arXiv:2402.08264v3 [cs.DM] 4 Jul 2024

On Iiro Honkala's Contributions to Identifying Codes

Olivier Hudry[*]

Télécom Paris, France

olivier.hudry@telecom-paris.fr

Junnila Ville

University of Turku, Finland

viljun@utu.fi

Antoine Lobstein

CNRS, Université Paris-Saclay, France

lobstein@lri.fr

Abstract. A set C of vertices in a graph $G = (V, E)$ is an identifying code if it is dominating and any two vertices of V are dominated by distinct sets of codewords. This paper presents a survey of Iiro Honkala's contributions to the study of identifying codes with respect to several aspects: complexity of computing an identifying code, combinatorics in binary Hamming spaces, infinite grids, relationships between identifying codes and usual parameters in graphs, structural properties of graphs admitting identifying codes, and number of optimal identifying codes.

Keywords: graph theory, combinatorics, identifying codes, domination, separation, twin-free graphs, complexity, binary Hamming spaces, infinite grids, classic parameters of graphs, number of optimal solutions.

1. Introduction

The graphs $G = (V, E)$ (V is the vertex set of G and E its edge set) that we shall consider in this article are undirected and, unless otherwise stated, connected. A *code* is any subset of V, whose elements are called *codewords*. For any integer $r \geqslant 1$, the *ball of radius r*, or the *r-ball*, of a vertex $v \in V$

---

[*]Address for correspondence: Département Informatique et Réseaux de Télécom Paris, France.

is the set $B_r(v) = \{u \in V : 0 \leqslant d(u, v) \leqslant r\}$, where the considered distance d is the one provided by the length of a shortest path between u and v (for $r = 1$, $B_1(v)$ is the same as the usual *closed neighbourhood* $N[v]$). Two distinct vertices v_1 and v_2 are said to be *r-twins* if they share the same r-ball: $B_r(v_1) = B_r(v_2)$.

Definition 1. For any code $C \subseteq V$ and any vertex $v \in V$, the r-identifying set $I_r(C; v)$ of v with respect to C is defined by $I_r(C; v) = B_r(v) \cap C$. Furthermore, for any subset S of V, we set: $I_r(C; S) = \bigcup_{v \in S} I_r(C; v)$.

In other words, $I_r(C; v)$ is the set of codewords which are at distance at most r from v. When there is no ambiguity, we may write $I_r(v)$ and $I_r(S)$ instead of $I_r(C; v)$ and $I_r(C; S)$ respectively.

Two vertices within distance r from each other are indifferently said to r-*dominate* or to r-*cover* each other. A vertex v is r-*dominated* or r-*covered by a code* C if it is r-dominated by at least one element of C, i.e., if its r-identifying set is not empty: $I_r(C; v) \neq \emptyset$. If every vertex is r-dominated by C, then C is said to be r-*dominating* or r-*covering*. Two vertices u and v are r-*separated* by a vertex w if we have $w \in B_r(u)$ and $w \notin B_r(v)$ or if we have $w \in B_r(v)$ and $w \notin B_r(u)$ (note that w can be equal to u or v). Two vertices u and v are r-*separated by a code* C if they are r-separated by at least one element of C: $I_r(C; u) \neq I_r(C; v)$. If every distinct vertices u and v are r-separated, then C is said to be r-*separating*.

The conjunction of the r-domination and of the r-separation properties leads to the definition of an r-identifying code:

Definition 2. An r-*identifying code* $C \subseteq V$, or simply an r-IdC, is an r-dominating and r-separating code. In other words, C fulfills the following two properties:

- $\forall v \in V : I_r(C; v) \neq \emptyset$;

- $\forall v_1 \in V, \forall v_2 \in V$ with $v_1 \neq v_2 : I_r(C; v_1) \neq I_r(C; v_2)$.

