



**HAL**  
open science

## Bounds for identifying codes in terms of degree parameters

Florent Foucaud, Guillem Perarnau

► **To cite this version:**

Florent Foucaud, Guillem Perarnau. Bounds for identifying codes in terms of degree parameters. 2011. hal-00578230v1

**HAL Id: hal-00578230**

**<https://hal.science/hal-00578230v1>**

Preprint submitted on 18 Mar 2011 (v1), last revised 21 Sep 2012 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bounds for identifying codes in terms of degree parameters\*

Florent Foucaud<sup>†</sup>

Guillem Perarnau<sup>‡</sup>

March 18, 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  $\gamma^{\text{ID}}(G)$  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  $G$  is connected, has  $n$  vertices and maximum degree  $d$ , then  $\gamma^{\text{ID}}(G) \leq n - \frac{n}{d} + O(1)$ . We use probabilistic tools to show that  $\gamma^{\text{ID}}(G) \leq n - \frac{n}{\Theta(d^3)}$  holds. For  $d$ -regular graphs and further classes of graphs, this bound is improved to  $\gamma^{\text{ID}}(G) \leq n - \frac{n}{\Theta(d^{3/2})}$ . Moreover, if  $G$  is regular and has no vertices sharing the same neighbours, we show that the conjecture holds up to a constant factor (i.e.  $\gamma^{\text{ID}}(G) \leq n - \frac{n}{\Theta(d)}$ ). In the second part, we prove that in any graph  $G$  of minimum degree  $\delta$  and girth 5,  $\gamma^{\text{ID}}(G) \leq \frac{3 \log \delta}{\delta} n$ . Finally, using the former result, we also determine that almost all  $d$ -regular graphs have their minimum identifying code of size  $\Theta\left(\frac{\log d}{d}\right) \cdot n$ .

## 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 [17]. These applications include the location of threats in facilities using sensors [23], error-detection schemes [17] and routing [18] in networks, as well as the structural analysis of RNA proteins [16].

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. [20, 7, 13, 10, 11]). We focus on degree-related graph parameters such as the maximum and minimum 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 [11]. 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  $\bar{d} = \bar{d}(G)$ . We denote by  $u \sim v$ , the adjacency between two vertices  $u$  and  $v$ , and  $u \not\sim v$  denotes their non-adjacency. The set of neighbours of some vertex  $v$  is called its *open neighbourhood* and denoted by  $N(v)$ , and its *closed neighbourhood*, i.e. the set of its neighbours and itself, is denoted by  $N[v]$ . If two distinct vertices  $u, v$  are such that  $N[u] = N[v]$ , they are called *twins*. If  $u \not\sim v$  and  $N(u) = N(v)$ ,  $u$  and  $v$  are called *weak twins*. The symmetric difference between two sets  $A$  and  $B$  is denoted by  $A\Delta B$ . We use *log* to denote the natural logarithm function. We also make use of the standard asymptotic notations  $o, O, \Theta, \Omega$  and  $\omega$ . Usually the asymptotics are taken either on variables  $d, \delta$  or  $n$ . If we write  $o(1)$  or  $O(1)$ , then by convention the asymptotic is taken only on  $n$ , the number of vertices.

---

\*This research is supported by the ANR Project IDEA - Identifying coDes in Evolving grAphs, ANR-08-EMER-007, 2009-2011. The second author wants to thank the FPU grant from the Ministerio de Educación de España.

<sup>†</sup>LaBRI - Université Bordeaux 1, 351 cours de la Libération, 33405 Talence cedex, France.

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

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  $\mathcal{C}$ .  $\mathcal{C}$  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 1** *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 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*.

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 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

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

Both bounds are tight [17, 13]. Moreover all graphs reaching these two bounds have been classified (see [20] for the lower bound and [10] for the upper bound).

When considering graphs of given maximum degree,  $d$ , it was shown in [17] that the (stronger) lower bound  $\gamma^{\text{ID}}(G) \geq \frac{2n}{d+2}$  holds. This bound is tight and a classification of all graphs reaching it has been proposed in [9]. Note that for any  $d$ , this family of graphs includes regular graphs and graphs of arbitrarily large girth.

It was conjectured in [11] that the following upper bound holds.

**Conjecture 2 ([11])** *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 other classes of graphs described in Appendix B). Therefore if Conjecture 2 holds, for any graph  $G$  on  $n$  vertices and of maximum degree  $d$  we would have  $\frac{2}{d+2}n \leq \gamma^{\text{ID}}(G) \leq \frac{(d-1)}{d}n + O(1)$ , with both bounds being tight.

Note that Conjecture 2 holds for graphs of maximum degree 2, i.e. when  $G$  is a path or a cycle (see [14]). It was shown in [10] 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 (up to a constant factor) if  $G$  is triangle-free: then, the bound  $\gamma^{\text{ID}}(G) \leq n - \frac{n}{3(d+1)}$  holds [11].

Identifying codes have been previously studied in two models of random graphs, that is the classic random graph model [12] and the model of random geometric graphs [22]. To our knowledge random regular graphs have not been studied in the context of identifying codes.

We summarize our results in Table 1 and compare them to the bounds for the dominating set problem (the table contains references for both the bound and its tightness). All bounds are asymptotically tight except for the one of the first entry of the table, which is related to Conjecture 2. 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, interestingly, for larger girth and for most regular graphs, the bounds for the two problems coincide asymptotically, as shown by the last lines of the table.

In order to prove our results, we use probabilistic techniques. For some results, we use the weighted version of Lovász' Local Lemma. However it is not just a basic application of this lemma as we first use it to show the existence of a valid identifying code, and then we use the Chernoff bound to show that this code is also not too large. We believe that the interplay between these two techniques is particularly interesting. We also make use of other probabilistic techniques such as the First Moment Method and the Alteration Method in order to give better bounds in more restricted cases. Finally, we work with

	Identifying codes	Dominating sets
girth 3	$n - \frac{n}{\Theta(d^{3/2})}$ Thm. 13 (Conj. 2: $n - \frac{n}{\Theta(d)}$ )	$\Theta\left(\frac{\log d}{d}\right) \cdot n$ [1], [24]
girth 3 and weak-twin-free	$n - \frac{n}{\Theta(d)}$ Thm. 16, Constr. 27	$\Theta\left(\frac{\log d}{d}\right) \cdot n$ [1], [24]
girth 4	$n - \frac{n}{\Theta(d)}$ [11], Constr. 28	$\Theta\left(\frac{\log d}{d}\right) \cdot n$ [1], [24]
girth 5	$\Theta\left(\frac{\log d}{d}\right) \cdot n$ Thm. 18, Thm. 20	$\Theta\left(\frac{\log d}{d}\right) \cdot n$ [1], [24]
random $d$ -regular graphs (w.h.p)	$\Theta\left(\frac{\log d}{d}\right) \cdot n$ Thm. 19, Thm. 20	$\Theta\left(\frac{\log d}{d}\right) \cdot n$ [1], [24]

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

the Configuration Model [6] in order to compute the identifying code number of a random regular graph with high probability.

We note that our results can be used to design probabilistic polynomial-time algorithms which construct identifying codes of the corresponding cardinalities. Indeed we use techniques that allow to randomly construct the identifying code, such as the Lovász Local Lemma, which is known to have an algorithmic version [21]. Moreover, the method of conditional probabilities [1] can be used to provide a deterministic polynomial-time algorithm to find the identifying code which existence is proven by Theorem 18.

The organization of this paper is as follows. In Section 2 we state some preliminary results which will be used throughout the paper. In Sections 3 and 4, we use the weighted version of Lovász' Local Lemma to give new upper bounds on  $\gamma^{\text{ID}}$  for graphs of maximum degree  $d$ . In particular we improve the known results in Section 3 and give a better bound for  $d$ -regular graphs having no weak twins (Section 4). This gives a new family of graphs for which Conjecture 2 holds up to a constant factor. In Section 5, we give a new upper bound of  $O\left(\frac{\log d}{\delta}\right) \cdot n$  for graphs having minimum degree  $\delta$  and girth at least 5. Finally, we extend the latter result to random regular graphs. We use the configuration model [6] to show that for large  $d$ , almost all  $d$ -regular graphs  $G$  have  $\gamma^{\text{ID}}(G) = \Theta\left(\frac{\log d}{d}\right) \cdot n$ . This is done in Section 6. We provide an appendix with the statements of some standard probabilistic tools we use, such as a modified version of the Weighted Local Lemma and the Chernoff bounds (Appendix A). A further part of the appendix is dedicated to various constructions of families of graphs which show the tightness of some of our results (Appendix B).

## 2 Preliminary results

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 3** *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 easy, but it is worth mentioning here.

**Observation 4** *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 following proposition upper-bounds the number of forced vertices in a twin-free graph and is crucial in some of our proofs.

