ARTICLE IN PRESS

Discrete Applied Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On the number of optimal identifying codes in a twin-free graph

Iiro Honkala[a,*], Olivier Hudry[b], Antoine Lobstein[c]

[a] University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland

[b] Institut Télécom - Télécom ParisTech & CNRS - LTCI UMR 5141, 46, rue Barrault, 75634 Paris Cedex 13, France

[c] CNRS - LTCI UMR 5141 & Institut Télécom - Télécom ParisTech, 46, rue Barrault, 75634 Paris Cedex 13, France

ARTICLE INFO

Article history:
Received 26 February 2014
Received in revised form 18 July 2014
Accepted 18 August 2014
Available online xxxx

Keywords:
Graph theory
Twin-free graphs
Identifiable graphs
Identifying codes

ABSTRACT

Let G be a simple, undirected graph with vertex set V. For $v \in V$ and $r \geq 1$, we denote by $B_{G,r}(v)$ the ball of radius r and centre v. A set $C \subseteq V$ is said to be an r-*identifying code* in G if the sets $B_{G,r}(v) \cap C$, $v \in V$, are all nonempty and distinct. A graph G which admits an r-identifying code is called r-*twin-free* or r-*identifiable*, and in this case the smallest size of an r-identifying code in G is denoted by $\gamma_r^{ID}(G)$.

We study the number of different optimal r-identifying codes C, i.e., such that $|C| = \gamma_r^{ID}(G)$, that a graph G can admit, and try to construct graphs having "many" such codes.

1. Introduction

We introduce basic definitions and notation for graphs (for which we refer to, e.g., [1,5]) and for identifying codes (see [12] and the bibliography at [13]).

We shall denote by $G = (V, E)$ a simple, undirected graph with vertex set V and edge set E, where an *edge* between $x \in V$ and $y \in V$ is indifferently denoted by $\{x, y\}$, $\{y, x\}$, xy or yx. The *order* of a graph is its number of vertices $|V|$.

A *path* $P_n = x_1x_2 \dots x_n$ is a sequence of n distinct vertices x_i, $1 \leq i \leq n$, such that x_ix_{i+1} is an edge for $i \in \{1, 2, \dots, n-1\}$. The *length* of P_n is its number of edges, $n - 1$.

A graph G is called *connected* if for any two vertices x and y, there is a path between them. It is called *disconnected* otherwise. In a connected graph G, we can define the *distance* between any two vertices x and y, denoted by $d_G(x, y)$, as the length of any shortest path between x and y, since such a path exists. This definition can be extended to disconnected graphs, using the convention that $d_G(x, y) = +\infty$ if there is no path between x and y.

For any vertex $v \in V$ and integer $r \geq 1$, the *ball of radius* r *and centre* v, denoted by $B_{G,r}(v)$, is the set of vertices within distance r from v:

$$B_{G,r}(v) = \{x \in V : d_G(v, x) \leq r\}.$$

Two vertices x and y such that $B_{G,r}(x) = B_{G,r}(y)$ are called (G, r)-*twins*; if G has no (G, r)-twins, that is, if

$$\forall x, y \in V \text{ with } x \neq y, \quad B_{G,r}(x) \neq B_{G,r}(y),$$

then we say that G is r-*twin-free* or r-*identifiable*. When there is no ambiguity about the graph G, we may use simply $B_r(v)$.

* Corresponding author. Fax: +358 22310311.
E-mail addresses: honkala@utu.fi (I. Honkala), hudry@telecom-paristech.fr (O. Hudry), lobstein@telecom-paristech.fr (A. Lobstein).

http://dx.doi.org/10.1016/j.dam.2014.08.020
0166-218X/© 2014 Elsevier B.V. All rights reserved.

Please cite this article in press as: I. Honkala, et al., On the number of optimal identifying codes in a twin-free graph, Discrete Applied Mathematics (2014), http://dx.doi.org/10.1016/j.dam.2014.08.020

Whenever two vertices x and y are within distance r from each other in G, i.e., $x \in B_r(y)$ and $y \in B_r(x)$, we say that x and y *r-cover* each other. When three vertices x, y, z are such that $x \in B_r(z)$ and $y \notin B_r(z)$, we say that z *r-separates* x and y in G (note that $z = x$ is possible). A set is said to r-separate x and y in G if it contains at least one vertex which does.

A *code* C is simply a subset of V, and its elements are called *codewords*. For each vertex $v \in V$, the *r-identifying set* of v, with respect to C, is the set of codewords r-covering v, and is denoted by $I_{G,C,r}(v)$:

$$I_{G,C,r}(v) = B_{G,r}(v) \cap C.$$

We say that C is an *r-dominating* code in G if all the sets $I_{G,C,r}(v),\ v \in V$, are nonempty (see, e.g., [8] or [9] on the large topic of domination).

We say that C is an *r-identifying* code [12] if all the sets $I_{G,C,r}(v),\ v \in V$, are nonempty and distinct: in other words, every vertex is r-covered by at least one codeword, and every pair of vertices is r-separated by at least one codeword. Or: given the (nonempty) identifying set $I_{G,C,r}(v)$ of an unknown vertex $v \in V$, we can uniquely recover v (we also say that we *r-identify* v).

It is quite easy to observe that a graph G admits an r-identifying code if and only if G is r-twin-free; this is why r-twin-free graphs are also called r-identifiable.