It is easy to check that a graph G admits an r-IdC if and only if it has no r-twins (then, V itself is an r-IdC), i.e.:

$$\forall v_1 \in V, \ \forall v_2 \in V \text{ with } v_1 \neq v_2 : \ B_r(v_1) \neq B_r(v_2).$$

For any r-twin-free graph G, we denote by $Id_r(G)$ the smallest possible cardinality of an r-IdC and we call this the r-*identification number* of G, or simply the r-Id number of G. An r-IdC with the smallest possible size, i.e. whose cardinality is equal to $Id_r(G)$, is said to be *optimal*.

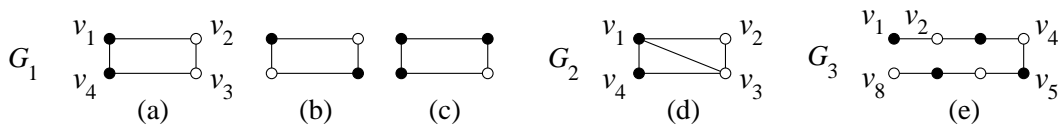**Example 1.1.** Consider the graphs provided by Figure 1.

Figure 1.   Different graphs and codes. Black vertices represent codewords.

- For G_1: clearly, G_1 requires at least two vertices in order to 1-cover all the vertices. But: (a) $\{v_1, v_4\}$ is not 1-identifying since v_1 and v_4 are not 1-separated; (b) $\{v_1, v_3\}$ is not 1-identifying since v_2 and v_4 are not 1-separated; so any 1-IdC contains at least three vertices; (c) $\{v_1, v_2, v_4\}$ is 1-identifying since all the vertices are 1-covered and any two distinct vertices are 1-separated. Hence $Id_1(G_1) = 3$. For $r > 1$, all the vertices of G_1 are r-twins.

- For G_2: v_1 and v_4 are not 1-separated by the code displayed in (d). Moreover, v_1 and v_3 are 1-twins: therefore they cannot be 1-separated. For $r > 1$, all the vertices of G_2 are r-twins. Thus, G_2 admits no r-IdC for any $r \geqslant 1$.

- For G_3: the code $\{v_1, v_3, v_5, v_7\}$ displayed in (e) is not 1-identifying since v_7 and v_8 are not 1-separated; but $\{v_1, v_3, v_5, v_6, v_7\}$ is 1-identifying. In contrast to the case $r = 1$, v_7 and v_8 are now 2-separated by $\{v_1, v_3, v_5, v_7\}$ (more precisely, by v_5), but not v_6 and v_7. For $r \geqslant 4$, v_4 and v_5 are r-twins.

A usual illustration of identifying codes comes from fault diagnosis in multiprocessor networks. Such a network can be represented as a graph whose vertices are processors and edges are links between processors. Assume that a processor is malfunctioning; we wish to locate the faulty processor. For this purpose, some processors (which will constitute the r-IdC) will be selected and assigned the task of testing the vertices at distance at most r. Whenever a selected processor (i.e., a codeword) detects a fault, it sends a binary message saying that one element in its r-ball is malfunctioning. We require that we can unambiguously identify the location of the malfunctioning processor, based only on the information which ones of the codewords sent a message; in this case, an identifying code is what we need.

The term "identifying code" is used in the 1998 paper [62] by M. G. Karpovsky, K. Chakrabarty and I. B. Levitin, which certainly marks the starting point for the blossoming of works on this topic, but the concept is already contained in [78] by N. S. V. Rao in 1993 (for references on identifying codes, see the ongoing bibliography at [58]).

The aim of this paper, in this issue devoted to Iiro Honkala, is to survey some of Iiro's contributions to the study of identifying codes. For this reason, when quoting a reference, we add a star (for instance [2]*) for denoting that this contribution is authored or co-authored by Iiro. For the same reason, some aspects of identifying codes are not covered here (for a wider survey on IdCs, see, e.g., [70]).

