Advances in Mathematics of Communications
Volume X, No. 0X, 200X, X–XX Web site: http://www.aimSciences.org

DISCRIMINATING CODES IN BIPARTITE GRAPHS: BOUNDS, EXTREMAL CARDINALITIES, COMPLEXITY

Emmanuel Charbit

Institut TELECOM - TELECOM ParisTech
46, rue Barrault, 75634 Paris Cedex 13 - France

Irène Charon

Institut TELECOM - TELECOM ParisTech
& Centre National de la Recherche Scientifique - LTCI UMR 5141
46, rue Barrault, 75634 Paris Cedex 13 - France

Gérard Cohen

Institut TELECOM - TELECOM ParisTech
& Centre National de la Recherche Scientifique - LTCI UMR 5141
46, rue Barrault, 75634 Paris Cedex 13 - France

Olivier Hudry

Institut TELECOM - TELECOM ParisTech
& Centre National de la Recherche Scientifique - LTCI UMR 5141
46, rue Barrault, 75634 Paris Cedex 13 - France

Antoine Lobstein

Centre National de la Recherche Scientifique - LTCI UMR 5141
& Institut TELECOM - TELECOM ParisTech
46, rue Barrault, 75634 Paris Cedex 13 - France

(Communicated by Aim Sciences)

Abstract. Consider an undirected bipartite graph $G = (V = I \cup A, E)$, with no edge inside I nor A. For any vertex $v \in V$, let $N(v)$ be the set of neighbours of v. A code $C \subseteq A$ is said to be discriminating if all the sets $N(i) \cap C$, $i \in I$, are nonempty and distinct.

We study some properties of discriminating codes. In particular, we give bounds on the minimum size of these codes, investigate graphs where minimal discriminating codes have size close to the upper bound, or give the exact minimum size in particular graphs; we also give an NP-completeness result.

1. Introduction

We consider undirected graphs, in which a *code* is simply a subset of vertices. We define different classes of codes which, in various meanings, can help to unambiguously determine a vertex. The motivations may come from processor networks where we wish to locate a faulty vertex under certain conditions, or from the need to identify an individual, given its set of attributes.

2000 *Mathematics Subject Classification:* Primary: 05C99; Secondary: 94B65, 05C35, 68Q17.
Key words and phrases: Graph theory, bipartite graphs, discriminating codes, identifying codes, locating-dominating codes, separating codes, complexity.

This work was done during a stay of the first author at TELECOM ParisTech.

Let $G = (V, E)$ be an undirected graph. For any vertex $v \in V$, let $N(v)$ denote the neighbourhood of v and $B_1(v) = N(v) \cup \{v\}$. More generally, for an integer $r \geq 1$, we define $B_r(v)$, the *ball* of radius r centred at v, as the set of vertices within distance r from v, where the distance $d(x, y)$ between two vertices x and y is the smallest possible number of edges in any path between them. Whenever two vertices v_1 and v_2 are such that $v_1 \in B_r(v_2)$ (or, equivalently, $v_2 \in B_r(v_1)$), we say that they r-*cover* each other. A set $X \subseteq V$ r-covers a set $Y \subseteq V$ if every vertex in Y is r-covered by at least one vertex in X.

The elements of a code $C \subseteq V$ are called *codewords*. For each vertex $v \in V$, we denote by

$$K_{C,r}(v) = C \cap B_r(v)$$

the set of codewords r-covering v. Two vertices v_1 and v_2 with $K_{C,r}(v_1) \neq K_{C,r}(v_2)$ are said to be r-*separated* by code C.

We now give several definitions for a code C, for which there are two particular sets of vertices, A and I: C must be included in A, and all vertices in I must be r-covered and pairwise r-separated by C. More specifically:

A code $C \subseteq V$ is called r-*locating-dominating* [14] if all the sets $K_{C,r}(v)$, $v \in V \setminus C$, are nonempty and distinct. In other words, $A = V$ and $I = V \setminus C$.

A code $C \subseteq V$ is called r-*identifying* [13] if all the sets $K_{C,r}(v)$, $v \in V$, are nonempty and distinct. In other words, $A = V$ and $I = V$.

For an abundant literature on r-locating-dominating and r-identifying codes, see [15]. The next definition is more recent.

We consider a bipartite graph $G = (V = I \cup A, E)$, with no edge inside A nor I. A code $C \subseteq A$ is called r-*discriminating* [5] if all the sets $K_{C,r}(v)$, $v \in I$, are nonempty and distinct.

Since any distance between a vertex in I and a vertex in A is odd, we shall only consider r-discriminating codes with odd r. Anyway, we shall deal with general values of r in Section 5 only, and before that, we shall concentrate on 1-discriminating codes (which we shall simply call discriminating when there is no ambiguity). In this setting, I can be viewed as a set of *individuals* and A as a set of *attributes*, with an edge between $i \in I$ and $a \in A$ if i owns a; a discriminating code C is a set of attributes such that no two individuals have the same set of attribute codewords, and each individual has a nonempty set of attribute codewords. If we drop the latter condition, we say that C is a *separating* code; in this case, at most one individual has the empty set as its set of attribute codewords, and in a given graph the cardinalities of the smallest separating and discriminating codes differ by at most one. Separating codes are of practical interest for the study of discriminating codes, because the two notions are very close, and separating codes present a nice symmetry with respect to the empty set and the whole set of attributes, that discriminating codes do not possess.

Discriminating codes do not always exist. A necessary and sufficient condition for existence is that any individual i has a nonempty neighbourhood and any two distinct individuals i_1, i_2 have distinct neighbourhoods. Indeed, if this condition holds, then $C = A$ is discriminating, and if not, then for all codes $C \subseteq A$, we have $K_{C,1}(i_1) = N(i_1) \cap C = N(i_2) \cap C = K_{C,1}(i_2)$ or $K_{C,1}(i) = \emptyset$.

One of the main issues now is, given a bipartite graph G and an integer r, to find r-discriminating codes with minimum cardinalities — if they exist. This problem is polynomial for instance when $r = 1$ and the graph is a tree [6], but in general, it is NP-hard (see [5], [6] for the case $r = 1$ and Section 5 here for the general case).

In Section 2, we give various bounds on the minimum cardinality of a 1-discriminating code, using 1-separating codes as a tool. In Section 3, we study graphs in which the minimal 1-discriminating codes have a size which is equal or close to the upper bound; from this, we also derive a result on 1-identifying codes in Subsection 3.4. We study particular graphs or families of graphs in Section 4, and the aforementioned result on complexity, with its full proof, is given in the last section, Section 5.

2. Bounds on 1-Discriminating Codes

Let $G=(V=I\cup A,E)$ be an undirected, bipartite graph, with $E\subseteq\{\{i,a\}:i\in I,a\in A\}$. Denote by $\overline{G}=(V,\overline{E})$ the "complementary" graph with $\overline{E}=\{\{i,a\}:i\in I,a\in A\}\setminus E$. Then obviously a code is separating in G if and only if it is separating in $\overline{G}$.

2.1. Lower Bounds. Let $G=(V=I\cup A,E)$ be a bipartite graph, and let C be a discriminating code in G. A first trivial bound is that

$$|C|\geq\lceil\log_2(|I|+1)\rceil, \tag{1}$$

and this bound can be achieved, for instance by a graph where A has n elements, I has 2^n-1 elements representing the nonempty subsets of $[1,n]$ and an edge links an attibute a_k to the subsets containing k.