**Proposition 5** *Let  $G$  be a twin-free graph on  $n$  vertices and of maximum degree  $d$ . Then there are at most  $\left(1 - \frac{1}{d+1}\right) \cdot n$  forced vertices in  $G$ .*

**Proof** Let us define a partial order  $\leq$  over the set of vertices of  $G$  such that  $u \leq v$  if  $N[u] \subseteq N[v]$ . We construct an oriented graph  $H$  on  $V(G)$  as a subgraph of the Hasse diagram of poset  $(V(G), \leq)$ . The arc set of  $H$  is the set of all arcs  $\overrightarrow{uv}$  where there exists some vertex  $x$  such that  $N[v] = N[u] \cup \{x\}$ . Note that then  $x$  is  $uv$ -forced, and we note  $x = f(\overrightarrow{uv})$ . If no edge forces some vertex we call this vertex *free*.

For a vertex  $v$  of  $V(G)$ , we define the set  $P(v)$  as the union of  $v$  itself and the set of all predecessors and successors of  $v$  in  $H$ . Note that since  $v$  has degree at most  $d$  in  $G$  and is adjacent to each of its predecessors and successors in  $H$ , we have  $P(v) \subseteq N[v]$ . Therefore  $|P(v)| \leq d + 1$ .

We also define the set  $Q(v)$  as follows:  $Q(v) = \{s \mid s \text{ is isolated in } H \text{ and } s \text{ is } vw\text{-forced for some } w \in V(G)\}$ . Since each  $vw$ -forced vertex involves some edge  $vw$  we have  $|Q(v)| \leq d$ .

First of all, we need to prove a few preliminary claims.

**Claim 6** *Let  $u, v, w, x, y$  be some vertices of  $G$  such that  $u, v, w$  are distinct,  $x = f(\overrightarrow{uv})$  and  $y = f(\overrightarrow{wb})$ . Then  $x \neq y$ .*

**Proof** Suppose not, then  $N[u] = N[w] = N[v] \setminus \{x\}$ , a contradiction since  $G$  is twin-free.  $\square$

**Claim 7** *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  $H$ , 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  $H$ . Since  $u \not\sim s$ ,  $u \not\sim t$ . Then  $v \not\sim t$  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  $y = s$ .  $\square$

**Claim 8** *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.*

**Proof** Suppose  $s$  has two in-neighbours  $t$  and  $t'$  in  $H$ . By Claim 7,  $v$  is both  $ts$ -forced and  $t's$ -forced, which is a contradiction to Claim 6.  $\square$

**Claim 9** *Let  $s$  be a forced vertex in  $G$  with  $s = f(\overrightarrow{uv})$ , and let  $t$  be a forced in-neighbour of  $s$  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 7,  $v = f(\overrightarrow{ts})$ , that is,  $v$  is the unique vertex such that  $v$  is adjacent to  $s$ , but not to  $t$ . Therefore  $x = v$ .  $\square$

**Claim 10** *Let  $s$  be a non-isolated sink in  $H$  which is forced in  $G$  with  $s = f(\overrightarrow{uv})$  for some vertices  $u$  and  $v$ . Then either  $s$  has a free predecessor, or there exists a free vertex  $w$  such that  $P(s) \subseteq N[w]$ .*

**Proof** If  $s$  has a free predecessor in  $H$  we are done. Otherwise, suppose all predecessors of  $s$  are forced. By Claim 8,  $s$  and its predecessors form a directed path  $\{t_0, \dots, t_k, s\}$  in  $H$ . Notice that by Claim 7,  $v = f(\overrightarrow{t_k s})$ . By our assumption we know that  $t_k$  is forced, say  $t_k = f(\overrightarrow{xy})$ . But now by Claim 9,  $x = v$ , that is,  $v$  has an out-neighbour,  $y$ . Let us call this out-neighbour,  $v_k$ . What we just showed is that  $t_k = f(\overrightarrow{vv_k})$ . Now, repeating these arguments for each other predecessor of  $s$  shows that there is a directed path  $\{u, v, v_k, \dots, v_0\}$  with  $t_k = f(\overrightarrow{vv_k})$  and for all  $0 \leq i \leq k - 1$ ,  $t_i = f(\overrightarrow{v_{i+1} v_i})$ . By applying Claim 9 on vertices  $v_1, v_0$  and  $t_0$ , if  $v_0$  is forced then  $t_0$  has an in-neighbour in  $H$ , a contradiction — hence  $v_0$  is free. Moreover note that since  $v_0 \sim t_0$ ,  $v_0$  is adjacent to all successors of  $t_0$  in  $H$ , that is, to all elements of  $P(s)$ . Therefore, putting  $w = v_0$ , we obtain the claim.  $\square$

To prove the result, we construct a set  $X = \{x_1, \dots, x_k\}$  of free vertices such that  $\bigcup_{i=1}^k A(x_i) = V(G)$ , where  $A(x_i)$  is a set of at most  $d + 1$  vertices. Then we have  $k \geq \frac{n}{d+1}$  and the claim of the proposition follows.

Let  $S$  be the set of sinks in  $H$ . Since each vertex is either a sink or has a successor which is a sink,  $\bigcup_{s \in S} P(s) = V(G)$ . Hence we will consider each sink  $s$  of  $H$  and ensure that  $P(s) \subseteq \bigcup_{i=1}^k A(x_i)$  in order to prove that  $\bigcup_{i=1}^k A(x_i) = V(G)$ .

We now describe a procedure to build set  $X$  while considering each sink of  $H$ . In this procedure, we first consider all free sinks. Then, we consider all forced sinks which have at least one in-neighbour in  $H$ . Finally, we consider all forced sinks which are isolated in  $H$ . We denote by  $s$  the currently considered sink.

**Case a.** Sink  $s$  is free. Then we put  $s$  in  $X$  and we set  $A(s) = N[s]$ . We have  $P(s) \subseteq A(s)$  and  $|A(s)| \leq d + 1$ .

For the remaining cases we suppose that  $s$  is forced, i.e. there exist two vertices  $u, v$  such that  $s = f(\overrightarrow{uv})$ .

**Case b.** Sink  $s$  has at least one in-neighbour in  $H$ , say  $t$ . We distinguish two subcases.

1.  $s$  has a free predecessor in  $H$ . Let  $p$  be its closest free predecessor. We put  $p$  in  $X$  and set  $A(p) = N[p]$ . By Claim 8, note that each predecessor of  $s$  which is also a successor of  $p$ , has only one in-neighbour. Hence  $P(s) \subseteq P(p) \subseteq A(p)$ .
2.  $s$  has no free predecessor in  $H$ . Then by Claim 10 there exists a free vertex  $w$  such that  $P(s) \subseteq N[w]$ . Note that up to this point of the procedure, for each vertex  $x \in X$ ,  $A(x) = N[x]$ . Hence if  $w$  already belongs to  $X$ , we have  $A(w) = N[w]$ . Otherwise, we add  $w$  to  $X$  and set  $A(w) = N[w]$ . In both cases since  $P(s) \subseteq N[w]$  we are done.

**Case c.** Sink  $s$  is isolated in  $H$ . Note that  $P(s) = \{s\}$ . We suppose that  $s$  does not belong to some set  $A(x)$  for  $x \in X$  (otherwise we are done). We claim that both  $v$  and  $u$  are free. Suppose  $u$  is forced and there exist  $x, y$  such that  $f(\overrightarrow{xy}) = u$ . Then  $y \sim u$ , hence  $y \sim v$ . Now since  $f(\overrightarrow{xy}) = u$ ,  $x \sim v$ . But  $u \not\sim x$ . The only vertex adjacent to  $v$  but not to  $u$  is  $s$ , hence  $s = x$  and  $s$  has  $y$  as an out-neighbour, a contradiction since  $s$  is a sink. A similar argument shows that  $v$  is free too. We now distinguish the following subcases.