We consider also a generalization of r-IdCs, which occurs when more than one vertex should be identified [62]: for a given integer ℓ, an $(r, \leqslant \ell)$-*identifying code*, or simply $(r, \leqslant \ell)$-*IdC*, in a graph $G = (V, E)$ is an r-dominating code C such that for all $X \subseteq V$, $Y \subseteq V$ with $X \neq Y$, $0 < |X| \leqslant \ell$, $0 < |Y| \leqslant \ell$, we have:

$$\bigcup_{u \in X} I_r(C; u) \neq \bigcup_{v \in Y} I_r(C; v).$$

As for r-IdCs, we say that an $(r, \leqslant \ell)$-IdC is *optimal* if its size is minimum. Then $Id_{(r, \leqslant \ell)}(G)$ denotes this minimum size.

This generalization, and other similar ones, have been studied mostly in the n-cube and in the grids; we refer to, e.g., [26], [27], [36], [37], [44]*, [50]*, [62], [63], [73].

The following theorem provides a general property fulfilled by r-twin-free graphs.

Theorem 1.2. ([1])

Let $r \geqslant 1$ be any integer. Any r-twin-free graph G with at least one edge admits P_{2r+1}, the path on $2r + 1$ vertices, as an induced subgraph. As a consequence, G has order $n \geqslant 2r + 1$ and the only r-twin-free graph with order $n = 2r + 1$ and with at least one edge is the path P_{2r+1}.

Theorem 1.3 gathers a lower bound from [62] and an upper bound from [5] for $Id_r(G)$, both sharp (see also [17] and [38]).

Theorem 1.3. For any $r \geqslant 1$ and any connected r-twin-free graph $G = (V, E)$ of order $n \geqslant 2r + 1$:

$$\lceil \log_2(n + 1) \rceil \leqslant Id_r(G) \leqslant n - 1.$$

In what follows, the structure of the paper is explained. In Section 2, the computational complexity of determining an r-identifying code of a given size. Then, in Section 3, we proceed with various results concerning binary Hamming spaces. In Section 4, the results for infinite grids — the square, triangular, king and hexagonal grids — are summarized. As the existence of an identifying code requires the underlying graph to be twin-free, the structural properties of such graphs are studied in Section 5. Although the question of determining an optimal identifying code in a graph is already rather difficult, its natural extension of calculating the number of optimal codes has also been previously studied as examined in Section 6. Finally, in Section 7, some variants and closely related concepts of identifying codes are briefly discussed.

2. Complexity for general graphs

The decision problem associated with the computation of $Id_r(G)$ is NP-complete for any fixed $r \geqslant 1$, as stated by Theorem 2.1.

Theorem 2.1. ([21]* for $r = 1$, [16] for $r > 1$) For any fixed $r \geqslant 1$, the following problem is NP-complete:

    Name: r-IdC
    Instance: A graph $G = (V, E)$ and an integer $k \leqslant |V|$.
    Question: Does G admit an r-IdC of size at most k?

The proof of Theorem 2.1 for $r = 1$ is based on a polynomial transformation from the NP-complete problem 3-SAT. Let $(U; C_1, C_2, ..., C_m)$ be any instance of 3-SAT, where U denotes the set of Boolean variables and where $C_1, C_2, ..., C_m$ are the clauses, each clause containing three literals. The transformation uses two kinds of components. The first one, associated with each Boolean variable x_j $(1 \leqslant j \leqslant |U|)$, contains six vertices named $x_j, \overline{x_j}, a_j, b_j, c_j$ and d_j and the six edges $\{a_j, b_j\}, \{b_j, x_j\}$, $\{b_j, \overline{x_j}\}, \{c_j, d_j\}, \{c_j, x_j\}, \{c_j, \overline{x_j}\}$. The second one, associated with each clause C_i $(1 \leqslant i \leqslant m)$, contains two vertices named α_i and β_i and the edge $\{\alpha_i, \beta_i\}$. Extra edges reflect the composition of any clause: if C_i contains the three literals x_j, x_h and x_k, we add the edges $\{\alpha_i, x_j\}, \{\alpha_i, x_h\}$, $\{\alpha_i, x_k\}$. Figure 2 illustrates this transformation by specifying a part of the graph thus obtained.