Let $\alpha_{\max}=\max\{\deg(a):a\in A\}$ and $\alpha_{\min}=\min\{\deg(a):a\in A\}$. If $\alpha_{\max}=\alpha_{\min}=\alpha$, we say that G is *attribute-regular* with degree α.

Theorem 1. *Let C be a discriminating code in a bipartite graph G. Then*

$$|C|\geq\max\left\{\frac{2|I|}{\alpha_{\max}+1},\frac{2|I|}{|I|-\alpha_{\min}+2}\right\}. \tag{2}$$

As a consequence, if G is attribute-regular with degree α, then

$$|C|\geq\max\left\{\frac{2|I|}{\alpha+1},\frac{2|I|}{|I|-\alpha+2}\right\}. \tag{3}$$

Proof. Because C is discriminating in G, at most $|C|$ individuals can be 1-covered by exactly one codeword, and the other individuals are 1-covered by at least two codewords. Therefore $|C|\times\alpha_{\max}\geq 1\times|C|+2\times(|I|-|C|)$, or:

$$|C|\geq\frac{2|I|}{\alpha_{\max}+1}. \tag{4}$$

A similar counting argument for a separating code D leads to $|D|\times\alpha_{\max}\geq 0\times 1+1\times|D|\;+\;2\times(|I|-|D|-1)$, or:

$$|D|\geq\frac{2|I|-2}{\alpha_{\max}+1}. \tag{5}$$

Now consider the bipartite graph $G^*=(V^*=I^*\cup A,E)$ where $I^*=I\cup\{i\}$, which is simply obtained from G by adding an isolated individual i. The code C is separating in G^*, and therefore also in the complementary graph $\overline{G^*}$, which has $|I|+1$ individuals and maximum degree $|I|+1-\alpha_{\min}$. Using (5), we obtain:

$$|C|\geq\frac{2|I|}{|I|-\alpha_{\min}+2}, \tag{6}$$

which, together with (4), yields (2). Trivially, (3) follows. □

Inequality (4) is tight: consider *perfect matchings*, i.e., graphs with $2p$ vertices, $I = \{i_k : 1 \leq k \leq p\}$, $A = \{a_k : 1 \leq k \leq p\}$, and $E = \{\{i_k, a_k\} : 1 \leq k \leq p\}$; here $\alpha_{\max} = \alpha = 1$ and $C = A$ is the only discriminating code, of size $|I|$. See also Section 4.1. Inequalities (4) and (6) are simultaneously achieved in the complementary graph of a perfect matching with $p = 3$: here, $|I| = 3, \alpha_{\max} = \alpha_{\min} = \alpha = 2$, and any set with two attributes is a discriminating code.

2.2. UPPER BOUNDS. We first give a lemma which characterizes minimal discriminating codes.

Lemma 1. *A discriminating code $C \subseteq A$ is minimal for inclusion if and only if*

$$\forall c \in C, \quad \begin{array}{c} \exists i, j \in I \ : \ \big(N(i) \cap C\big) \Delta \big(N(j) \cap C\big) = \{c\} \\ or \quad \exists k \in I \ : \ N(k) \cap C = \{c\} \end{array}, \tag{7}$$

where Δ stands for symmetric difference.

Proof. If condition (7) holds, then removing any codeword from C leads to a code which is not discriminating anymore: i and j are not 1-separated by the code anymore, or k is not 1-covered by the code anymore. If condition (7) is violated by some $c_0 \in C$, then $C \setminus \{c_0\}$ is still discriminating. $\square$

Define $\mathbf{C}$ the *incidence matrix* of a code C, whose $|C|$ rows represent the codewords and $|I|$ columns the individuals. We use Lemma 1 to give an upper bound on $|C|$ with a very short proof.

Lemma 2. *If a discriminating code $C \subseteq A$ is minimal for inclusion, then the rows of $\mathbf{C}$ are linearly independent over any field, and*

$$|C| = \mathrm{rank}(\mathbf{C}) \leq \mathrm{rank}(\mathbf{A}). \tag{8}$$

As a consequence,

$$|C| \leq |I|. \tag{9}$$

Proof. By Lemma 1, for any row c in $\mathbf{C}$, there exist two columns i, j in $\mathbf{C}$ differing *exactly* on c, or there is one column k in $\mathbf{C}$ whose *only* "1" is on row c. Therefore any combination of rows involving c leads to a nonzero configuration on columns i, j, or on column k. The inequalities follow because $\mathrm{rank}(\mathbf{C}) \leq \mathrm{rank}(\mathbf{A}) \leq |I|$. $\square$

Note that the inequality $|C| \leq |I|$ is tight, since it is again reached in a perfect matching: if $I = \{i_k : 1 \leq k \leq p\}$, $A = \{a_k : 1 \leq k \leq p\}$, $E = \{\{i_k, a_k\} : 1 \leq k \leq p\}$, then $C = A$ is the only discriminating code.

From G and a separating code C which is minimal for inclusion, we now define a new graph, $H(G, C)$, which will be very important in the sequel. First we state the following lemma, which can be proved using exactly the same kind of argument as for Lemma 1.

Lemma 3. *A separating code $C \subseteq A$ is minimal for inclusion if and only if*

$$\forall c \in C, \ \ \exists i, j \in I \ : \ \big(N(i) \cap C\big) \Delta \big(N(j) \cap C\big) = \{c\}. \tag{10}$$

With this lemma, and following [3], we can define, for a minimal separating code C, the graph $H(G, C)$ in the following way: $H(G, C)$ has vertex set I, and for every codeword $c \in C$, we take exactly one pair of vertices $i, j \in I$ which is such that $\big(N(i) \cap C\big) \Delta \big(N(j) \cap C\big) = \{c\}$, and we put in $H(G, C)$ the edge $\{i, j\}$ labelled by c.

For short, we shall call such an edge a c-edge. Note that in general, $H(G,C)$ is not necessarily unique. When there is no ambiguity on G and C, we shall denote $H = H(G,C)$.

Example 1. Let G be the following graph with $2p+1$ vertices:

$$I = \{i_k : 0 \leq k \leq p\},\ A = \{a_k : 1 \leq k \leq p\},\ E = \{\{i_k, a_k\} : 1 \leq k \leq p\};$$

in other words, we have added an isolated individual, i_0, to a perfect matching. Then $C = A$ is the only separating code, and there is only one graph H, where i_0 is linked to every $i_k, 1 \leq k \leq p$, by the a_k-edge. Such a graph is called a *star* with centre i_0.

Lemma 4. [3] *There is no cycle in $H(G,C)$.*

Proof. Assume on the contrary that there is a cycle $i_1, i_2, \dots, i_\ell, i_1$, where the edge $\{i_k, i_{k+1}\}$, $1 \leq k \leq \ell-1$, is labelled by c_k and the edge $\{i_\ell, i_1\}$ is labelled by c_ℓ. This means that $(N(i_k)\cap C)\Delta(N(i_{k+1})\cap C) = \{c_k\}$ and $(N(i_\ell)\cap C)\Delta(N(i_1)\cap C) = \{c_\ell\}$.

Consider the c_1-edge $\{i_1, i_2\}$. Without loss of generality, we can assume that $c_1 \in N(i_1)$, $c_1 \notin N(i_2)$. Because

$$c_1 \neq c_2 \ \text{ and } \ \{c_2\} = (N(i_2) \cap C)\Delta(N(i_3) \cap C),$$

c_1 does not belong to $N(i_3)$ either. Repeating this argument leads to $c_1 \notin N(i_4)$, $\dots$, $c_1 \notin N(i_\ell)$ and ultimately $c_1 \notin N(i_1)$, a contradiction. □

Corollary 1. *If C is a separating code which is minimal for inclusion, then*

$$|C| \leq |I| - 1. \tag{11}$$

Proof. A graph H has $|I|$ vertices, $|C|$ edges, and we have just proved that H is a forest. □

The bound (11) is tight: consider the graph G in Example 1. On the way, Corollary 1 obviously implies the bound for discriminating codes $|C| \leq |I|$ in Lemma 2.