When G is r-twin-free, we denote by $\gamma_r^{ID}(G)$ the smallest cardinality of an r-identifying code in G. Any r-identifying code C such that $|C| = \gamma_r^{ID}(G)$ is said to be *optimal*. The search for an optimal r-identifying code in given graphs or families of graphs is an important part of the studies devoted to identifying codes. In general, this problem is NP-hard [3].

Lemma 1. *If $r \geq 1$ and G is an r-twin-free graph of order n, then*

$$\gamma_r^{ID}(G) \geq \lceil \log_2(n+1) \rceil.$$

Proof. Any r-identifying code C in G must provide n nonempty r-identifying sets to the n vertices of G, therefore $2^{|C|} - 1 \geq n$. □

One application of identifying codes is the following: we place ourselves in the case $r = 1$ and assume that we have to protect a museum, or any other type of premises, using smoke detectors. The museum can be viewed as a graph, where the vertices represent the rooms, and the edges, the doors between rooms. The detectors are located in some of the rooms and give the alarm whenever there is smoke in their room or in one of the adjacent rooms. If there is smoke in one room and if the detectors are located in rooms corresponding to a 1-identifying code, then, only by knowing which detectors gave the alarm, we can identify the room where someone is smoking.

In this paper, following [15,14] where the notion of "completely different codes" is discussed in the framework of infinite lattices, we are interested in finding graphs which have a large number of different optimal r-identifying codes (Section 3). Typically, we shall construct graphs of order n which admit $2^{\alpha \cdot n}$ different optimal r-identifying codes, and we want to have α as close to 1 as possible. Our results are:

- $\approx 2^{0.77003n}$ for $r = 1$ (Theorem 11);
- $\lfloor 2^{(\frac{1+\log_2 5}{5} - \varepsilon)n} \rfloor$ for $r \geq 1$ and any $\varepsilon > 0$ (Corollary 16), knowing that $\frac{1+\log_2 5}{5} \approx 0.664$.

Considering again our example of the watching of a museum, this means that we want not only to use the smallest possible number of detectors, but also to have a large number of choices for their locations.

Note that if we are interested in the number of r-identifying codes that a graph G can admit (without requiring optimality), then, using Lemma 1, we obtain the upper bound

$$\Sigma_{i=\lceil \log_2(n+1) \rceil}^{n} \binom{n}{i} = 2^n - \Sigma_{i=0}^{\lceil \log_2(n+1) \rceil - 1} \binom{n}{i}.$$

On the other hand, consider the graph $G = (V, E)$ with $V = Z_q \times Z_q$ and $E = \{\{x = (a, b), y = (c, d)\} : a = c \text{ or } b = d\}$, i.e., G is a $q \times q$ square array of vertices in which any two vertices are adjacent if and only if they are on the same row or column (identifying codes in these graphs are studied in [7]). Then any subset of V containing at least two vertices in every row and every column is a 1-identifying code: a vertex v which has in its 1-identifying set two codewords belonging to the same row is itself in this row, the same is true for columns, and the position of v is uniquely determined (note that these codes are not optimal, see [7]). Now there are exactly $(q+1)2^{q^2-q}$ subsets of vertices with at most one vertex in, say, the first row, and therefore at most $q^2(q+1)2^{q^2-q}$ codes which are not 1-identifying, so almost all of the 2^{q^2} subsets of V are 1-identifying: if we set $n = q^2$, we have at least

$$2^n \left(1 - \frac{n(\sqrt{n}+1)}{2^{\sqrt{n}}}\right)$$

1-identifying codes in G.

In comparison, some results are known for 1-dominating codes with respect to *minimality* (for inclusion): it has been proved in [6] that graphs with n vertices exist which admit $2^{0.651n}$ different minimal 1-dominating codes, and that any graph with n vertices admits at most $2^{0.779n}$ different minimal 1-dominating codes; the upper bound is obviously valid also for optimal codes, and the lower bound is obtained by a construction which works for optimal codes too. See also [4]. The clique on n vertices is an obvious example where all the $2^n - 1$ nonempty subsets of vertices are 1-dominating codes.

In [11], we study not the number, but the ensemble of all the optimal r-identifying codes in a graph, in particular the distances between optimal codes; the same is done for 1-dominating codes, and for 1-locating-dominating codes – which we do not define here – in the short note [10].

2. Basic facts

Before we proceed, we need some additional definitions, as well as some easy but useful lemmas. In particular, the following trivial lemma will often be used implicitly.

Lemma 2. *Let $G = (V, E)$ be a graph. If C is r-identifying in G, so is any $D \supseteq C$.* □

Definition 3. Let $G = (V, E)$ be a graph. Its r-*th power*, or r-*th transitive closure*, is the graph denoted by $G^r = (V, E^r)$ and defined by $E^r = \{xy : x \in V, y \in V, d_G(x, y) \leq r\}$.

Lemma 4. *Let $r \geq 1$ and $s \geq 1$ be two integers. A code C is (rs)-identifying in G if and only if it is r-identifying in G^s.*

Proof. For any two vertices $u, v \in V$, we have $d_{G^s}(u, v) \leq r$ if and only if $d_G(u, v) \leq rs$. □

Lemma 5 *([2, Lemma 1(i) and Remark 3])***.** *Let $r \geq 1$, let G be a path, and let C be a subset of vertices of G. If all vertices are r-covered by C and all pairs of consecutive vertices are r-separated by C, then C is r-identifying in G.* □

Definition 6. Let $G = (V, E)$ be a graph. A code C is said to be r-*separating* if all the sets $I_{G,C,r}(v),\ v \in V$, are distinct; we say that C is r-*separating-only* if all the sets $I_{G,C,r}(v),\ v \in V$, are distinct *and* one of them is empty.

