] N. Bertrand, I. Charon, O. Hudry and A. Lobstein. 1-identifying codes on trees. *Australasian Journal of Combinatorics* 31:21–35, 2005.

[6] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics* 4:311–316, 1980.

[7] B. Bollobás. *Random graphs*. 2nd edition, Cambridge University Press, 2001.

[8] I. Charon, O. Hudry and A. Lobstein. Extremal cardinalities for identifying and locating-dominating codes in graphs. *Discrete Mathematics* 307(3-5):356–366, 2007.

[9] J.F. Fink and M.S. Jacobson, n-domination in graphs, *Graph Theory with Applications to Algorithms and Computer Science*, John Wiley and Sons, New York (1985), 282–300.

[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. Klasing, A. Kosowski and A. Raspaud. On the size of identifying codes in triangle-free graphs. Submitted for publication, 2010. Available online at `http://arxiv.org/abs/1010.5975`.

[13] A. Frieze, R. Martin, J. Moncel, M. Ruszinkó and C. Smyth. Codes identifying sets of vertices in random networks. *Discrete Mathematics* 307(9-10):1094–1107, 2007.

[14] 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.

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

[16] A. Harutyunyan, P. Horn and J. Verstraete. Independent dominating sets in graphs of girth five. Submitted for publication, 2009. Available online at `http://www.mathcs.emory.edu/~phorn/`.

[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] 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.

[19] 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.

[20] M. Molloy and B. Reed. *Graph Colouring and the Probabilistic Method*. Springer, 1st edition, December 2001.

[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] T. Müller and J.-S. Sereni. Identifying and locating-dominating codes in (random) geometric networks. *Combinatorics, Probability and Computing* 18(6):925–952, 2009.

[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] S. Thomassé and A. Yeo. Total domination of graphs and small transversals of hypergraphs. *Combinatorica* 27(4):473–487, 2007.

[25] N. Wormald. The asymptotic distribution of short cycles in random regular graphs. *Journal of Combinatorial Theory Series B* 31(2):168–182, 1981.