Lemma 5. *Consider, in a connected component T of the graph $H(G,C)$, the c-edge $\{i,j\}$, with $c \in N(i)$, $c \notin N(j)$.*

(i) Then, in this component, all vertices v such that $d(i,v) < d(j,v)$ belong to $N(c)$ and no vertex w such that $d(i,w) > d(j,w)$ belongs to $N(c)$. In other words, a c-edge divides the component into two parts: on one side of the edge, all vertices belong to $N(c)$, on the other side, no vertex does.

(ii) In any of the connected components not containing the c-edge $\{i,j\}$, either all the vertices belong to $N(c)$ or no vertex belongs to $N(c)$.

Proof. Consider in T a neighbour v of i, $v \neq j$ (respectively, a neighbour w of j, $w \neq i$). Then, because $c \in N(i)$ and $(N(i)\cap C)\Delta(N(v)\cap C)$ is a singleton not equal to $\{c\}$, we have $c \in N(v)$. Similarly, $c \notin N(w)$. Step by step, we obtain (i).

For (ii), consider two vertices v, w in a connected component $S \neq T$. If $v \in N(c)$, using the above argument leads, step by step, to $w \in N(c)$; similarly, if $v \notin N(c)$, then $w \notin N(c)$. □

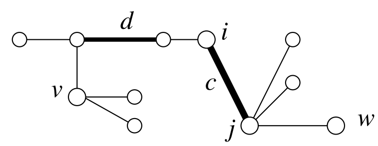


FIGURE 1. The tree H

3. CODES WITH SIZES CLOSE TO THE UPPER BOUND

We have just seen in the previous section that a minimum discriminating code has at most $|I|$ elements. In this section, we are interested in characterizing "bad graphs", that is, graphs in which any discriminating code has a "large" size, namely $|I|$ or $|I|-1$.

What we actually do is to consider graphs where $|A|=|I|$ or $|A|=|I|-1$, and A is the only discriminating code. To do this, we first study, in Subsections 3.1 and 3.2, graphs where $|A|=|I|-1$ or $|A|=|I|-2$, and A is the only separating code, from which we derive results on discriminating codes in Subsection 3.3. Ultimately, in Subsection 3.4, we shall also obtain results on graphs which are "bad" with respect to identifying codes.

Let $G=(V=I\cup A,E)$ be a bipartite graph such that $|A|=|I|-1$ or $|I|-2$ and $C=A$ is the only *separating* code.

3.1. THE CASE $|A|=|I|-1$.

Since $|C|=|I|-1$, the graph H is a tree, and we can polish Lemma 3.

Lemma 6. *For every codeword $c\in C$, there exists a* unique *pair i,j of individuals such that*

$$\big(N(i)\cap C\big)\Delta\big(N(j)\cap C\big)=\{c\}.$$

Proof. Suppose on the contrary that we have two distinct pairs $\{i,j\}$ and $\{v,w\}$ such that

$$\big(N(i)\cap C\big)\Delta\big(N(j)\cap C\big)=\big(N(v)\cap C\big)\Delta\big(N(w)\cap C\big)=\{c\},$$

and construct the graph H where the c-edge is $\{i,j\}$ — see Figure 1.

We can assume, without loss of generality, that $N(c)$ contains i and v, and neither j nor w. By Lemma 5(i), in H the vertices i,v are on one side of the c-edge, and j,w are on the other side. We assume that, for instance, $i\neq v$, which means that the shortest path between i and v contains at least one d-edge. Now v is on one side of this edge, and w is on the other side, so that d also belongs to $(N(v)\cap C)\Delta(N(w)\cap C)$, a contradiction. □

Corollary 2. *The tree H is unique.*

Corollary 3. *The number of labelled bipartite graphs with $|I|=n$ individuals and admitting a minimal separating code C such that $|A|=|C|=n-1$ is equal to*

$$n^{n-2}\times(n-1)!\times 2^{n-1}. \tag{12}$$

Proof. Cayley's formula [4] states that there are n^{n-2} trees with n labelled vertices; this in turn generates $n^{n-2}(n-1)!$ trees with n labelled vertices and $n-1$ labelled edges. Each of these doubly labelled trees can be considered as a graph H. We have seen in Corollary 2 that two different trees cannot be produced by the same

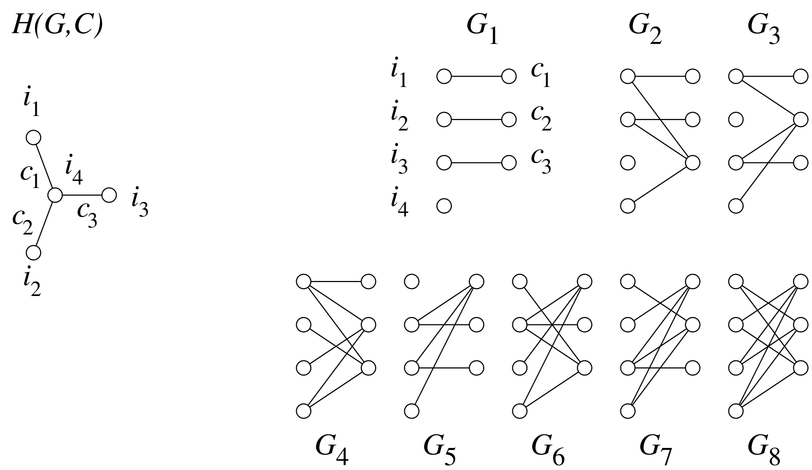


Figure 2. The star with four vertices as graph $H(G, C)$

graph. Finally, from such a doubly labelled tree we can recover 2^{n-1} graphs G, where the set of n individuals is labelled and the set of $n-1$ attributes is labelled, in the following way: we know how each of the $n-1$ c-edges divides H into two parts, the vertices which belong to $N(c)$ and those which do not. All we have to do is a binary choice, vertex by vertex and independently, in order to determine which part of I we link to c in G and which part we do not.

By construction, these $2^n - 1$ graphs all give the same graph H, and they are such that $C = A$ is the only separating code: a code $C^* \neq A$ would have at most $|I| - 2$ elements and its graph $H(G, C^*)$ would not be a tree, but a forest with at least two components. □

Figures 2 and 3 illustrate how to recover 8 graphs G when the graph H is the star with centre i_4 and leaves i_1, i_2, i_3, or the path i_1, i_2, i_3, i_4. For instance, in Figure 2, in the first four graphs, G_1, G_2, G_3, G_4, the vertex i_1 is linked to c_1, and it is not in the last four.

If we restrict ourselves to attribute-regular graphs with degree α, which implies that all codewords c have $|N(c)| = \alpha$, then the number of graphs given by (12) reduces and actually we can describe all these graphs. In H, we consider a leaf i_1, linked to a vertex i_2 by a c_1-edge. We distinguish between two cases.

(i) $c_1 \in N(i_1)$; then, by Lemma 5(i), no vertex other than i_1 belongs to $N(c_1)$, i.e., $N(c_1) = \{i_1\}$ and $\alpha = 1$.