1. Vertex  $u$  does not yet belong to  $X$ . Then we add  $u$  to  $X$ , and set  $A(u) = Q(u)$ . Recall that  $P(s) = \{s\} \subseteq Q(u)$  and  $|Q(u)| \leq d$ . Note that doing this also solves the case of all other isolated sinks of  $Q(u)$ .
2. Vertex  $u$  belongs to  $X$ , but vertex  $v$  does not. Then we add  $v$  to  $X$  and set  $A(v) = Q(u)$ .
3. Vertex  $u$  belongs to  $X$ , and vertex  $v$  has been put in  $X$  in Case a or Case b. Then we have  $A(v) = N[v]$ . Since  $s \sim v$   $P(s) = \{s\} \subseteq A(v)$  and we are already done.
4. Vertex  $u$  belongs to  $X$ , and vertex  $v$  has been put in  $X$  in Case c.1 (resp. Case c.2). This means that when  $v$  was originally added to  $X$ , we had  $A(v) = Q(v)$  (resp.  $A(v) = Q(u')$  for some in-neighbour of  $v$  with  $u \neq u'$ ). We re-set  $A(v) \leftarrow A(v) \cup Q(u)$ . Note that this re-setting may happen several times — potentially, once for each in-neighbour of  $v$  in  $H$ . If  $\{u_1, \dots, u_k\}$  denotes the set of these in-neighbours, clearly in the end of the process we have  $A(v) \subseteq Q(v) \cup \bigcup_{i=1}^k Q(u_i)$  (resp.  $A(v) \subseteq \bigcup_{i=1}^k Q(u_i)$ ). We claim that  $|\bigcup_{i=1}^k Q(u_i)| \leq |Q(v) \cup \bigcup_{i=1}^k Q(u_i)| \leq d$ . Indeed, to each element  $q$  of  $Q(v)$ , one can assign a unique outgoing arc of  $v$ ,  $\overrightarrow{vq}$ , to  $q$ , as the arc such that  $f(\overrightarrow{vq}) = q$ . Similarly, to each element  $q$  of some set  $Q(u_i)$ , one can uniquely assign the incoming arc of  $v$ ,  $\overrightarrow{u_i v}$ , to  $q$ . Since there are at most  $d$  arcs incident to  $v$ , this proves our claim and we are done.

□

Note that Proposition 5 is tight. Indeed, consider the graph  $A_k$  on  $2k$  vertices defined in [10] as follows:  $V(A_k) = \{x_1, \dots, x_{2k}\}$  and  $E(A_k) = \{x_i x_j \mid |i - j| \leq k - 1\}$ .  $A_k$  can be seen as the  $(k - 1)$ -th power of the path  $P_{2k}$ . Now, in the graph  $A_k$  with an extra universal vertex  $x$  (i.e.  $x$  is adjacent to all vertices of  $V(A_k)$ ), one can check that all vertices but  $x$  are forced. This graph has  $n = 2k + 1$  vertices, maximum degree  $2k$  and exactly  $\frac{1}{n} = \frac{1}{d+1}$  non-forced vertices. 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 2. Indeed in this case  $n - \frac{n}{d+1} = n - \frac{n}{d} + O(1)$ . So we ask the following question:

**Question 11** *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  $n \cdot \left(1 - \frac{1}{d+1}\right)$  forced vertices?*

Answering this question in positive would provide counterexamples to Conjecture 2. Note that for the similar question where we replace  $d + 1$  by  $d$ , the answer is positive by Construction 26 of the appendix. For any  $d$ , this construction provides arbitrarily large graphs having  $\left(1 - \frac{1}{d}\right) \cdot n$  forced vertices.

Finally, the next proposition shows an upper bound on the number of weak twins in a graph.

**Proposition 12** *Let  $G$  be a graph on  $n$  vertices having maximum degree  $d$ , then  $G$  has at most  $\frac{n(d-1)}{2}$  pairs of weak twins.*

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

Note that the bound of this lemma 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 weak twins.

### 3 Upper bounds for the general case

In this section, we improve the known upper bounds by using the Weighted Local Lemma, stated in Lemma 24 of the appendix. We note that this proof is not just a basic application of the Local Lemma. Indeed we first build a random set, and use the Weighted Local Lemma to show that with non-zero probability, it is an identifying code. However we then have to show that with non-zero probability, this code is not too large and moreover, that these two events are correlated.

**Theorem 13** *Let  $G$  be a twin-free graph on  $n$  vertices having maximum degree  $d$  and no isolated vertices. Let  $f(G)$  denote the proportion of non-forced vertices of  $G$ . Then*

$$\gamma^{ID}(G) \leq \begin{cases} \left(1 - \frac{f(G)}{7.5 \cdot d^{3/2}}\right) \cdot n & \text{if } f(G) \geq 9d^{-1/2} \\ \left(1 - \frac{f(G)^2}{67.5 \cdot d}\right) \cdot n & \text{if } f(G) < 9d^{-1/2} \end{cases}$$

**Proof** First of all we note that in this proof, we assume that  $d$  is sufficiently large as it makes all computations easier. However we note that it is possible to do the computations for small values of  $d$  too.

Let  $F$  be the set of forced vertices of  $G$ , and  $V' = V \setminus F$ . 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  $F \cup (V' \setminus S)$  is an identifying code of  $G$ . Using the Weighted Local Lemma, we compute a lower bound on the (non-zero) probability that this holds for some set  $S_0$ . Finally, we use the Chernoff bound to show that with non-zero probability, the size of this set  $S_0$  is large enough for our purposes. This shows that such a set  $S_0$  exists, and it can be used to build an identifying code that has a sufficiently large size.

Let  $p = \frac{1}{k(d) \cdot d^{3/2}}$  ( $k(d)$  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  $\text{Bin}(n \cdot f(G), p)$  and has expected value  $\mathbb{E}[|S|] = p \cdot n \cdot f(G) = \frac{n \cdot f(G)}{k(d) \cdot d^{3/2}}$ .

Let us now define our “bad” events. These are of four types. An illustration is given in Figure 1.

- **Type A:** for each vertex  $u \in V'$ , let  $A_u$  be the event that  $N[u] \subseteq S$ .
- **Type B:** for each pair  $\{u, v\}$  of adjacent vertices of  $V'$ , let  $B_{u,v}$  be the event that  $(N[u] \Delta N[v]) \subseteq S$ .
- **Type C:** for each pair  $\{u, v\} \in V'$  of vertices at distance two from each other such that  $u$  and  $v$  are not weak twins, let  $C_{u,v}$  be the event that  $(N[u] \Delta N[v]) \subseteq S$ .
- **Type D:** for each pair  $\{u, v\} \in V'$  of weak twins, let  $D_{u,v}$  be the event that  $(N[u] \Delta N[v]) \subseteq S$ .

All the events of type  $B_{u,v}$  where  $|N[u] \Delta N[v]| = 1$  do not happen because  $F$  belongs to the code. Observe that the events  $C_{u,v}$  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,  $V \setminus S$  is a dominating set of  $G$ . If no event of type  $B$  occurs, all pairs of adjacent vertices are separated by  $V \setminus S$ . If no event of type  $C$  or  $D$  occurs, all pairs of vertices at distance 2 from each other are separated. Thus by Observation 3,  $V \setminus S$  is also a separating set of  $G$ , and therefore it is an identifying code of  $G$ .

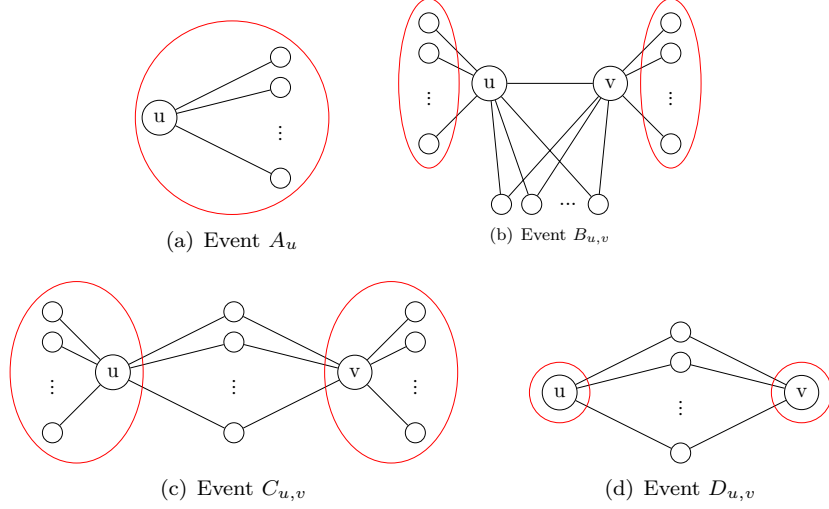


Figure 1: The “bad” events. The circled vertices belong to set  $S$ .

Let us call  $E_{IC}$  the event that no event of type  $A$ ,  $B$ ,  $C$  or  $D$  occurs. Using the Weighted Local Lemma, we want to show that  $Pr(E_{IC}) > 0$ , but first of all we have to check the hypothesis of the lemma.

We define the weight  $t_i$  of each event  $E_i$  as the number of vertices participating in it (i.e. the number of vertices that have to belong to set  $S$  when  $E_i$  holds, as defined previously). Note that since  $G$  has maximum degree  $d$  and no isolated vertices, for each vertex  $u$ , we have  $2 \leq |N[u]| \leq d + 1$ . An event of type  $B$  has size at least 2, otherwise there would be a unique vertex belonging to this event. This vertex would be  $uv$ -forced and would belong to  $F$ , and therefore not to  $S$ . Thus we get the following bounds

$$2 \leq t_A \leq d + 1 \quad 2 \leq t_B \leq 2d - 2 \quad 3 \leq t_C \leq 2d \quad t_D = 2$$

where  $t_i = t_X$  ( $X \in \{A, B, C, D\}$ ) if  $E_i$  is of type  $X$ .