We denote by $\chi_r^S(G)$ the smallest size of an r-separating-only code in G, when such a code exists, and we say that C is an *optimal* r-separating-only code if $|C| = \chi_r^S(G)$.

Lemma 7. *If an r-separating-only code exists in a graph G, then an r-identifying code also exists, but the converse is not true. Moreover, we have:*

$$\gamma_r^{ID}(G) \geq \chi_r^S(G) \geq \gamma_r^{ID}(G) - 1, \tag{1}$$

and if $\chi_r^S(G) = \gamma_r^{ID}(G) - 1$, then all optimal r-separating codes are r-separating-only.

Proof. The paths P_3 and P_4 are examples of graphs admitting 1-identifying codes but no 1-separating-only code. If C is an r-separating-only code in G, with the vertex v not r-covered by C, then $C \cup \{v\}$ is r-identifying in G; this argument also shows that $\gamma_r^{ID}(G) \leq |C| + 1$.

The first inequality and the last assertion are obvious. □

Lemma 8. *Let G_0 be a graph of order n_0, admitting S_0 different optimal r-identifying codes and σ_0 different optimal r-separating-only codes. For $p \geq 1$, let G be the graph consisting of p copies of G_0. Then G has $n = pn_0$ vertices, admits*

$$(S_0)^p = 2^{\frac{\log_2 S_0}{n_0} n}$$

different optimal r-identifying codes, which are of size $p\gamma_r^{ID}(G_0)$, and admits

$$p\sigma_0(S_0)^{p-1}$$

different optimal r-separating-only codes, of size $\chi_r^S(G_0) + (p-1)\gamma_r^{ID}(G_0)$.

Proof. The only way to construct an r-identifying code in G is to take an r-identifying code in each copy, and for optimality to take an optimal code in each copy.

The only way to construct an r-separating-only code in G is to take one r-separating-only code in one copy, and one r-identifying code in each of the other copies, and for optimality to take an optimal code in each copy.

In both cases, all the choices are independent. □

The next proposition and lemma deal with the case $r = 1$.

Proposition 9. *Let G_0 be a graph of order $n_0 \geq 2$, admitting S_0 different optimal 1-identifying codes and σ_0 different optimal 1-separating-only codes. We assume that*

$$\chi_1^S(G_0) = \gamma_1^{ID}(G_0) - 1. \tag{2}$$

For $p \geq 2$, let $G = (V, E)$ be the graph consisting of p copies of G_0. Let x be a vertex not belonging to V and $G(x)$ be the graph with vertex set $V(x) = V \cup \{x\}$ and edge set $E(x) = E \cup \{\{x, y\} : y \in V\}$; we say that x is a universal *vertex. We denote by $S(G(x))$ (respectively, $\sigma(G(x))$) the number of different optimal 1-identifying (respectively, 1-separating-only) codes of $G(x)$.*

Then the graph $G(x)$ has $n = pn_0 + 1$ vertices and the set $\mathcal{C}$ of optimal 1-identifying codes in $G(x)$ is equal to the union of

- (a) *the set $\mathcal{C}_1$ of all the optimal 1-identifying codes in G, and of*
- (b) *the set $\mathcal{C}_2$ of all the sets $\{x\} \cup C$, where C is an optimal 1-separating-only code in G.*

These codes have size $p\gamma_1^{ID}(G_0)$. The number $S(G(x))$ verifies

$$S(G(x)) = (S_0)^p + p\sigma_0(S_0)^{p-1}. \tag{3}$$

Also, the set $\mathcal{C}_3$ of optimal 1-separating-only codes in $G(x)$ is equal to

- (c) *the set $\mathcal{C}_4$ of all the codes which are optimal 1-separating-only in G.*

These codes have size $p\gamma_1^{ID}(G_0) - 1$, and the number $\sigma(G(x))$ verifies

$$\sigma(G(x)) = p\sigma_0(S_0)^{p-1}. \tag{4}$$

Proof. (i) Let C be a 1-identifying code in G: the sets $I_{G,C,1}(v),\ v \in V$, are all different and nonempty. Now let us consider the same code C in $G(x)$; we have $I_{G(x),C,1}(x) = C$, and this nonempty set is different from all the other 1-identifying sets, because there are at least two copies of G_0, so one copy cannot contain all the codewords. This proves that C is also 1-identifying in $G(x)$. Hence $\gamma_1^{ID}(G(x)) \leq \gamma_1^{ID}(G)$.

(ii) Let C be an optimal 1-identifying code in $G(x)$, containing x. Since x cannot 1-separate any pair of vertices in G, its only purpose as a codeword is to 1-cover some vertices not 1-covered by any other codeword; because these vertices are 1-separated by C, only one of them, which we denote by v, can be such that $I_{G(x),C,1}(v) = \{x\}$. Then $C^* = C \setminus \{x\} \cup \{v\}$ is also optimal and 1-identifying in $G(x)$. Now $C^* \subseteq V$, and C^* is 1-identifying in G. Therefore, $\gamma_1^{ID}(G) \leq \gamma_1^{ID}(G(x))$.