(ii) $c_1 \in N(i_2)$; then, by Lemma 5(i), all vertices, except i_1, belong to $N(c_1)$, i.e., $N(c_1) = I \setminus \{i_1\}$ and $\alpha = |I| - 1$.

For our purposes, we use a restricted definition of isomorphism: two bipartite graphs

$$G_1 = (V_1 = I_1 \cup A_1, E_1) \text{ and } G_2 = (V_2 = I_2 \cup A_2, E_2)$$

are said to be *isomorphic* if there is a one-to-one onto function $f : V_1 \to V_2$ such that $f(I_1) = I_2, f(A_1) = A_2$, and $\{u, v\} \in E_1$ if and only if $\{f(u), f(v)\} \in E_2$; then we can conclude with the following theorem.

Theorem 2. *The only bipartite attribute-regular graphs $(I \cup A, E)$ with $|I| = n$ individuals and admitting a minimal separating code C such that $|A| = |C| = n - 1$*

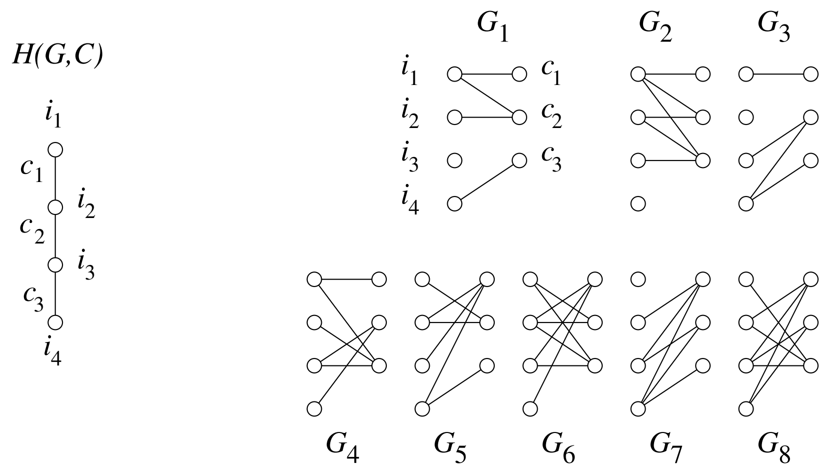


FIGURE 3. The path with four vertices as graph $H(G,C)$

are isomorphic to the graph G consisting of a perfect matching plus an isolated individual, and its complementary graph $\overline{G}$.

This graph G was described in Example 1. See the graphs G_1 and $G_8 = \overline{G_1}$ in the example of Figure 2 when $n = 4$.

3.2. THE CASE $|A| = |I| - 2$. In this subsection, we assume that the graph G is attribute-regular with degree α — otherwise, the investigation gets too complicated. So we have: $\forall c \in A, |N(c)| = \alpha$. It is straightforward to see that the case $\alpha = 1$, and, using complementation, the case $\alpha = |I| - 1$, are impossible. So from now on, $2 \leq \alpha \leq |I| - 2$.

Since $|C| = |I| - 2$, the graph H is a forest with two connected components, S and T, with $|S| + |T| = |I|$. We distinguish between two cases.

(i) $|S| \geq 2, |T| \geq 2$. Consider in S a leaf i_S together with its c_S-edge.

If $i_S \in N(c_S)$, then, using Lemma 5 and the fact that $\alpha \geq 2$, all vertices in T belong to $N(c_S)$, and $\alpha = |T| + 1$. Consider in T a leaf i_T together with its c_T-edge. If $i_T \notin N(c_T)$, then either $|N(c_T)| = |T| - 1 = \alpha - 2$ or $|N(c_T)| = |T| - 1 + |S| = |I| - 1 > \alpha$, in both cases a contradiction. Therefore $i_T \in N(c_T)$, $S \subset N(c_T)$, and $\alpha = |S| + 1$, and finally

$$|S| = |T| = \alpha - 1 = \frac{|I|}{2} = |I| - \alpha + 1.$$

We now show that S is a star (and so is T). Assume on the contrary that in S we have a path i_1, i_2, i_3, i_4, with the c-edge $\{i_2, i_3\}$ and, without loss of generality, $i_1, i_2 \in N(c)$, $i_3, i_4 \notin N(c)$. If $T \subset N(c)$, then $|N(c)| \geq |T| + 2 = \alpha + 1$, and if $T \cap N(c) = \emptyset$, then $|N(c)| \leq |S| - 2 = \alpha - 3$, in both cases a contradiction.

So the graph $H(G,C)$ consists of two stars S and T with $\alpha - 1$ vertices, and inside each star, any c-edge $\{i, j\}$, where i is the centre and j is a leaf, is such that the leaf j and all the $\alpha - 1$ vertices in the other star belong to $N(c)$. We are now able to build the corresponding family of graphs G with parameter α or $|I|$, see Figure 4, with $|I_1| = |I_2| = |A_1| + 1 = |A_2| + 1 = \alpha - 1$.

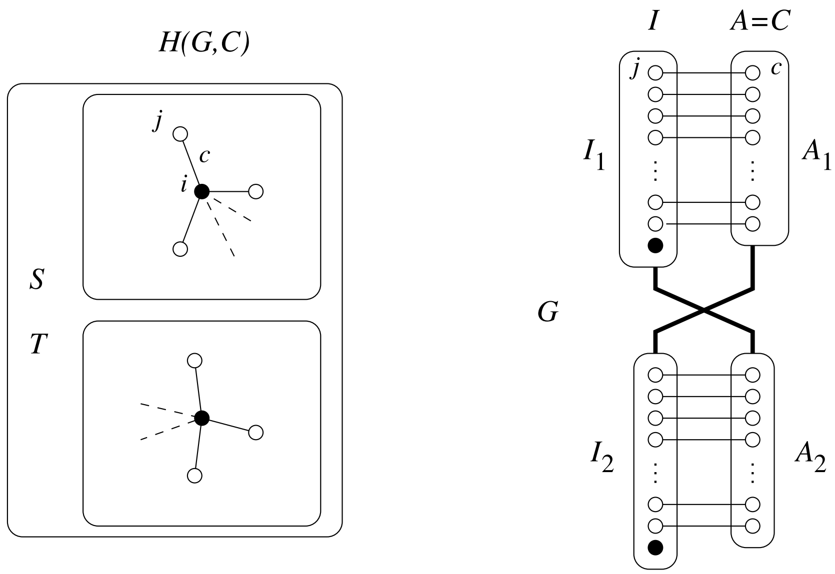


FIGURE 4. The union of two stars of same size as graph $H(G, C)$; the bold lines mean that all possible edges exist between I_1 and A_2, and between I_2 and A_1

FIGURE 5. Trees containing paths i_1, i_2, i_3, i_4, i_5 or i_1, i_2, i_3, i_4 can be ruled out by examining the d-edge

If $i_S \notin N(c_S)$, then we consider $\overline{G}$, the complementary graph of G, and we are back to case (i). We can conclude that we shall obtain the graphs which are complementary to the graphs described in Figure 4.

(ii) $|T| = 1$, $|S| = |I| - 1$. We set $T = \{t\}$. Consider in S a leaf i together with its c-edge.

If $i \in N(c)$, then, using Lemma 5 and $\alpha \geq 2$, $t \in N(c)$ and $\alpha = 2$.