Thus we have  $Pr(A_u) \leq p^2$ ,  $Pr(B_{u,v}) \leq p^2$ ,  $Pr(C_{u,v}) \leq p^3$  and  $Pr(D_{u,v}) \leq p^2$ .

In order to apply the Weighted Local Lemma, we need to know the number of dependencies among the different type of events. The entry  $(X, Y)$  of the following event-intersection table for  $X, Y \in \{A, B, C, D\}$ , shows an upper bound for the number of events of type  $Y$  which some event of type  $X$  may depend on. The proof of correctness of this table is based on a tedious case analysis. We only prove the correctness of entry  $(A, A)$  and omit the other cases.

**Proof of correctness of entry  $(A, A)$ :** Consider some event  $A_u$  of type  $A$  for some vertex  $u \in V'$ . For some other vertex  $v \in V'$ , the fact that  $A_u$  depends on event  $A_v$  means that there is a nonempty intersection between  $N[u]$  and  $N[v]$ . Therefore  $v$  must be either a neighbour of  $u$ , or a vertex at distance 2 of  $u$ . There are at most  $d + d(d - 1) = d^2$  such vertices, which completes the proof of this case.

	$A$	$B$	$C$	$D$
$A$	$d^2$	$d^3 - 2d^2 + 2d$	$d^4 + d^2 - 2d$	$d^2 - d + 1$
$B$	$2(d^2 - d + 1)$	$2d^3 - 4d^2 + 4d - 1$	$2d(d^3 - 3d^2 + 4d - 2)$	$d^2 - d$
$C$	$2d^2 + 2$	$2d(d^2 - 2d + 2)$	$2d^4 - 6d^3 + 9d^2 - 5d - 1$	$d^2 + d - 2$
$D$	$d + 2$	$d^2 + d$	$d^3 - d$	$2d - 4$

For two events  $E_i$  and  $E_j$ , we note  $i \sim j$  if  $E_i$  and  $E_j$  are not mutually independent. In order to apply the Weighted Local Lemma (Lemma 24), we need the following condition for each event  $E_i$  (we recall that  $\gamma(x) = \frac{1 - \sqrt{1 - 2x}}{x}$ ).

$$\sum_{i \sim j} (2p)^{t_j} \leq \frac{\ln 2}{\gamma((2p)^{t_i})} t_i \quad (1)$$



This leads to the following inequalities

$$d^2(2p)^2 + (d^3 - 2d^2 + 2d)(2p)^2 + (d^4 + d^2 - 2d)(2p)^3 + (d^2 - d + 1)(2p)^2 \leq \frac{2 \ln 2}{\gamma((2p)^2)} \quad (2)$$

$$2(d^2 - d + 1)(2p)^2 + (2d^3 - 4d^2 + 4d - 1)(2p)^2 + 2d(d^3 - 3d^2 + 4d - 2)(2p)^3 + (d^2 - d)(2p)^2 \leq \frac{2 \ln 2}{\gamma((2p)^2)} \quad (3)$$

$$(2d^2 + 2)(2p)^2 + 2d(d^2 - 2d + 2)(2p)^2 + (2d^4 - 6d^3 + 9d^2 - 5d - 1)(2p)^3 + (d^2 + d - 2)(2p)^2 \leq \frac{3 \ln 2}{\gamma((2p)^3)} \quad (4)$$

$$(d + 2)(2p)^2 + (d^2 + d)(2p)^2 + (d^3 - d)(2p)^3 + (2d - 4)(2p)^2 \leq \frac{2 \ln 2}{\gamma((2p)^2)} \quad (5)$$

Observe that Inequality (3) implies Inequalities (2), (4) and (5). Moreover Inequality (3) is implied by the following one.

$$8d^3 p^2 \leq \frac{2 \ln 2}{\gamma(4p^2)} \quad (6)$$

Note that the Taylor expansion of  $\gamma(x)^{-1}$  in  $x = 0$  is  $\frac{x}{1-\sqrt{1-2x}} = 2 + o(1)$ . Hence when we plug the exact value of  $\gamma(4p^2)$  into Inequality 6, we observe that we need  $p = O(d^{-3/2})$  in order to solve it.

Hence, we fix  $p = \frac{1}{k(d) \cdot d^{3/2}}$ . Note that this implies that  $k(d) = \Omega(1)$ . In fact, in our proof we will always have  $k(d) \geq 1$ . Since  $d \geq 3$ , we know that  $p \leq 1/5$  and therefore  $\gamma(4p^2) < 1.096$ , which implies  $k(d) \geq 2.52$ .

Under this condition the Weighted Local Lemma can be applied. Let  $\{E_1, \dots, E_M\}$  be the set of “bad” events, and let  $t_i$  be such that  $Pr(E_i) \leq p^{t_i}$ . The Weighted Local Lemma not only shows that the probability of  $E_{IC}$  is strictly positive, but  $Pr(E_{IC}) \geq \prod_{i=1}^M (1 - (2p)^{t_i})$ .

Recall that  $t_A \geq 2$ ,  $t_B \geq 2$ ,  $t_C \geq 3$  and  $t_D \geq 2$ . Let  $M_X$  be the number of events of type  $X$ , where  $X \in \{A, B, C, D\}$ . We have

$$Pr(E_{IC}) \geq \prod_{i=1}^{M_A} (1 - (2p)^2) \prod_{i=1}^{M_B} (1 - (2p)^2) \prod_{i=1}^{M_C} (1 - (2p)^3) \prod_{i=1}^{M_D} (1 - (2p)^2)$$

Moreover,  $M_A \leq n \cdot f(G)$  since by definition there is one event of type  $A$  for each vertex of  $V'$ .  $M_B \leq \frac{nd}{2}$  since there is at most one event  $B$  for each edge in  $G$ .  $M_C$  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  $\frac{n \cdot f(G) \cdot d \cdot (d-1)}{2}$ . Finally,  $M_D$  is at most the number of pairs of weak twins in  $V'$ , that is  $n \cdot f(G) \cdot \frac{d-1}{2}$  by Proposition 12. Hence, we have

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

Using Lemma 24 and the facts that  $\gamma(4p^2) \leq 1.096$  and  $p = \frac{1}{k(d) \cdot d^{3/2}}$

$$\begin{aligned} Pr(E_{IC}) &\geq \exp \left\{ -\gamma(4p^2) \left( f(G)(2p)^2 + \frac{d}{2}(2p)^2 + \frac{f(G)d(d-1)}{2}(2p)^3 + \frac{f(G)(d-1)}{2}(2p)^2 \right) n \right\} \\ &\geq \exp \left\{ -1.096 \cdot \frac{2}{k(d)^2 d^2} \left( \frac{2f(G)}{d} + 1 + \frac{2f(G)}{k(d) \cdot d^{1/2}} + f(G) \right) n \right\} \end{aligned}$$

Using the fact that  $f(G) \leq 1$ , one can check that

$$Pr(E_{IC}) \geq \exp \left\{ -\frac{5}{k(d)^2 d^2} n \right\}$$

The Weighted Local Lemma shows that  $S$  has the desired properties with probability  $Pr(E_{IC}) > 0$ , implying that such a set exists. Let  $S_0$  be such a set. Note that we have no guarantee on the size of  $S_0$ . In fact, if  $S_0 = \emptyset$  then  $V \setminus S_0$  is always an identifying code. Therefore we need to estimate the probability that  $|S_0|$  is below its expected size minus some proportion  $\beta \cdot n \cdot f(G)$  of vertices. For convenience we set  $\beta = \frac{1}{c(d)}$  where  $c(d)$  can be an arbitrary positive function of  $d$ . To do this we use the Chernoff bound of Theorem 25 (see Appendix A) by putting  $a = \beta \cdot n \cdot f(G)$ . Let  $E_{BIG}$  be the event that  $|S_0| - p \cdot n > -\frac{n \cdot f(G)}{c(d)}$ . We get

$$\begin{aligned}
Pr(\overline{E_{\text{BIG}}}) &\leq \exp\left\{-\frac{\left(\frac{n \cdot f(G)}{c(d)}\right)^2}{2pn \cdot f(G)}\right\} \\
&= \exp\left\{-\frac{k(d)f(G)d^{3/2}}{2c(d)^2}n\right\}
\end{aligned}$$

Now we have

$$\begin{aligned}
Pr(E_{\text{IC}} \text{ and } E_{\text{BIG}}) &= 1 - Pr(\overline{E_{\text{IC}}} \text{ or } \overline{E_{\text{BIG}}}) \\
&\geq 1 - Pr(\overline{E_{\text{IC}}}) - Pr(\overline{E_{\text{BIG}}}) \\
&= 1 - (1 - Pr(E_{\text{IC}})) - Pr(\overline{E_{\text{BIG}}}) \\
&= Pr(E_{\text{IC}}) - Pr(\overline{E_{\text{BIG}}}) \\
&\geq \exp\left\{-\frac{5}{k(d)^2 d^2}n\right\} - \exp\left\{-\frac{k(d)f(G)d^{3/2}}{2c(d)^2}n\right\}
\end{aligned}$$

Thus,  $Pr(E_{\text{IC}} \text{ and } E_{\text{BIG}}) > 0$  if  $c(d) < \frac{k(d)^{3/2}d^{7/4}f(G)^{1/2}}{\sqrt{10}}$ . We fix  $c(d) = (1 - \epsilon)\frac{k(d)^{3/2}d^{7/4}f(G)^{1/2}}{\sqrt{10}}$ , for some very small constant  $\epsilon$ .