(iii) By (i) and (ii), any optimal 1-identifying code in G is an optimal 1-identifying code in $G(x)$, the inclusion $\mathcal{C}_1 \subseteq \mathcal{C}$ is proved, and

$$\gamma_1^{ID}(G(x)) = \gamma_1^{ID}(G). \tag{5}$$

(iv) Let C be a 1-separating-only code in G: the sets $I_{G,C,1}(v),\ v \in V$, are all different and nonempty, except for a unique vertex v_0, which is such that $I_{G,C,1}(v_0) = \emptyset$. Now $C^* = C \cup \{x\}$ is 1-identifying in $G(x)$, because $I_{G(x),C^*,1}(x) = C^*$ and $I_{G(x),C^*,1}(v_0) = \{x\}$ are unique 1-identifying sets; in particular, no vertex other than x can have C^* as its 1-identifying set, because, G_0 being of size at least two, the codewords of C cannot be all in one of the p copies ($p \geq 2$).

Now, using, for G, Lemma 8 with $r = 1$, the assumption (2), and equality (5), we have:

$$\begin{aligned}\chi_1^S(G) &= \chi_1^S(G_0) + (p-1)\gamma_1^{ID}(G_0) = p\gamma_1^{ID}(G_0) - 1 \\ &= \gamma_1^{ID}(G) - 1 = \gamma_1^{ID}(G(x)) - 1.\end{aligned} \tag{6}$$

Therefore, if C is an optimal 1-separating-only code in G, then $C \cup \{x\}$ is an optimal 1-identifying code in $G(x)$, of size $p\gamma_1^{ID}(G_0)$, and the inclusion $\mathcal{C}_2 \subseteq \mathcal{C}$ is proved.

(v) We now prove that $\mathcal{C} \subseteq \mathcal{C}_1 \cup \mathcal{C}_2$. Let C be an optimal 1-identifying code in $G(x)$. If C does not contain x, then C is 1-identifying also in G, and by (5), C is optimal in G. If C contains x, then $C \setminus \{x\}$ is 1-separating-only in G (otherwise, x is a useless codeword and C is not optimal in $G(x)$), and by (5) and the right-hand-side inequality in (1), it is optimal in G.

(vi) We have just proved that $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$. Therefore, the number of optimal 1-identifying codes in $G(x)$ is equal to the number of optimal 1-identifying codes in G, plus the number of optimal 1-separating-only codes in G, the latter number being equal to $p\sigma_0(S_0)^{p-1}$. This proves equality (3).

(vii) Let C be a 1-separating-only code in G: the sets $I_{G,C,1}(v),\ v \in V$, are all different and nonempty, except for a unique vertex v_0, which is such that $I_{G,C,1}(v_0) = \emptyset$. Now $I_{G(x),C,1}(x) = C$ is different from all the other 1-identifying sets, because, G_0 being of size at least two, the codewords cannot be all in one of the p copies ($p \geq 2$); so C is 1-separating-only in $G(x)$. Using the right-hand-side inequality in (1), and (6), we have:

$$\chi_1^S(G(x)) \geq \gamma_1^{ID}(G(x)) - 1 = \chi_1^S(G),$$

which shows that if C is an optimal 1-separating-only code in G, then it is an optimal 1-separating-only code in $G(x)$ (and $\chi_1^S(G(x)) = \chi_1^S(G)$); this proves the inclusion $\mathcal{C}_4 \subseteq \mathcal{C}_3$. Such codes have size $p\gamma_1^{ID}(G_0) - 1$, as we already observed.

(viii) We now prove that $\mathcal{C}_3 \subseteq \mathcal{C}_4$. Let C be an optimal 1-separating-only code in $G(x)$. Then C does not contain x, C is 1-separating-only in G, and the same argument on cardinalities as before shows that it is optimal in G.

(ix) The equality (4): $\sigma(G(x)) = p\sigma_0(S_0)^{p-1}$ immediately follows. □

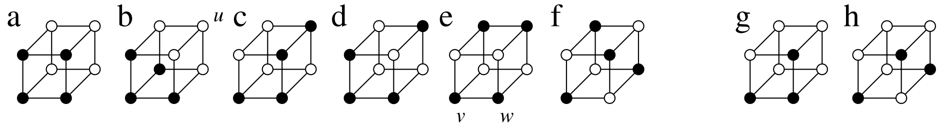

Fig. 1. (a)–(f): The three-dimensional cube G_0 and its different non-isomorphic 4-subsets; (g)–(h): the non-isomorphic 1-separating-only codes of size three in G_0.

We now study the properties of the cube, which will be used in subsequent constructions.

Lemma 10. *Let $G_0 = (V_0, E_0)$ be the three-dimensional cube. This graph admits $S_0 = 56$ different optimal 1-identifying codes and $\sigma_0 = 32$ different optimal 1-separating-only codes.*

Proof. It is straightforward to observe that $\gamma_1^{ID}(G_0) = 4$, and that, up to symmetries, the six different types of 4-subsets of V_0 are given in Fig. 1(a)–(f). If the black vertices represent the codewords, then (a), (c), (d) and (f) give 1-identifying codes, the numbers of which are 6, 24, 24 and 2, respectively; (b) and (e) give 8 and 6 codes, respectively, which are not 1-identifying: in (b), because u is not 1-covered by any codeword, and in (e) because, among others, v and w are not 1-separated by any codeword.