Now there is no path $i_1, i_2, i_3, i_4, i_5, i_6$ in S, otherwise, because of the d-edge $\{i_3, i_4\}$, there would be at least three vertices, i_1, i_2, i_3 or i_4, i_5, i_6, in $N(d)$, contradicting $|N(d)| = \alpha = 2$. Applying the same argument to trees containing paths of length 4 or less, we can rule out the two trees given in Figure 5, and conclude that the only possibility left is that S is an *extended star*, consisting of a centre z and branches of length one, z, i, or length two, z, v, w, see Figure 6. Let p be the number of leaves linked to z, and q the number of leaves not linked to z ($p \geq 0$, $q \geq 0$, $2q + p = |S| - 1$); then q is also the number of vertices linked to a leaf and to z. The case $|S| = 3$ is treated in Figure 7. If $|S| = 4$, the case of the star will be included in the general case with $p = 3$, $q = 0$, and the case of the path is described in Figure 8. In the remaining cases, that is, when S is a star with four vertices and

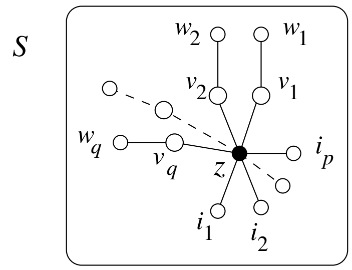


FIGURE 6. The tree S is an extended star with branches of lengths 1 or 2

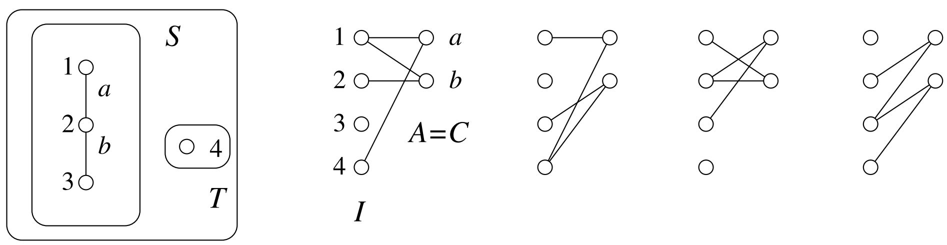


FIGURE 7. The forest $S \cup T$ gives four isomorphic graphs when $|S| = 3$, with $\alpha = 2$

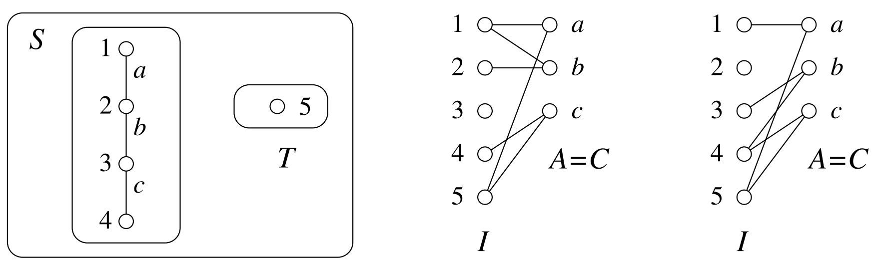


FIGURE 8. The forest $S \cup T$ gives two isomorphic graphs when $|S| = 4$ and S is a path ($\alpha = 2$)

when $|S| \geq 5$ (see Figure 9 for notation), using repeatedly Lemma 5 and the fact that $\alpha = 2$, we obtain that

– for each of the p edges c_k, $N(c_k) = \{i_k, t\}$,
– for each of the q edges d_k, $N(d_k) = \{w_k, t\}$,
– for each of the q edges e_k, $N(e_k) = \{v_k, w_k\}$,

and we can build the corresponding family of graphs G with parameters p, q, see Figure 9.

If $i \notin N(c)$, then, similarly to the case $i_S \notin N(c_S)$ at the end of case (i), we obtain the graphs which are complementary to the graphs described in Figures 7, 8 and 9; therefore, we have proved the following result.

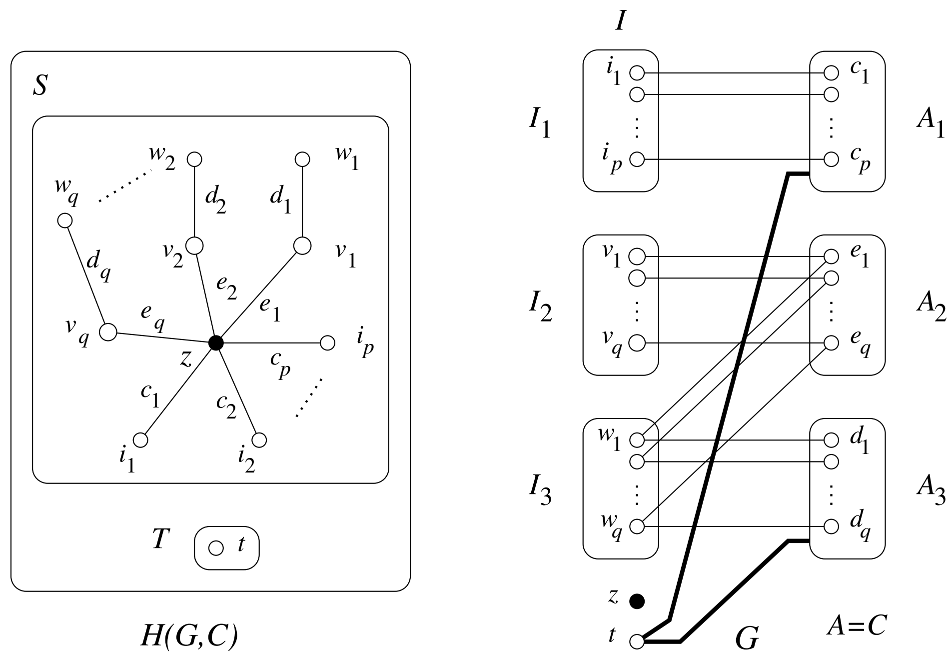


FIGURE 9. The forest $S \cup T$ and the graphs G when $|S| \geq 5$ or when S is the star with 4 vertices; the bold lines mean that all possible edges exist between t and A_1, and t and A_3

Theorem 3. *The only bipartite attribute-regular graphs $(I \cup A, E)$ with $|I| = n$ individuals, $|A| = n - 2$ attributes, and admitting A as only separating code, are isomorphic to the families of graphs G described in Figures 4 and 9 and their complementary graphs $\overline{G}$, plus the particular cases given in Figures 7 and 8 and their complementary graphs.*

Note that the particular cases given in Figures 7 and 8 are actually included in the general case, but are different if considered as *labelled* graphs, because they do not have a vertex playing the same role as the isolated individual z.

3.3. 1-DISCRIMINATING CODES WITH SIZES $|I|$, $|I| - 1$. Using the results just obtained on separating codes, we can now exhibit bipartite graphs $G = (V = I \cup A, E)$ where $|A| = |I|$ or $|I| - 1$ and $C = A$ is the only discriminating code.

3.3.1. *The Case* $|A| = |I|$. We obtain all bipartite graphs with $|I| = n$ individuals, $|A| = n$ attributes, and admitting A as only discriminating code, by considering all bipartite graphs with $|I| = n+1$ individuals, $|A| = n$ attributes, admitting A as only separating code and containing an individual without attribute, and by removing this individual.

Surveying the graphs G described in the proof of Corollary 3, we see that, for each labelled tree with $n + 1$ vertices, $n + 1$ graphs G have an isolated individual. For instance, in Figure 2, the graph G_1 (respectively, G_2, G_3, G_5) deprived of the individual i_4 (respectively, i_3, i_2, i_1) has a minimal discriminating code of size 3; in Figure 3, the same holds for G_1 (respectively, G_2, G_3, G_7) deprived of the individual i_3 (respectively, i_4, i_2, i_1).