Now we have to check that  $E_{\text{BIG}}$  implies that the size of  $S_0$  is still large enough.

$$\begin{aligned}
|S_0| &\geq \mathbb{E}(|S|) - \frac{n \cdot f(G)}{c(d)} \\
&= \frac{n \cdot f(G)}{k(d) \cdot d^{3/2}} - \frac{n \cdot f(G)}{c(d)} \\
&= \left(\frac{1}{k(d)} - \frac{\sqrt{10}}{(1 - \epsilon)k(d)^{3/2}d^{1/4}f(G)^{1/2}}\right) \frac{n \cdot f(G)}{d^{3/2}}
\end{aligned} \tag{7}$$

Let us now distinguish between two cases depending on the behaviour of the function  $f(G)$ .

**Case 1.**  $f(G) = \omega(d^{-1/2})$ .

Plugging this into Equation (7) together with the fact that  $k(d) \geq 2.52$  (see Equation (6)), we get

$$|S_0| \geq \left(1 - \frac{\sqrt{10}}{(1 - \epsilon)k(d)^{1/2}d^{1/4}f(G)^{1/2}}\right) \frac{n \cdot f(G)}{2.52 \cdot d^{3/2}} = (1 - o_d(1)) \frac{n \cdot f(G)}{2.52 \cdot d^{3/2}} \geq \frac{n \cdot f(G)}{7.5 \cdot d^{3/2}}$$

which completes the first part of the statement as  $d$  is taken to be large enough.

**Case 2.**  $f(G) = O(d^{-1/2})$ .

Since  $|S|$  must be positive, from Equation (7) we need  $k(d)^{3/2}d^{1/4}f(G)^{1/2} > k(d)\sqrt{10}$  and therefore  $k(d) = \frac{c_0}{f(G)\sqrt{d}}$  where  $c_0 > 10^{(1)}$ . This expression is optimized when  $c_0 = \frac{45}{2}$ .

Let us set

$$k(d) = \begin{cases} 2.52 & \text{if } f(G) = O(d^{-1/2}) \text{ and } f(G) \geq 9d^{-1/2} \\ \frac{45}{2\sqrt{d} \cdot f(G)} & \text{if } f(G) = O(d^{-1/2}) \text{ and } f(G) < 9d^{-1/2} \end{cases}$$

Note that in both cases, for large enough  $d$ ,  $k(d) \geq 2.52$ .

Now we can see that

$$|S_0| \geq \begin{cases} \frac{f(G)}{7.5 \cdot d^{3/2}}n & \text{if } f(G) = O(d^{-1/2}) \text{ and } f(G) \geq 9d^{-1/2} \\ \frac{f(G)^2}{67.5 \cdot d}n & \text{if } f(G) = O(d^{-1/2}) \text{ and } f(G) < 9d^{-1/2} \end{cases}$$

Finally, considering all cases, we get an identifying code  $\mathcal{C} = F \cup (V' \setminus S)$  such that

$$|\mathcal{C}| \leq \begin{cases} \left(1 - \frac{f(G)}{7.5 \cdot d^{3/2}}\right)n & \text{if } f(G) \geq 9d^{-1/2} \\ \left(1 - \frac{f(G)^2}{67.5 \cdot d}\right)n & \text{if } f(G) < 9d^{-1/2} \end{cases}$$

□

---

<sup>1</sup>Note that if  $f(G) = \omega(d^{-1/2})$ , we get  $k(d) = o(1)$ . In that case the conditions for applying the Weighted Local Lemma (see Equation (6)) are not satisfied, so this approach cannot be used in Case 1.

The general theorem applies to graphs where  $f(G)$  as defined in the statement of Theorem 13, is constant. In particular for regular graphs  $f(G) = 1$  because a forced vertex implies the existence of two vertices with distinct degrees.

**Corollary 14 (Graphs with constant number of forced vertices)** *Let  $G$  be a twin-free graph on  $n$  vertices having maximum degree  $d$  and  $f(G)$  not depending on  $d$ . Then  $\gamma^{ID}(G) \leq n - \frac{n}{\Theta(d^{3/2})}$ . In particular if  $G$  is  $d$ -regular,  $\gamma^{ID}(G) \leq n - \frac{n}{7.5 \cdot d^{3/2}}$ .*

By Proposition 5 we know that  $f(G) \geq \frac{1}{d+1}$ . Thus we obtain the following general result.

**Corollary 15 (General case)** *Let  $G$  be a twin-free graph on  $n$  vertices with maximum degree  $d$ . Then  $\gamma^{ID}(G) \leq n - \frac{n}{67.5 \cdot d(d+1)^2} = n - \frac{n}{\Theta(d^3)}$ .*

## 4 A tight upper bound for regular graphs without weak twins

In this section we show that Conjecture 2 holds asymptotically for the class of regular graphs without weak twins. Moreover, as shown by Construction 27 of the appendix, for any  $d$  there are arbitrarily large graphs from this class that need  $n - \frac{n}{\Theta(d)}$  vertices in any identifying code. Therefore this result is asymptotically tight for this class. Moreover we believe that this class of graphs is interesting as it is still very large, and is a natural extension of twin-free graphs.

We also remark that the technique used in this section with non-regular graphs, would give a weaker bound than the ones of the previous section.

**Theorem 16** *Let  $G$  be a  $d$ -regular graph on  $n$  vertices without weak twins. Then  $\gamma^{ID}(G) \leq n - \frac{n}{41d}$ .*

**Proof** The proof for this theorem is based on the same ideas than the one of Theorem 13. Again we assume that  $d$  is sufficiently large. Observe that if  $G$  has no weak twins, there appear no events of type  $D$ . Moreover, since  $G$  is regular, there are no forced vertices in  $G$ . Thus, it is not necessary to add them beforehand as we did in the proof of Theorem 13. In this case, in order to be more accurate, we will split the events of type  $B$  and  $C$  depending on their size. Hence, the defined events are,

- **Type A:** for each vertex  $u$ , let  $A_u$  be the event that  $N[u] \subseteq S$ .
- **Type  $B^i$  ( $2 \leq i \leq 2d - 2$ ):** for each pair  $\{u, v\}$  of adjacent vertices, let  $B_{u,v}^i$  be the event that  $|N[u] \Delta N[v]| = i$  and  $(N[u] \Delta N[v]) \subseteq S$ .
- **Type  $C^i$  ( $4 \leq i \leq 2d$ ):** for each pair  $\{u, v\} \in V$  of vertices at distance two from each other, let  $C_{u,v}^i$  be the event that  $|N[u] \Delta N[v]| = i$  and  $(N[u] \Delta N[v]) \subseteq S$ .

Observe that since  $G$  is regular, the events  $B^i$  and  $C^i$  are just defined for even  $i$ . The size of the events is

$$t_A = d + 1 \quad t_{B^i} = i \quad t_{C^i} = i$$

The cell  $(X, Y)$  of the following table shows an upper bound for the number of events of type  $Y$  that can intersect with an event of type  $X$ .

	A	...	$B^j$	...	$C^j$	...
A	$d^2$	...	$b_j^A d^2$	...	$c_j^A d^3$	...
$\vdots$	$\vdots$		$\vdots$		$\vdots$	
$B^i$	$i(d+1)$	...	$b_j^{B^i} d^2$	...	$c_j^{B^i} d^3$	...
$\vdots$	$\vdots$		$\vdots$		$\vdots$	
$C^i$	$i(d+1)$	...	$b_j^{C^i} d^2$	...	$c_j^{B^i} d^3$	...
$\vdots$	$\vdots$		$\vdots$		$\vdots$	

The constants  $b_j^E$  and  $c_j^E$  are the number of events of type  $B^j$  and  $C^j$  respectively that intersect with a fixed event  $E$ .