The optimal 1-separating-only codes have size three, and there are two types of them: see Fig. 1(g)–(h), where there are $(4 \cdot 6) + 8 = 32$ such optimal codes. □

3. The number of optimal *r*-identifying codes

3.1. The case $r = 1$

We are going to use the three-dimensional cube, studied in Lemma 10, as our first-level brick, with which we shall construct a graph consisting of k_1 copies of it together with a universal vertex x_1. This new graph will then be used as a second-level brick with which we shall construct a graph consisting of k_2 copies of it together with a universal vertex x_2, and so on. At each level, the graph is connected, we know the number of vertices and, thanks to Proposition 9, we know the number of optimal 1-identifying codes and the number of 1-separating-only codes, as well as the fact that the size of an optimal 1-identifying code is equal to the size of an optimal 1-separating-only code plus one, which allows us to go to the next level and compute again these numbers.

Our construction starts with k_1 copies of the cube, $k_1 \geq 2$, yielding a graph with $8k_1 + 1$ vertices, admitting at least

$$56^{k_1} + 32 \cdot k_1 \cdot 56^{k_1-1} \tag{7}$$

optimal 1-identifying codes of size $4k_1$ and

$$32 \cdot k_1 \cdot 56^{k_1-1} \tag{8}$$

optimal 1-separating-only codes of size $4k_1 - 1$. At the next level, we have, for $k_2 \geq 2$, a graph with $k_2(8k_1+1)+1$ vertices, admitting at least

$$(56^{k_1} + 32 \cdot k_1 \cdot 56^{k_1-1})^{k_2} + k_2(56^{k_1} + 32 \cdot k_1 \cdot 56^{k_1-1})^{k_2-1} \cdot 32 \cdot k_1 \cdot 56^{k_1-1} \tag{9}$$

optimal 1-identifying codes of size $4k_1k_2$ and

$$k_2(56^{k_1} + 32 \cdot k_1 \cdot 56^{k_1-1})^{k_2-1} \cdot 32 \cdot k_1 \cdot 56^{k_1-1} \tag{10}$$

optimal 1-separating-only codes of size $4k_1k_2 - 1$, and so on; we go on until we stop at a certain level h, where we know the number of vertices and the number of optimal 1-identifying codes, which depend on the k_j's, $1 \leq j \leq h$, so that we can, in principle, optimize h and the k_j's in order to obtain the graph where the number of optimal 1-identifying codes, written as $2^{\alpha \cdot n}$ where n is the order of the final graph, has the largest α.

Once this optimal graph, or a suboptimal one, is obtained, with a fixed order n_0, we can take p copies of it and obtain an infinite family of graphs, with order $n = pn_0$, admitting, by Lemma 8, $(2^{\alpha \cdot n_0})^p = 2^{\alpha \cdot n}$ optimal 1-identifying codes.

Already the expressions in (7)–(10) show that it is difficult, both analytically and computationally, to extract the best coefficient α. If we choose to consider ten levels, then the best is to take $k_1 = k_2 = \cdots = k_9 = 3$ and $k_{10} = 4$, for which we reach $\alpha \approx 0.77003$ (for 669,221 vertices). Finally, this graph can be made connected using a universal vertex without changing significantly the result on α, and so the following theorem holds.

Theorem 11. *There exist infinitely many connected graphs with n vertices admitting approximately $2^{0.77003 \cdot n}$ different optimal 1-identifying codes.* □

Please cite this article in press as: I. Honkala, et al., On the number of optimal identifying codes in a twin-free graph, Discrete Applied Mathematics (2014), http://dx.doi.org/10.1016/j.dam.2014.08.020

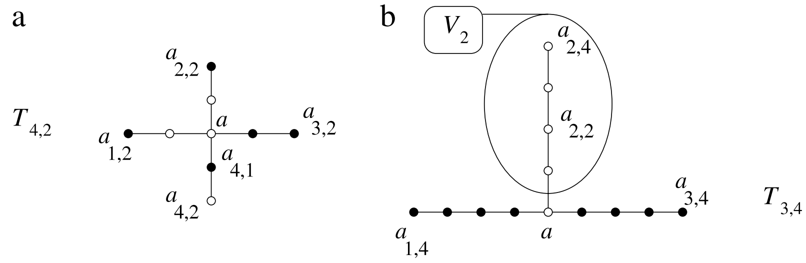

Fig. 2. The trees $T_{4,2}$ and $T_{3,4}$ in the proof of Theorem 13. Codewords are in black, non-codewords are in white. In (a), C is 2-identifying; in (b), C is not 4-identifying, only because $C \cap V_2 = \emptyset$, so $a_{2,4}$ is not 4-covered by C.