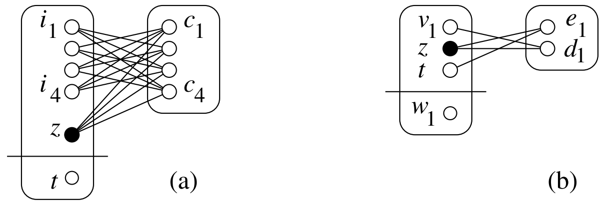


FIGURE 10. The complementary graphs of the graphs G given in Figure 9, particular cases: (a) $p = 4, q = 0$, (b) $p = 0, q = 1$

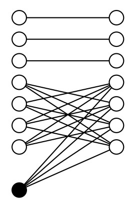


FIGURE 11. Another graph requiring $|I| - 1$ codewords to constitute a discriminating code

In the particular case of attribute-regular graphs, mentioned at the end of Subsection 3.1, we obtain only the graph G of Example 1 deprived of the isolated individual i_0, that is, a perfect matching.

3.3.2. *The Case* $|A| = |I| - 1$. In order to obtain bipartite graphs with $|I| = n$ individuals, $|A| = n - 1$ attributes, and admitting A as only discriminating code, we consider the graphs we know with $|I| = n+1$ individuals, $|A| = n-1$ attributes, admitting A as only separating code and containing an individual without attribute, and we remove this individual.

Using the results of Subsection 3.2, we see that neither the family of graphs in Figure 4 nor their complementaries have isolated individuals. As far as Figure 9 is concerned, the family of graphs described there have an isolated individual, namely z, whereas their complementaries do not, unless we are in some particular cases: (a) the individual t can be isolated if $q = 0$, and (b) w_1 can be isolated if $p = 0, q = 1$, see Figure 10. We can observe that case (a) is the complementary graph of the graph G in Example 1.

Since in Subsection 3.2, we considered only attribute-regular graphs, here also we obtain only attribute-regular graphs, and not all graphs requiring $|I|-1$ codewords. One example which we do not reach is given in Figure 11.

3.3.3. *Summary.* The following table recapitulates the results of Subsection 3.3.

	1-discriminating codes with $\|A\| = \|C\| = \|I\|$	1-discriminating codes with $\|A\| = \|C\| = \|I\| - 1$
any graph	• some graphs of Cor. 3 minus one individual (= all)	• graph of Fig. 11 • ...?
attribute-regular graphs	• perfect matchings (= all)	• graphs of Fig. 9 minus z • particular cases, *cf.* Fig. 10 (= all)

3.4. A RESULT AND A CONJECTURE ON 1-IDENTIFYING CODES. The following is known about "bad" graphs $G^* = (V^*, E^*)$ with respect to identifying codes: if G^* is connected and finite and admits a 1-identifying code C^*, then $|C^*| \leq |V^*| - 1$, see [2], [10], [12]. Using the *r-transitive closure*, or *r-th power*, of G^*, which is the graph with vertex set V^* and edges between any two vertices at distance at most r in G^*, it is straightforward to see that the same is true for r-identifying codes. On the other hand, if G^* is finite and not connected, then the only family of graphs in which the only identifying code is the whole set of vertices is the family of *independent*, or *stable*, sets (graphs for which $E^* = \emptyset$).

Note that in the case of infinite graphs, there are connected examples where all the vertices are necessary in an identifying code [9], [10].

The bound $|C^*| \leq |V^*| - 1$ is tight: when $r = 1$, it is easy to observe that the bound is reached by the star with at least three elements and the complete graph minus a maximum matching. It is more difficult to find examples of graphs reaching the bound when $r > 1$, see [9], [10]. In both cases, we do not know whether these graphs are the only ones to achieve the bound $|C^*| \leq |V^*| - 1$. Using the previous subsection, we can improve slightly our knowledge, in the case $r = 1$.

We consider a graph $G^* = (V^*, E^*)$ and we construct the bipartite graph $G = (V = I \cup A, E)$ in the following way:

$$I = V^*,\ A = \{B_1(j) : j \in V^*\},\ E = \{\{i, B_1(j)\} : i \in B_1(j), j \in V^*\}.$$

We call this graph the 1-*ball membership graph* of G^* (1-BM graph for short). We shall see in Section 5 that there is a 1-identifying code of size at most k in G^* if and only if there is a 1-discriminating code of size at most k in G; and that if C^* is a 1-identifying code in G^*, then $C = \{B_1(c^*) : c^* \in C^*\} \subseteq A$ is 1-discriminating in G.

Now we assume that G^* admits a minimum 1-identifying code C^* of size $|V^*| - 1$ which is regular with degree $\alpha - 1$, that is: $\forall c^* \in C^*, \deg(c^*) = \alpha - 1$. Setting $C^* = V^* \setminus \{x_0\}$, we construct the 1-BM graph of G^* and from it we remove the vertex $B_1(x_0)$ and its edges, to obtain a graph $G' = (V' = I \cup A', E')$ where $I = V^*$ and $A' = \{B_1(c^*) : c^* \in C^*\}$.

Because C^* is a minimum 1-identifying code in G^*, we know that A' is a minimum 1-discriminating code in the 1-BM graph of G^*; in G', A' is the only 1-discriminating code, has size $|I| - 1$ and is regular with degree α, which is the size of a ball of radius 1 centred at any codeword $c^* \in C^*$. We can therefore apply the results of Subsection 3.3.2, summarized in the rightmost downmost square of the table in Subsection 3.3.3.

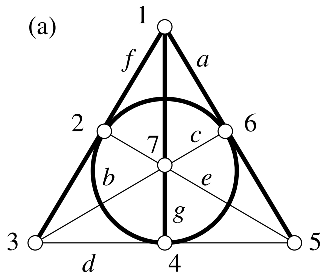
It is then tedious but straightforward to check that, going back to G^*, the only possibilities are the star with at least three vertices and the complete graph minus a maximum matching. So, we have the following proposition.

Proposition 1. *The only connected graphs with n vertices such that there is a* regular *minimum 1-identifying code with size $n - 1$ are the star and the complete graph minus a maximum matching.*

We believe that this is still true when we drop the regularity condition on the code.

Conjecture 1. *The only connected graphs with n vertices such that there is a minimum 1-identifying code with size $n - 1$ are the star and the complete graph minus a maximum matching.*

(a)



(b)

$$\begin{array}{c} \\ a\\ b\\ c\\ d\\ e\\ f\\ g \end{array}\begin{array}{c} \;1\;2\;3\;4\;5\;6\;7 \\ \begin{bmatrix} 1&0&0&0&1&1&0\\ 0&1&0&1&0&1&0\\ 0&0&1&0&0&1&1\\ 0&0&1&1&1&0&0\\ 0&1&0&0&1&0&1\\ 1&1&1&0&0&0&0\\ 1&0&0&1&0&0&1 \end{bmatrix} \end{array} = \boldsymbol{A}$$

incidence matrix

FIGURE 12. Two different representations of the Fano graph; bold lines represent the codewords of a minimum 1-discriminating code

4. Miscellaneous Families of Bipartite Graphs

The family of bipartite *planar* graphs, and in particular *trees*, has beeen investigated in [6], where, among other results, a linear-time algorithm is described, which, given a tree, outputs a minimum 1-discriminating code. In [7], strong links between discriminating and identifying codes are shown in the binary *n-hypercube* (or binary Hamming space of dimension n).