Observe that for any event  $E$  of size  $i$

$$\sum_{j=2}^{2d-2} b_j^E \leq id^2 \quad \sum_{j=4}^{2d} c_j^E \leq id^3 \quad (8)$$

The conditions for the different types of events given by Equation (1) in the proof of Theorem 13, can now be translated as follows

$$d^2(2p)^{d+1} + \sum_{j=2}^{2d-2} b_j^X(2p)^j + \sum_{j=4}^{2d} c_j^X(2p)^j \leq \frac{(d+1)\ln 2}{\gamma((2p)^{d+1})} \quad (X \text{ of type } A) \quad (9)$$

$$i(d+1)(2p)^{d+1} + \sum_{j=2}^{2d-2} b_j^X(2p)^j + \sum_{j=4}^{2d} c_j^X(2p)^j \leq \frac{i \ln 2}{\gamma((2p)^i)} \quad (X \text{ of type } B^i) \quad (10)$$

$$i(d+1)(2p)^{d+1} + \sum_{j=2}^{2d-2} b_j^X(2p)^j + \sum_{j=4}^{2d} c_j^X(2p)^j \leq \frac{i \ln 2}{\gamma((2p)^i)} \quad (X \text{ of type } C^i) \quad (11)$$

Using (8), the equations (9), (10) and (11) are implied by

$$(d+1)(2p)^{d+1} + d^2(2p)^2 + d^3(2p)^4 \leq \frac{\ln 2}{\gamma((2p)^2)}$$

Taking  $p = \frac{1}{k^d}$  for some  $k > 4$ , all the previous equations are satisfied. As in the previous proof,  $E_{IC}$  denotes the event that  $V \setminus S$  is a valid identifying code (i.e. no “bad” event occurs), and  $E_{BIG}$  is the event that the size of  $S$  is from the same order that the expected one. The Weighted Local Lemma gives the following lower bound

$$Pr(E_{IC}) \geq \exp \left\{ -\gamma(4p^2)n \left( (2p)^{d+1} + \frac{d}{2}(2p)^2 + \frac{d(d-1)}{2}(2p)^4 \right) \right\} \geq \exp \left\{ -\frac{3}{k^2 d} n \right\}$$

Whereas the Chernoff bound gives

$$Pr(\overline{E_{BIG}}) \leq \exp \left\{ -\frac{k}{2c^2 d} n \right\}$$

Now, using the same arguments than in the proof of Theorem 13, this leads to  $c < \frac{k^{3/2}}{\sqrt{6}}$ , and the size of the code is optimized when  $k = 41$ . Hence we get  $\gamma^{ID}(G) \leq |V \setminus S| \leq n - \frac{n}{41d}$ .  $\square$

## 5 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 the results of this section in Section 6 about random regular graphs.

Despite being different than our previous proofs, the ones of this section have also a probabilistic flavour. We realized that for graphs of girth 5, applying the (Weighted) Local Lemma does not lead to a tight result. However, using the first moment 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  $V \setminus D$ ,  $|N(v) \cap D| \geq 2$  [8]. 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  $G[D]$ , the subgraph induced by  $D$ , 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 separating.

First of all, note that all the vertices that do not lay in  $D$  are separated. Indeed they are dominated at least twice each, and as  $g(G) > 4$  they cannot be dominated by the two same vertices.

Similarly, a vertex  $x$  of  $D$  and a vertex  $y$  of  $V \setminus D$  are separated since  $y$  has two vertices which dominate it, but they cannot both dominate  $x$  (otherwise there would be a triangle or a  $C_4$  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.  $\square$

The following theorem makes use of the previous proposition and uses the first moment method.

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

**Proof** Let  $S \subseteq V$  be a random subset of vertices, where each vertex  $v \in V$  is added to  $S$  uniformly at random with probability  $p$  (where  $p$  will be determined later). For every vertex  $v \in V$ , 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 non-2-dominated vertices, i.e.  $T = \{v \mid X_v = 1\}$ . Note that  $X = \sum X_v = |T|$ . Let us estimate the size of  $T$ . Observing that  $|N[v] \cap S| \sim \text{Bin}(\delta + 1, p)$  we get

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

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

$$\begin{aligned} \mathbb{E}[|D|] &\leq \mathbb{E}[|S|] + \mathbb{E}[X] \\ &\leq np + n(1+\delta p)e^{-\delta p} \\ &= n \left( p + (1+\delta p)e^{-\delta p} \right) \end{aligned}$$

Let us set  $p = \frac{\log \delta}{\delta}$ . Then we get  $\mathbb{E}[|D|] \leq (1 + o_\delta) \frac{2 \log \delta}{\delta} n$ . This shows that there exists at least one 2-dominating set  $D$  with size at most  $(1 + o_\delta) \frac{2 \log \delta}{\delta} n$ .

**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  $G[D]$  (i.e., they are adjacent and both of degree 1 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 get a 2-dominating set of size at most  $|D| + \frac{|D|}{2} = (1 + o_\delta) \frac{3 \log \delta}{\delta} n$  having the desired property. Now applying Lemma 17 completes Case 1.

**Case 2:** (sparse case) In the case where  $\bar{d}(G) = o_\delta(\delta \log \delta)$ , 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  $N[u] \Delta N[v] \subseteq V \setminus S$ . Using the facts that  $p = \frac{\log \delta}{\delta}$  and  $1-x \leq e^{-x}$  if  $x \leq 1$ , let us calculate the expected value of  $Y = \sum_{uv \in E} Y_{uv}$ .

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

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

$$\mathbb{E}(|\mathcal{C}|) \leq \mathbb{E}(|S|) + \mathbb{E}(X) + \mathbb{E}(Y) \leq (1 + o_\delta) \frac{2 \log \delta}{\delta} n + (1 + o_\delta(1)) \frac{n\bar{d}}{2\delta^2}$$

Using that  $\bar{d} = o_\delta(\delta \log \delta)$ ,

$$\mathbb{E}(|\mathcal{C}|) \leq (1 + o_\delta(1)) \frac{2 \log \delta}{\delta} n$$

Then there exists some choice of  $S$  such that  $|\mathcal{C}| \leq (1 + o_\delta(1)) \frac{2 \log \delta}{\delta} n$ . This directly implies that  $\gamma^{\text{ID}}(G) \leq (1 + o_\delta(1)) \frac{2 \log \delta}{\delta} n$  and completes the proof.  $\square$

Note that Theorem 18 is asymptotically tight. Indeed, it is shown in the next section (Corollary 22) that for large enough  $d$ , almost all  $d$ -regular graphs have both girth 5 and their minimum identifying code of size  $\Theta(\frac{\log d}{d}n)$ . Moreover, the theorem cannot be extended much in the sense that if we drop the condition on girth 5, we have arbitrarily large  $d$ -regular triangle-free graphs having large minimum identifying codes. For example, such a graph is  $G$  given by Construction 28 of the appendix, and has  $\gamma^{\text{ID}}(G) = (1 - \frac{1}{\Theta(d)})n$ . 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.  $\gamma^{\text{ID}}(T_{d,h}) = \lceil n - \frac{n}{d-1+\frac{1}{d}} \rceil$  [4]).

## 6 Identifying codes of random regular graphs

An interesting question that arises, is which is the value for parameter  $\gamma^{\text{ID}}$  for most  $d$ -regular graphs? We know some lower and upper bounds for the identifying code size, 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  $K_{dn}$  at random (see [6] for 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  $G$  has neither loops nor multiedges. It is well-known (see e.g. [6]) that

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

Then any property that holds with probability 1 for  $\mathcal{G}^*(n, d)$  as  $n \rightarrow \infty$ , will be directly proven for  $\mathcal{G}(n, d)$ . In this case we will say that the property holds with high probability (w.h.p.).

**Theorem 19** *Let  $G \in \mathcal{G}(n, d)$  then w.h.p.  $\gamma^{\text{ID}}(G) \leq (1 + o_d(1)) \frac{2 \log d}{d} n$ .*

**Proof** First of all we have to show that almost each random regular graph is identifiable, i.e.  $G \in \mathcal{G}(n, d)$  is twin-free w.h.p.

Observe that the number of perfect matchings of  $K_m$  is  $m!! = m(m-2)(m-4) \dots 2$ . Fix a vertex  $u$  of  $G$  and let  $\{v_1, \dots, v_d\} = N(u)$ . We will give the probability that  $u$  and  $v = v_1$  are twins, i.e.  $N[u] = N[v]$ . The number of perfect matchings of  $K_{dn}$  such that in the resulting graph  $G$  of  $\mathcal{G}(n, d)$ ,  $v_2 \in N(v)$ , is at most  $(d-1)(d-1)(nd-2d-2)!!$ . Analogously the number of perfect matchings with  $v_2, v_3 \in N(v)$  is at most  $(d-1)(d-1)(d-2)(d-1)(nd-2d-4)!!$ . Thus we have

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

As we have at most  $\frac{nd}{2}$  possible pair of twins (one for each edge), the probability of having some twins is

$$\Pr(\exists \text{ a pair of twins}) \leq \frac{nd}{2} \left(\frac{d}{n}\right)^{d-1} \rightarrow 0$$

since  $d \geq 3$ . Therefore, almost all random regular graphs are identifiable.