Remark 12. Assume that we have a graph G_0 with n_0 vertices, admitting $S_0 = 2^{\alpha \cdot n_0}$ optimal 1-identifying codes for some $\alpha \in [0, 1]$, and that G_0 also admits σ_0 optimal 1-separating-only codes, the size of which is assumed to be equal to $\gamma_1^{ID}(G_0) - 1$; actually, for our purpose, it is sufficient to have just one 1-separating-only code of size $\gamma_1^{ID}(G_0) - 1$. If we take p copies of G_0, we obtain a graph G with $n = pn_0$ vertices, admitting $2^{\alpha \cdot n}$ optimal 1-identifying codes. But if we add a universal vertex x to G, then, thanks to Proposition 9, the graph $G(x)$, whose order is pn_0+1, has $(S_0)^p + p\sigma_0(S_0)^{p-1} = 2^{\beta(n+1)}$ solutions; now for p large enough (namely $p > S_0/\sigma_0$), and because $\sigma_0 > 0$, we have:

$$n_0 \log_2\left(1 + p\frac{\sigma_0}{S_0}\right) > n_0 \geq \alpha n_0 = \log_2(S_0),$$

which implies that

$$\frac{p\log_2(S_0) + \log_2\left(1 + p\frac{\sigma_0}{S_0}\right)}{pn_0 + 1} > \frac{p\log_2(S_0)}{pn_0},$$

which in turn implies that

$$\frac{\log_2\left((S_0)^p + p\sigma_0(S_0)^{p-1}\right)}{n+1} > \frac{\log_2\left((S_0)^p\right)}{n},$$

i.e., $\beta > \alpha$. This means that, provided that some conditions on G_0 hold, we can always slightly improve on the coefficient α.

Note that these conditions are satisfied by the three-dimensional cube and by the graphs constructed using this cube in the way described above.

Open problem 1. Improve significantly on Theorem 11.

Open problem 2. Find a nontrivial upper bound on the number of different optimal 1-identifying codes that a graph can have.

One upper bound is obviously $\binom{n}{\gamma_1^{ID}(G)}$, for which an upper bound is $\binom{n}{\frac{n}{2}}$; for large n, the number $\binom{n}{\gamma_1^{ID}(G)}$ can be approximated by Stirling's formula, which uses the binary entropy of $\gamma_1^{ID}(G)/n$. This in turn gives an interval of the form $\mathcal{I} = [\frac{n}{2} - \lambda n, \frac{n}{2} + \lambda n]$ where $\gamma_1^{ID}(G)$ must lie if we want to obtain a coefficient better than $\alpha \approx 0.77003$. Unfortunately, this interval is too large to be of interest, since we obtain $\mathcal{I} \approx [0.22n, 0.78n]$. We conjecture however that the graphs G with maximum number of optimal 1-identifying codes have $\gamma_1^{ID}(G)$ close to $n/2$.

If we go back to the upper bound $2^{0.779n}$ for the number of different optimal 1-dominating codes mentioned at the end of the Introduction, obviously, because optimality is required, we cannot apply it directly to 1-identifying codes, although any r-identifying code is r-dominating. This bound is obtained through a recursive algorithm which lists all minimal 1-dominating codes of a graph on n vertices, and is based on induction; the technique seems extremely difficult to adapt to the case of 1-identifying codes.

3.2. The general case

Theorem 13. *For every $r \geq 2$ and $k \geq 3$, there exists a tree $T_{k,r}$, of order $n = kr + 1$, admitting*

$$S_{k,r} = k\left(\binom{k}{2}^{r-1} - \binom{k-1}{2}^{r-1}\right) \tag{11}$$

different optimal r-identifying codes, of size $2(r-1) + k - 1$.

Proof. The tree $T_{k,r}$ is constructed as follows (see Fig. 2): its set of vertices is the set

$$V_{k,r} = \{a_{i,j} : 1 \le i \le k,\ 1 \le j \le r\} \cup \{a\},$$

and its set of edges is

$$E_{k,r} = \bigcup_{i=1}^{k} \{\{a, a_{i,1}\}, \{a_{i,j}, a_{i,j+1}\} : 1 \le j \le r-1\}.$$

Note that $T_{k,r}$ can be seen as the star with its k branches subdivided $r-1$ times. We denote by V_i the set $\{a_{i,j} : 1 \le j \le r\}$. We say that a vertex $a_{i,j}$ is of level j. Therefore, the k leaves of the tree are the vertices of level r.
First, we observe that, in order to pairwise r-separate the k vertices of level one, any r-identifying code contains at least $k-1$ vertices of level r. Next, we can see that, once these $k-1$ leaves (or more) have been chosen as codewords in C,

(1) the vertex a is r-covered by C, and is r-separated by C from all the other vertices because $k \ge 3$;
(2) every vertex in V_i is r-separated by C from every vertex in V_ℓ, $i \ne \ell$.

So all what is left to do is:

(1) in every set V_i, to pairwise r-separate the vertices $v_{i,j}$, $1 \le j \le r$; when doing this, the code will necessarily r-cover all the vertices in V_i, except maybe the leaf.
(2) if necessary after this first step, to r-cover the leaf which may have not been taken in the code.