Here, we give the exact size or density of a minimum 1-discriminating code in three particular bipartite graphs, namely the *Fano graph* and two *infinite* graphs, the *square* and the *hexagonal* grids.

4.1. The Fano Graph.

The Fano graph, for which we can efficiently use Lemma 2, consists of seven individuals $1, 2, \dots, 7$ and seven attributes $a, b, \dots, g$, and an individual owns an attribute if and only if the individual belongs to the curve labelled by the attribute, see Figure 12(a).

Proposition 2. *In the Fano graph, any minimum* 1*-discriminating code has size four.*

Proof. Let C be any minimum discriminating code in the Fano graph. If $\mathbf{F}_2$ denotes the field $\{0, 1\}$, it is straigthforward to see that the $\mathbf{F}_2$-rank of the incidence matrix $\mathbf{A}$ of the Fano graph, given in Figure 12(b), is four, therefore by Lemma 2, $|C| \leq 4$. On the other hand, we can use inequality (3) with $\alpha = 3$, and obtain that $|C| \geq \lceil \frac{2\times 7}{4} \rceil = 4$. An example of such a code is $\{a, b, f, g\}$. $\square$

4.2. The Infinite Square Grid.

The two-dimensional infinite square grid has vertex set Z^2 and edge set

$$\{\{(i, j), (i, j \pm 1)\}, \{(i, j), (i \pm 1, j)\} : i \in Z, j \in Z\}.$$

Without loss of generality, we can choose the set A of attributes to be $\{(i, j) : i \in Z, j \in Z, i+j \text{ odd}\}$, and the set I of individuals to be $\{(i, j) : i \in Z, j \in Z, i+j \text{ even}\}$, see Figure 13(a).

Taking the limit of (4) with $\alpha = \alpha_{\max} = 4$, we see that a 1-discriminating code has density at least 0.4 among the attributes. On the other hand, we now give a periodic construction of a discriminating code using 40% of the attributes: consider, as in Figure 14, a pattern of five columns, take as codewords the two attributes of each row containing exactly two attributes, and repeat this pattern. It is easy to see that the code thus constructed is 1-discriminating.

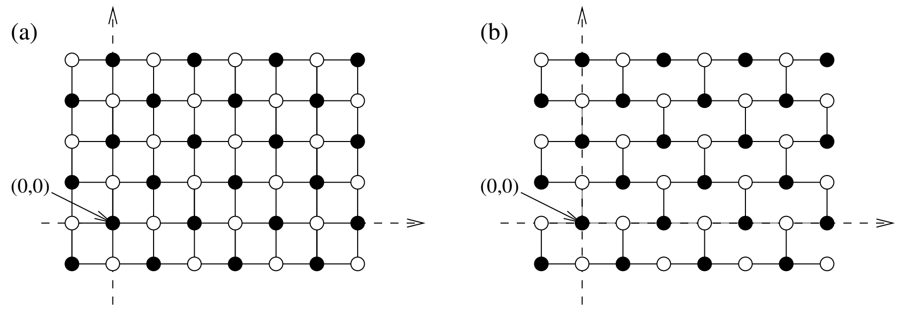


FIGURE 13. Partial representations of the infinite square and hexagonal grids; individuals are represented by black squares, attributes by white squares

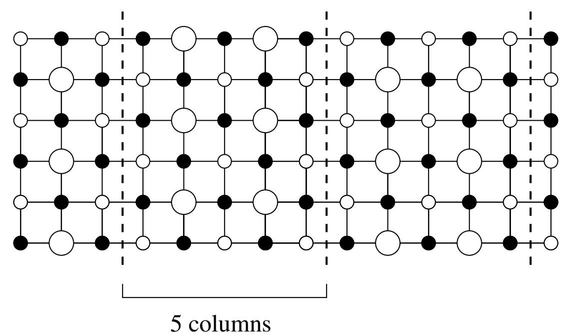


FIGURE 14. Partial representation of a periodic discriminating code in the infinite square grid; individuals are represented by black squares, attributes by white squares, large white squares are codewords

4.3. THE INFINITE HEXAGONAL GRID. The two-dimensional infinite hexagonal grid has vertex set Z^2 and edge set

$$\{\{(i,j),(i,j+(-1)^{i+j})\},\{(i,j),(i\pm 1,j)\}: i\in Z, j\in Z\}.$$

Without loss of generality, we can choose the set A of attributes to be $\{(i,j): i\in Z, j\in Z, i+j \text{ odd}\}$, and the set I of individuals to be $\{(i,j): i\in Z, j\in Z, i+j \text{ even}\}$, see Figure 13(b).

Taking the limit of (4) with $\alpha=\alpha_{\max}=3$, we see that a 1-discriminating code has density at least 0.5 among the attributes. On the other hand, we now give a periodic construction of a discriminating code using half of the attributes: simply take as codewords all the attributes on every second row. It is easy to check on Figure 15 that the code thus constructed is 1-discriminating.

5. COMPLEXITY

For notions of complexity and NP-completeness, we refer to [11]. We consider only odd integers $r,\ r\geq 1$, as mentioned in the Introduction.

Theorem 4. *The following decision problem is NP-complete:*
Name: *r-Discrimination (r-DISC).*

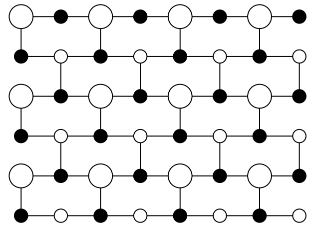


FIGURE 15. Partial representation of a periodic discriminating code in the infinite hexagonal grid; individuals are represented by black squares, attributes by white squares, large white squares are codewords

Instance: *A bipartite graph $G = (V = I \cup A, E)$, an integer k.*
Question: *Is there an r-discriminating code $C \subseteq A$ of size at most k?*

Proof. Since we shall consider two graphs $G = (V = I \cup A, E)$ and $G^* = (V^*, E^*)$, in order to avoid ambiguities we shall use the notation $d_G(x_1, x_2)$ for $x_1, x_2 \in V$ and $d_{G^*}(y_1, y_2)$ for $y_1, y_2 \in V^*$, and we draw the attention of the reader to the fact that we shall use $B_1(x)$ only for vertices x in V^*, which will be made equal to I.

First, we see that r-DISC belongs to NP, by observing that, given a set $C \subseteq A$, it is polynomial, with respect to the number of vertices of G, to check whether C is, or is not, an r-discriminating code (of convenient size); details are left to the reader. Next, we polynomially reduce the following NP-complete problem to r-DISC:

Name: r-Identification (r-ID).
Instance: A graph $G^* = (V^*, E^*)$, an integer k^*.
Question: Is there an r-identifying code of size at most k^* in G^*?

It has been proved in [8] that r-ID is NP-complete. We consider an instance of r-ID and, starting from G^*, we construct the bipartite graph G in the following way: $I = V^*$, $A = \{B_1(v) : v \in V^*\}$, $E = \{\{i, B_1(v)\} : i \in B_1(v), v \in V^*\}$; we set $k = k^*$. We can see that this is exactly the 1-ball membership graph described in Subsection 3.4. Note that this reduction is polynomial with respect to the size of the instance (G^*, k^*) of r-ID, and that it does not depend on r.

We claim that there is an r-identifying code of size at most k in G^* if and only if there is an r-discriminating code of size at most k in G.