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

$$|\mathcal{C}| \leq (1 + o_d(1)) \frac{2 \log d}{d} n$$

that separates any pair of vertices except from the ones where both vertices lie in a triangle or in a  $C_4$ .

Classical results on random regular graphs (independently, [5] and [25]) state that the random variables that count the number of cycles of length  $k$ ,  $C_k$ , tend in distribution to independent Poisson variables with parameter  $c_k = \frac{1}{2k}(d-1)^k$ .

In general we have to handle the case of these cycles. Observe that

$$\mathbb{E}(C_3) = \frac{(d-1)^3}{6} \quad \mathbb{E}(C_4) = \frac{(d-1)^4}{8}$$

i.e. a constant number of triangles and  $C_4$  are expected.

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

$$\Pr(C_3 > k) \leq \frac{(d-1)^3}{6k} \quad \Pr(C_4 > k) \leq \frac{(d-1)^4}{8k}$$

For  $k = \log n$ , the probability of having more than  $k$  triangles is  $o(1)$ . Then w.h.p., for constant  $d$ , we have at most  $\log n$  cycles of length 3 and 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  $v_{ij} \in N(u_i) \setminus N(u_j)$ . 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  $T = \{u_1, u_2, u_3, u_4\}$  is an induced  $C_4$ , adding  $T$  to  $\mathcal{C}$  we will separate all the elements in  $T$ . Observe that it is not necessary to consider non induced cycles because then all the vertices belong to triangles and they have been identified in the previous step.

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

$$\gamma^{\text{ID}}(G) \leq (1 + o_d(1)) \frac{2 \log d}{d} n + 7 \log n = (1 + o_d(1)) \frac{2 \log d}{d} n$$

Observe that the  $o_d(1) \frac{2 \log d}{d} n$  term contains also the  $7 \log n$  term.

Note that for the case where  $d \rightarrow \infty$ , it is not even necessary to use the argument of the small cycles, since as shown by the value of the variable  $C_k$ , almost every random regular graph has girth at least 5.  $\square$

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.  $\frac{d-1}{d}n$  for the graphs of Construction 27 of the appendix), almost all  $d$ -regular graphs have 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 [15]. 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  $(1 + o_d(1)) \frac{\log d}{d} n$ .*

**Proof** We will proceed by contradiction. Given a set of vertices  $X$  of size  $m$ , we will compute the probability that  $X$  dominates  $Y = V \setminus X$ . Recall that  $G$  has been obtained from the configuration model by selecting a random perfect matching of  $K_{dn}$ . Let  $y \in Y$  fixed, then the event  $A_y = \{N(X) \cap y \neq \emptyset\}$ , i.e.  $y$  is dominated by  $X$ , satisfies

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

where we have used the fact that  $1 - x \leq e^{-x}$  if  $x \leq 1$ .

The probability that  $X$  is covering  $Y = \{y_1, \dots, y_{n-m}\}$  is

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

We claim that  $\Pr(A_{y_i} | \cap_{j=0}^{i-1} A_{y_j}) \leq \Pr(A_{y_i})$ . The probability that  $y_i$  is not dominated by  $X$  taking into account that  $y_1, \dots, y_{i-1}$  are dominated, is

$$\begin{aligned} \Pr(\overline{A_{y_i}} | \cap_{j=0}^{i-1} A_{y_j}) &= \left(1 - \frac{d}{nd - 2(i-1) - 1}\right) \left(1 - \frac{d}{nd - 2(i-1) - 3}\right) \dots \left(1 - \frac{d}{nd - 2md + 1}\right) \\ &\geq \left(1 - \frac{d}{nd - 1}\right) \left(1 - \frac{d}{nd - 3}\right) \dots \left(1 - \frac{d}{nd - 2md + 1}\right) \\ &= \Pr(\overline{A_{y_i}}) \end{aligned}$$

and by reversing,  $\Pr(A_{y_i} | \cap_{j=0}^{i-1} A_{y_j}) \leq \Pr(A_{y_i})$ , hence the events are positively correlated.

Then

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

If  $m \leq \frac{\log d - c \log \log d}{d}$  for certain  $c > 2$

$$\Pr(\cap_{y \in Y} A_y) \leq e^{-\frac{(\log d)^c}{d}n}$$

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

$$\binom{n}{m} \leq \frac{n^m}{m!} \leq \left(\frac{en}{m}\right)^m \sim \left(\frac{de}{\log d}\right)^{(\log d - c \log \log d)n/d}$$

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

Therefore, applying the union bound

$$\begin{aligned} \Pr(\exists X \text{ of size } m \text{ and dominating}) &\leq \left(\frac{de}{\log d}\right)^{(\log d - c \log \log d)n/d} e^{-\frac{(\log d)^c}{d}n} \\ &\sim \exp\left\{\frac{\log d - c \log \log d}{d}(\log d + 1 - \log \log d)n - \frac{(\log d)^c}{d}n\right\} \\ &\sim \exp\left\{\frac{(\log d)^2}{d}n - \frac{(\log d)^c}{d}n\right\} \rightarrow 0 \end{aligned}$$

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

Since any identifying code is also a dominating set, we get the following result as an immediate corollary.

**Corollary 21** *Let  $G \in \mathcal{G}(n, d)$  then w.h.p.  $\gamma^{ID}(G) \geq (1 + o_d(1))\frac{\log d}{d}n$ .*

Together with the result of Theorem 19, this yields the following result.

**Corollary 22** *Let  $d$  be a sufficiently large integer. For almost all  $d$ -regular graphs  $G$ ,  $\gamma^{ID}(G) = \Theta\left(\frac{\log d}{d}\right)n$ .*

It must be stressed that when identifying random regular graphs, we need something not far from a 2-dominating set. This leads us to the following conjecture.

**Conjecture 23** *Let  $d$  be a sufficiently large integer. For almost all  $d$ -regular graphs  $G$ ,  $\gamma^{ID}(G) = (1 + o_d(1))\frac{2 \log(d)}{d}n$ .*



## References

- [1] N. Alon, 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, I. Charon, O. Hudry and A. Lobstein. 1-identifying codes on trees. *Australasian Journal of Combinatorics* 31:21–35, 2005.
- [5] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics* 4:311–316, 1980.
- [6] B. Bollobás. *Random graphs*. 2nd edition, Cambridge University Press, 2001.
- [7] 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.
- [8] 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.
- [9] F. Foucaud. Identifying codes in special graph classes. Master thesis, Université Bordeaux 1, France, June 2009. Under the supervision of R. Klasing and A. Raspaud. Available online at <http://www.labri.fr/perso/foucaud/Research/MastersThesis/>.
- [10] 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.
- [11] 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>.
- [12] A. Frieze, R. Martin, J. Moncel, M. Ruzinkó and C. Smyth. Codes identifying sets of vertices in random networks. *Discrete Mathematics* 307(9-10):1094–1107, 2007.
- [13] S. Gravier and J. Moncel. On graphs having a  $V \setminus \{x\}$  set as an identifying code. *Discrete Mathematics* 307(3-5):432–434, 2007.
- [14] S. Gravier, J. Moncel and A. Semri. Identifying codes of cycles, *European Journal of Combinatorics* 27(5):767–776, 2006.
- [15] 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/>.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] M. Molloy and B. Reed. *Graph Colouring and the Probabilistic Method*. Springer, 1st edition, December 2001.
- [20] 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.
- [21] R. A. Moser and G. Tardos. A constructive proof of the general Lovász Local Lemma. Manuscript, 2009. Available online at <http://arxiv.org/abs/0903.0544>.

- [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.

## APPENDIX

### A Probabilistic tools

We 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 which fits quite good to our problem. The Weighted Local Lemma is stated e.g. in [19]. There, it is claimed that the numerical values of this version can be slightly improved for certain applications. As we want to optimize the constant in our results, we will need to state a tighter version than the one of [19].

**Lemma 24 (Weighted Local Lemma [19], modified version)** *Let  $\mathcal{E} = \{E_1, \dots, E_M\}$  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 integers  $t_1, \dots, t_M \geq 1$  and a real  $p < 1$  such that for each  $1 \leq i \leq M$ :*

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

where  $\gamma(x) = \frac{1 - \sqrt{1 - 2x}}{x}$ .

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

The main difference between this lemma and the Weighted Local Lemma as stated in [19], is that in the latter case  $\gamma(x) = 2 \ln 2$ .

**Proof** Assume that the generalized version of the local lemma holds: if there are some reals  $0 < x_i \leq 1$ , such that for any event  $E_i$

$$Pr(E_i) \leq x_i \prod_{i \sim j} (1 - x_j) \tag{12}$$

then,  $Pr(\bigcap_{i=1}^M E_i) \geq \prod_{i=1}^M (1 - x_i) > 0$ .