Next, using and adapting Lemma 5, we observe that, in order to r-separate the vertices inside V_i, it is sufficient to r-separate the $r-1$ pairs of consecutive vertices $a_{i,r}$ and $a_{i,r-1}$, $a_{i,r-1}$ and $a_{i,r-2}, \ldots, a_{i,2}$ and $a_{i,1}$. Moreover, in order to r-separate $a_{i,r-j+1}$ and $a_{i,r-j}$, $1 \le j \le r-1$, the only candidates are the vertices $a_{\ell,j}$, with $\ell \ne i$. Since this has to be done for every V_i, we have to choose two codewords with level j, for each $j \in \{1, \ldots, r-1\}$. This shows that $\gamma_r^{ID}(T_{k,r}) \ge 2(r-1) + k - 1$. Conversely, any code C which contains

(i) $k-1$ leaves,
(ii) two vertices of level j, for each j in $\{1, \ldots, r-1\}$, and
(iii) is such that there is at least one codeword in each set V_i (so that all leaves are r-covered by C)

is an (optimal) r-identifying code in $T_{k,r}$. We can satisfy (i) in k ways, (ii) in $\binom{k}{2}^{r-1}$ ways, whereas (iii) forbids the codes C with i_0 such that $C \cap V_{i_0} = \emptyset$. There are $k\binom{k-1}{2}^{r-1}$ such forbidden configurations, and therefore there are $k\left(\binom{k}{2}^{r-1} - \binom{k-1}{2}^{r-1}\right)$ different optimal r-identifying codes in $T_{k,r}$. $\square$

Corollary 14. *For every $p \ge 1$, $r \ge 2$ and $k \ge 3$, there exists a forest $F_{p,k,r}$, of order $n = p(kr+1)$, admitting*

$$S_{p,k,r} = 2^{\frac{n}{kr+1}\log_2\left(k\left(\binom{k}{2}^{r-1} - \binom{k-1}{2}^{r-1}\right)\right)} \tag{12}$$

different optimal r-identifying codes, of size $p(2(r-1)+k-1)$.

Proof. Simply consider p copies of the tree $T_{k,r}$ and apply Lemma 8. $\square$

For $r = 2$, equality (12) reduces to $S_{p,k,2} = 2^{\frac{n}{2k+1}\log_2(k^2-k)}$, whose maximum over integers is reached for $k = 4$, yielding

$$S_{p,4,2} = 2^{\frac{\log_2 12}{9}n} \approx 2^{0.398n}$$

different optimal 2-identifying codes.

For $r = 3$, equality (12) reaches its maximum over integers for $k = 5$, yielding

$$S_{p,5,3} = 2^{\frac{\log_2 320}{16}n} \approx 2^{0.520n}$$

different optimal 3-identifying codes.

Table 1 shows that for the first values of r but $r = 2$, $k = 5$ provides for the maximum number of r-identifying codes in $F_{p,k,r}$.

However, as we are now going to see, for all values of r, including $r = 1$, we can come as close to $2^{\frac{1+\log_2 5}{5}n} \approx 2^{0.664n}$ as we want; to do this, we first carry out computations on the exponent of 2 in (12) for a fixed k and a growing r: if $\alpha = \frac{n}{kr+1}\log_2\left(k\left(\binom{k}{2}^{r-1} - \binom{k-1}{2}^{r-1}\right)\right)$, then

$$\begin{aligned}
\alpha &= \frac{n}{kr+1}\log_2\left(\frac{k(k-1)^{r-1}}{2^{r-1}}\left(k^{r-1} - (k-2)^{r-1}\right)\right) \\
&= \frac{n}{kr+1}\log_2\left(\frac{k^r(k-1)^{r-1}}{2^{r-1}}\left(1 - \left(\frac{k-2}{k}\right)^{r-1}\right)\right) \\
&= \frac{n}{kr+1}\log_2\frac{k^r(k-1)^{r-1}}{2^{r-1}} + n\frac{\log_2\left(1 - \left(\frac{k-2}{k}\right)^{r-1}\right)}{kr+1}.
\end{aligned}$$

Table 1
For each value of k and r ($2 \le r \le 6$, $r = 10$, and $3 \le k \le 6$) and with $p = \frac{n}{kr+1}$, we give in the first line of the cell the number $S_{k,r}$ given by (11), and an approximation of the number $S_{p,k,r}$, as given by (12), in the second line. A bullet indicates the highest value of $S_{p,k,r}$ for a given r.

r	$k=3$	$k=4$	$k=5$	$k=6$
2	6 $2^{0.369n}$	12 • $2^{0.398n}$	20 $2^{0.392n}$	30 $2^{0.378n}$
3	24 $2^{0.459n}$	108 $2^{0.5196n}$	320 • $2^{0.5202n}$	750 $2^{0.502n}$
4	78 $2^{0.483n}$	756 $2^{0.563n}$	3920 • $2^{0.568n}$	14 250 $2^{0.553n}$
5	240 $2^{0.494n}$	4860 $2^{0.583n}$	43 520 • $2^{0.593n}$	243 750 $2^{0.577n}$
6	726 $2^{0.501n}$	30 132 $2^{0.596n}$	461 120 • $2^{0.607n}$	3 956 250 $2^{0.592n}$
10	59 046 $2^{0.511n}$	40 232 052 $2^{0.617n}$	4 949 611 520 • $2^{0.632n}$	224 660 156 250 $2^{0.618n}$

Now we can see that for every $\varepsilon_1 > 0$, there exists r_1 such that for all $r \ge r_1$, we have

$$\alpha > \frac{n}{kr+1}\log_2\left(\frac{k^r(k-1)^{r-1}}{2^{r-1}}\right) - n\varepsilon_1 = \frac{n}{kr+1}\left(r\log_2 k + (r-1)\log_2\frac{k-1}{2}\right) - n\varepsilon_1$$
$$= n\frac{\log_2 k + \log_2(k-1) - 1}{k} - n\left(\frac{\log_2 k}{k(kr+1)} + \frac{k+1}{k(kr+1)}\log_2\frac{k-1}{2}\right) - n\varepsilon_1.$$

For every $\varepsilon_2 > 0$, there exists r_2 such that for all $r \ge \max(r_1, r_2)$, we have

$$\alpha > n\frac{\log_2 k + \log_2(k-1) - 1}{k} - n\varepsilon_1 - n\varepsilon_2.$$

The factor $\frac{\log_2 k+\log_2(k-1)-1}{k}$ of n is the largest for $k = 5$, when it is equal to $\frac{1+\log_2 5}{5} \approx 0.664$, so that we have established the following result.