This will be a direct consequence of the following fact (remember that r is odd):

$$\forall x, v \in V^* = I,\ d_G(x, B_1(v)) \le r \iff d_{G^*}(x, v) \le r. \tag{13}$$

Indeed, if in G there is a path $x, u_1, v_1, u_2, v_2, \ldots, u_p, v_p, B_1(v)$ with $2p + 2 \le r + 1$, $u_i \in A$ and $v_j \in I$, then, if $u_i = B_1(z_i)$, $z_i \in V^*$, the path $x, z_1, v_1, z_2, v_2, \ldots, z_p, v_p, v$ in G^* shows that $d_{G^*}(x, v) \le d_G(x, B_1(v)) \le r$. This is illustrated in Figure 16, with a path on 8 vertices in G (case $p = 3$).

We now prove the reciprocal; assume that there is in G^* a path $x, z_1, z_2, \ldots, z_p, v$, with $p+2 \le r+1$. Note that if p is odd, then $p \le r-2$, since r is odd. If p is even, then the path $x, B_1(z_1), z_2, \ldots, z_{p-2}, B_1(z_{p-1}), z_p, B_1(v)$ exists in G and has at most $p+2$ vertices, with $p \le r - 1$; if p is odd, then the path $x, B_1(z_1), z_2, \ldots, B_1(z_{p-2}), z_{p-1}$, $B_1(z_p), z_p, B_1(v)$ exists in G and has at most $p + 3$ vertices, with $p \le r - 2$. So in

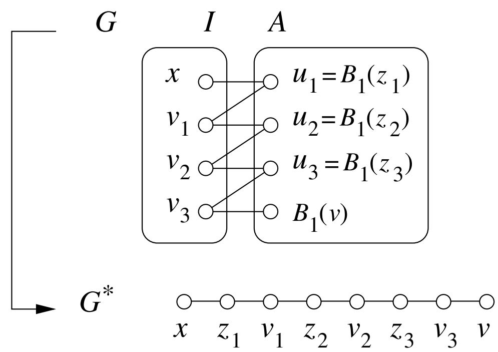


Figure 16. A path on 8 vertices in G and its corresponding path in G^*

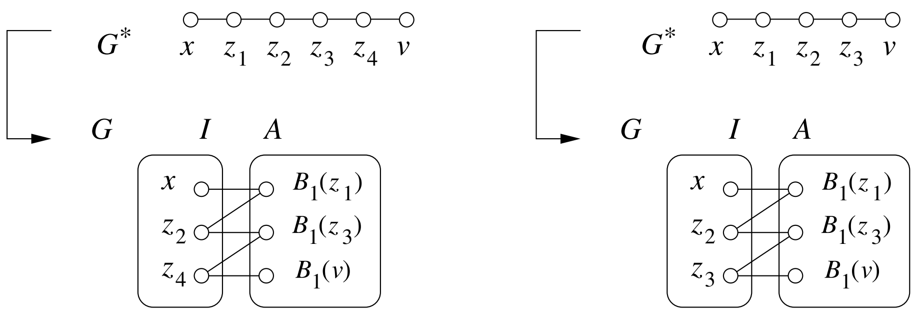


Figure 17. Paths on 6 and 5 vertices in G^* and their corresponding paths in G

both cases, $d_G(x, B_1(v)) \leq r$. This is illustrated in Figure 17, with paths on 6 and 5 vertices in G^* (cases $p = 4$ and $p = 3$, respectively).

We are now ready to prove that an r-identifying code C^* of size at most k exists in G^* if and only if an r-discriminating code C of size at most k exists in G.

If C^* exists, we take $C = \{B_1(v) : v \in C^*\} \subseteq A$. Then, since any vertex $x \in V^*$ is within distance r from at least one codeword $v \in C^*$, by (13) we have: $d_G(x, B_1(v)) \leq r$, for all $x \in I$. Also, since, given any two vertices $x, y \in V^*$, there is a codeword $v \in C^*$ such that, say, $d_{G^*}(x, v) \leq r$ and $d_{G^*}(y, v) > r$, we have, using (13): $d_G(x, B_1(v)) \leq r$ and $d_G(y, B_1(v)) > r$, for all $x, y \in I$. This proves that C is r-discriminating.

Conversely, assume that we have an r-discriminating code $C \subseteq A$ in G; there exists $C^* \subseteq V^*$ such that $C = \{B_1(v) : v \in C^*\}$, and we claim that C^* is an r-identifying code in G^*. Since any vertex $x \in I$ is within distance r from at least one codeword $B_1(v) \in C$, by (13) we have: $d_{G^*}(x, v) \leq r$, for all $x \in V^*$, with $v \in C^*$. Also, since, given any two vertices $x, y \in I$, there is a codeword $B_1(v) \in C$ such that, say, $d_G(x, B_1(v)) \leq r$ and $d_G(y, B_1(v)) > r$, we have, using (13): $d_{G^*}(x, v) \leq r$ and $d_{G^*}(y, v) > r$, for all $x, y \in V^*$, with $v \in C^*$. This proves that C^* is r-identifying. □

Remark. It can be proved [1] that the problem r-ID is still NP-complete when restricted to graphs with maximum degree at most three. Therefore the problem r-DISC is NP-complete even when restricted to graphs where attributes and individuals have degree at most four.

Acknowledgements

The authors wish to thank a referee for careful reading and useful remarks.

<irene.charon@telecom-paristech.fr; gerard.cohen@telecom-paristech.fr; olivier.hudry@telecom-paristech.fr; antoine.lobstein@telecom-paristech.fr>

References

[1] D. Auger, Personal Communication, 2007.

[2] N. Bertrand, Codes identifiants et codes localisateurs-dominateurs sur certains graphes, Mémoire de stage de maîtrise, ENST, Paris, 2001.

[3] J. A. Bondy, *Induced subsets*, J. Combin. Theory Ser. B, **12** (1972), 201–202.

[4] A. Cayley, *A theorem on trees*, Quart. J. Math., **23** (1889), 376–378.

[5] E. Charbit, I. Charon, G. Cohen and O. Hudry, *Discriminating codes in bipartite graphs*, Electron. Notes Discrete Math., **26** (2006), 29–35.

[6] I. Charon, G. Cohen, O. Hudry and A. Lobstein, *Discriminating codes in (bipartite) planar graphs*, European J. Combin., **29** (2008), 1353–1364.

[7] I. Charon, G. Cohen, O. Hudry and A. Lobstein, *Links between discriminating and identifying codes in the binary Hamming space*, Lect. Notes Comput. Sci., **4851** (2007), 267–270.

[8] I. Charon, O. Hudry and 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.

[9] I. Charon, O. Hudry and A. Lobstein, Extremal cardinalities for identifying and locating-dominating codes in graphs, Rapport interne Télécom Paris-2003D006, Paris, 2003.

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

[11] M. R. Garey and D. S. Johnson, Computers and Intractability, a Guide to the Theory of NP-Completeness, Freeman, New York, 1979.

[12] S. Gravier and J. Moncel, *On graphs having a* $V \setminus \{x\}$ *set as an identifying code*, Discrete Math., **307** (2007), 432–434.

[13] M. G. Karpovsky, K. Chakrabarty and L. B. Levitin, *On a new class of codes for identifying vertices in graphs*, IEEE Trans. Inform. Theory, **44** (1998), 599–611.

[14] D. F. Rall and P. J. Slater, *On location-domination numbers for certain classes of graphs*, Congr. Numer., **45** (1984), 97–106.

[15] http://www.infres.enst.fr/~lobstein/bibLOCDOMetID.html

Received May 2008; revised October 2008.