First of all observe that  $1 - x \geq \exp \{-\gamma(x)x\}$ :

$$1 - x - \exp \{-\gamma(x)x\} = 1 - x - (1 - \gamma(x)x + \frac{\gamma(x)^2 x^2}{2} - O(x^3)) \geq (\gamma(x) - 1)x + \frac{\gamma(x)^2 x^2}{2}$$

and the former expression is non-negative for the chosen  $\gamma(x)$ .

Thus, given the initial hypothesis of the lemma we must show (12). Setting  $x_i = (2p)^{t_i}$

$$\begin{aligned}
x_i \prod_{i \sim j} (1 - x_j) &= (2p)^{t_i} \prod (1 - (2p)^{t_j}) \\
&\geq (2p)^{t_i} \exp \left\{ -\gamma((2p)^{t_i}) \sum (2p)^{t_j} \right\} \\
&\geq (2p)^{t_i} \exp \left\{ -\gamma((2p)^{t_i}) \frac{\ln 2}{\gamma((2p)^{t_i})} t_i \right\} \\
&\geq (2p)^{t_i} \exp \{-\ln 2 \cdot t_i\} \\
&= p^{t_i} \geq \Pr(E_i)
\end{aligned}$$

Therefore,  $\Pr(\bigcap_{i=1}^M E_i) \geq \prod_{i=1}^M (1 - x_i) = \prod (1 - (2p)^{t_j})$  □

Note that  $\gamma(x) \rightarrow 1$  as  $x \rightarrow 0$ . Therefore if we have a good bound on  $p$ , we will get a tighter version of the Local Lemma. However, the gamma function does not give an asymptotic improvement.

We also use the well-known Chernoff bound (see e.g. [19]).

**Theorem 25 (Chernoff bound)** *Let  $X$  be a random variable of  $n$  independent trials of probability  $p$ , and let  $a > 0$  be a real number. Then  $\Pr(X - np \leq -a) \leq e^{-\frac{a^2}{2pn}}$ .*

## B Constructions

This section gathers some constructions which are used for showing the tightness of some of our upper bounds. Some of these constructions were already known.

**Construction 26** *Given any  $d_H$ -regular graph  $H$  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 26 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  $d_H n_H = (1 - \frac{1}{d})n$  forced vertices. In fact taking all these forced vertices gives an identifying code, therefore  $\gamma^{\text{ID}}(\mathcal{C}_1(H)) = (1 - \frac{1}{d})n$ . An example of this construction is given in Figure 2, where  $H$  is the hypercube of dimension 3  $H_3$ , and the black vertices are those which belong to a minimum identifying code of  $\mathcal{C}_1(H_3)$ .

The following construction is very similar, but yields regular graphs. It can be found in [9].

**Construction 27** [9] *Given any  $d_H$ -regular graph  $H$  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,  $\gamma^{\text{ID}}(\mathcal{C}_2(H)) = (1 - \frac{1}{d})n$ . An example of this construction is given in Figure 3, where  $H$  is the complete graph  $K_5$ , and the black vertices form a minimum identifying code of  $\mathcal{C}_2(K_5)$ .

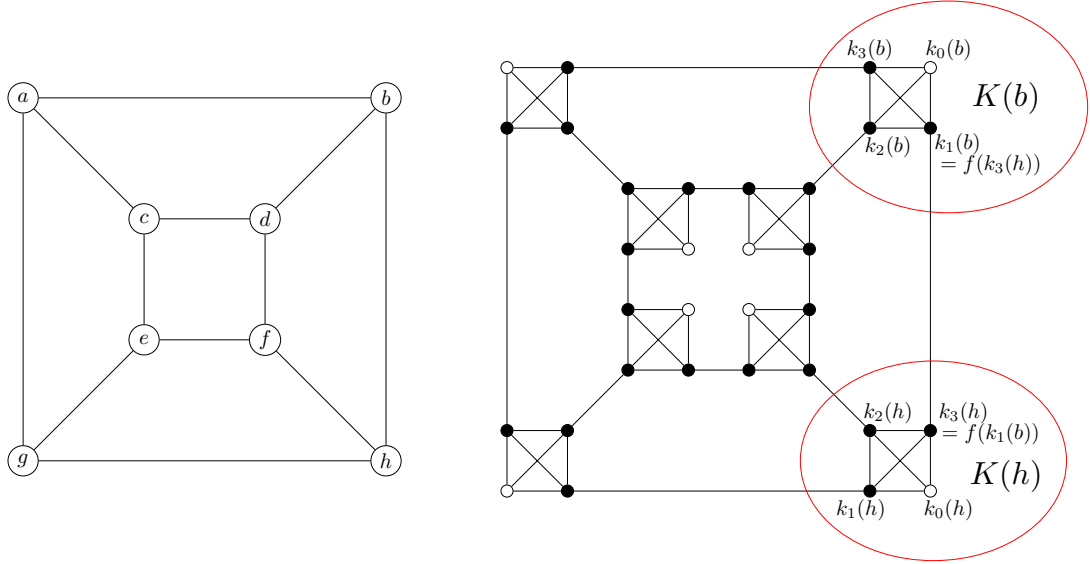


Figure 2: The graphs  $H_3$  and  $\mathcal{C}_1(H_3)$

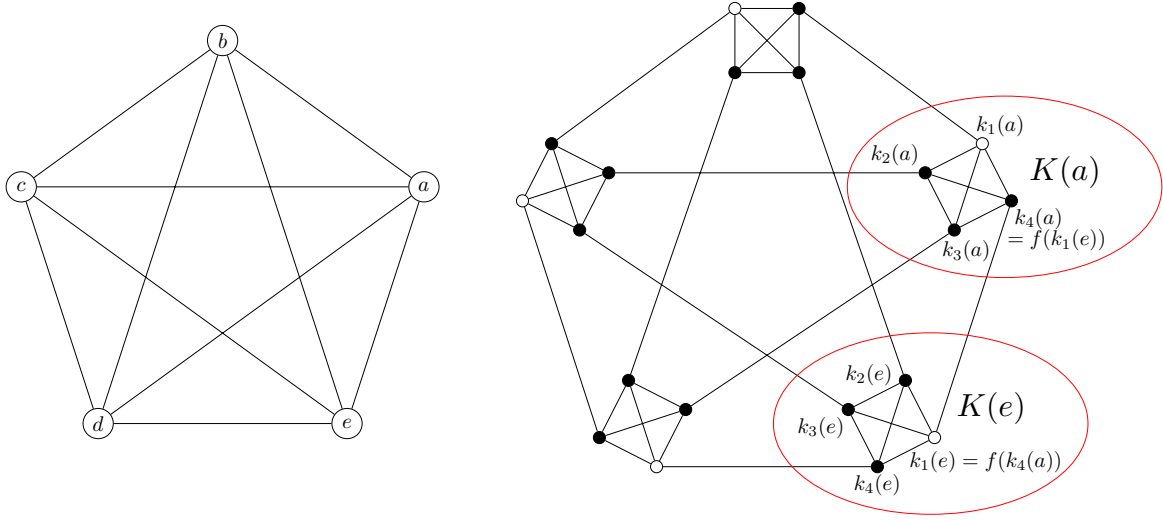


Figure 3: The graphs  $K_5$  and  $\mathcal{C}_2(K_5)$

**Construction 28** [9] 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  $\{c_0, \dots, c_{2k-1}\}$  be a set of  $2k$  vertices with a 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  $\gamma^{\text{ID}}(\mathcal{C}_3(2k, d)) = k + 2k(d - 2) = (1 - \frac{1}{2d/3})n$ . An example of this construction is given in Figure 4, where  $2k = 8$ ,  $d = 3$ , and the black vertices form a minimum identifying code of  $\mathcal{C}_3(8, 3)$ .

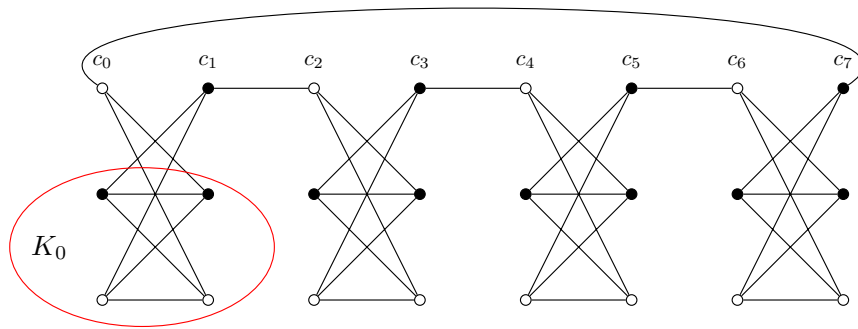


Figure 4: The graph  $\mathcal{C}_3(8, 3)$