Theorem 15. *For every real $\varepsilon > 0$, there exists an integer r_0 such that for every integer $r \ge r_0$, there exist infinitely many forests with $n = p(5r+1)$ vertices, admitting $\lfloor 2^{(\frac{1+\log_2 5}{5}-\varepsilon)n}\rfloor$ different optimal r-identifying codes.* □

The main result of this section is now simply a consequence of Theorem 15, valid for all values of r, including $r = 1$, and can be obtained by connected graphs.

Corollary 16. *Let $r \ge 1$ be an integer and $\varepsilon > 0$ be a real. There exist infinitely many connected graphs with n vertices admitting $\lfloor 2^{(\frac{1+\log_2 5}{5}-\varepsilon)n}\rfloor$ different optimal r-identifying codes.*

Proof. Let R_0 satisfy the conditions of Theorem 15 with $p = 1$: for $R \ge R_0$, there exists a tree T, i.e., a connected graph, with order $n = 5R+1$, admitting $\lfloor 2^{(\frac{1+\log_2 5}{5}-\varepsilon)n}\rfloor$ optimal R-identifying codes. Choose $R \ge R_0$ a multiple of r: $R = qr$, and consider the q-th transitive closure of T, T^q; then T^q is connected, has $n = 5R+1 = 5qr+1$ vertices and admits $\lfloor 2^{(\frac{1+\log_2 5}{5}-\varepsilon)n}\rfloor$ optimal r-identifying codes, since by Lemma 4, any R-identifying code in T is r-identifying in T^q when $R = qr$, and vice versa. □

Open problem 3. Improve on Corollary 16.

Open problem 4. Find a nontrivial upper bound on the number of different optimal r-identifying codes that a graph can have.

Note that any upper bound for $r = 1$ (cf. Open problem 2) is an upper bound for $r \ge 1$, and more generally, any upper bound for r_0 is an upper bound for all the multiples of r_0, thanks to Lemma 4.

Acknowledgement

The first author gratefully acknowledges the financial support of the Finnish Cultural Foundation.

References

[1] C. Berge, Graphes, Gauthier-Villars, Paris, 1983, English translation: Graphs, North-Holland Publishing Co.: Amsterdam, 1985.
[2] N. Bertrand, I. Charon, O. Hudry, A. Lobstein, Identifying and locating-dominating codes on chains and cycles, European J. Combin. 25 (2004) 969–987.
[3] I. Charon, O. Hudry, A. Lobstein, Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard, Theoret. Comput. Sci. 290 (2003) 2109–2120.
[4] J.F. Couturier, P. Heggernes, P. Van't Hof, D. Kratsch, Minimal dominating sets in graph classes: combinatorial bounds and enumeration, in: Lecture Notes in Computer Science, vol. 7147, 2012.
[5] R. Diestel, Graph Theory, Springer-Verlag, Berlin, 2005.
[6] F.V. Fomin, F. Grandoni, A.V. Pyatkin, A.A. Stepanov, Combinatorial bounds via measure and conquer: bounding minimal dominating sets and applications, ACM Trans. Algorithms 5 (2008) Article 9.
[7] S. Gravier, J. Moncel, A. Semri, Identifying codes of Cartesian product of two cliques of the same size, Electron. J. Combin. 15 (1) (2008) N4.
[8] T.W. Haynes, S.T. Hedetniemi, P.J. Slater, Fundamentals of Domination in Graphs, Marcel Dekker, New York, 1998.
[9] M.A. Henning, A. Yao, Total Domination in Graphs, Springer-Verlag, New York, 2013.
[10] I. Honkala, O. Hudry, A. Lobstein, On the ensemble of optimal dominating and locating-dominating codes in a graph, submitted for publication.
[11] I. Honkala, O. Hudry, A. Lobstein, On the ensemble of optimal identifying codes in a twin-free graph, submitted for publication.
[12] M.G. Karpovsky, K. Chakrabarty, L.B. Levitin, On a new class of codes for identifying vertices in graphs, IEEE Trans. Inform. Theory IT-44 (1998) 599–611.
[13] A. Lobstein, A bibliography on watching systems, identifying, locating-dominating and discriminating codes in graphs. http://perso.telecom-paristech.fr/~lobstein/debutBIBidetlocdom.pdf.
[14] M. Pelto, A definition of uniqueness for optimal identifying and covering codes in infinite lattices (2012) submitted for publication.
[15] M. Pelto, On identifying and locating-dominating codes in the infinite king grid, Ph.D. Thesis, University of Turku, Finland, 2012, 133pp.

Please cite this article in press as: I. Honkala, et al., On the number of optimal identifying codes in a twin-free graph, Discrete Applied Mathematics (2014), http://dx.doi.org/10.1016/j.dam.2014.08.020