By setting $k = 3|U| + m$, it is proved in [21]* that the instance of 3-SAT is satisfiable if and only if there exists a 1-IdC of size at most k in the constructed graph (in fact the size is then exactly k and
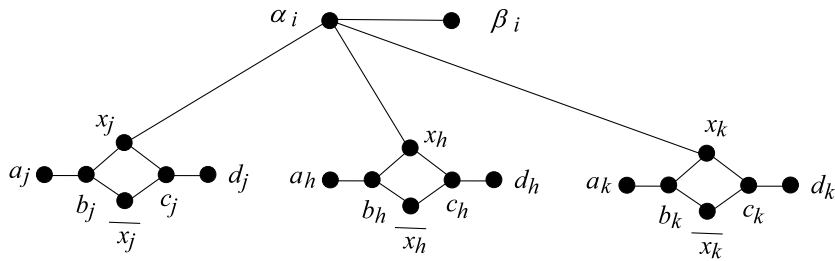
Figure 2. The subgraph induced by a clause $C_i = x_j \vee x_h \vee x_k$ for the transformation from 3-SAT to 1-IdC.

such a code can be provided by all the vertices α_i with $1 \leqslant i \leqslant m$, the vertices x_j or $\overline{x_j}$ according to the Boolean value taken by the variable associated with x_j and $\overline{x_j}$ for $1 \leqslant j \leqslant |U|$, and all the vertices b_j and c_j for $1 \leqslant j \leqslant |U|$). This construction is generalized in [16] to any integer r.

The NP-completeness of the problem r-IdC involves that, given G and r, determining $Id_r(G)$ or finding an optimal r-IdC of G are NP-hard problems (for other results related to complexity issues, see [3], [32], [33], [55] and [56]).

Specific families of graphs have been studied, including the binary Hamming spaces considered in the next section. For complexity results dealing with identification in these graphs, see [52]* and [53]* and, below, Subsection 3.1.

3. Binary Hamming spaces

For the whole section, let n be a positive integer. The *binary Hamming space* $\mathbb{F}^n$ is the n-fold Cartesian product of the binary field $\mathbb{F} = \{0, 1\}$. The 2^n elements of $\mathbb{F}^n$, i.e. vectors with n binary coordinates, are called *words*. A subset of $\mathbb{F}^n$ is called a *code of length* n. Let $\mathbf{x}$ and $\mathbf{y}$ be words belonging to $\mathbb{F}^n$; the *Hamming distance* $d(\mathbf{x}, \mathbf{y})$ between words $\mathbf{x}$ and $\mathbf{y}$ is the number of coordinate places in which they differ. The *Hamming ball* of radius r centred at $\mathbf{x}$ is denoted by $B_r(\mathbf{x})$ and consists of the words that are within Hamming distance r from $\mathbf{x}$.

The Hamming space may also be seen as a graph (observe then that n does not denote the number of vertices of the graph, but still the length of the Hamming space): $\mathbb{F}^n$ is the vertex set; if $\mathbf{x}$ and $\mathbf{y}$ are words belonging to $\mathbb{F}^n$ (i.e. two vertices of the associated graph), there is an edge between $\mathbf{x}$ and $\mathbf{y}$ if $\mathbf{x}$ and $\mathbf{y}$ differ by one coordinate, i.e. if their Hamming distance is equal to 1. As this graph is completely defined by $\mathbb{F}^n$, we will denote it by $\mathbb{F}^n$, omitting its edge set. In $\mathbb{F}^n$, the shortest path distance is the same as the Hamming distance and, thus, the Hamming ball of radius r is the same as the r-ball defined in Section 1. Hence the same notation is used in this section for the distance d and the ball B_r as in Section 1.

The set of non-zero coordinates of the word $\mathbf{x}$ is called the *support* of $\mathbf{x}$ and is denoted by $supp(\mathbf{x})$. The *weight* of $\mathbf{x}$ is the cardinality of the support of $\mathbf{x}$ and is denoted by $w(\mathbf{x})$.

The size of a Hamming ball of radius r in $\mathbb{F}^n$ does not depend on the choice of the centre and it is denoted by $V(n, r)$. Furthermore, we have:

$$V(n, r) = \sum_{i=0}^{r} \binom{n}{i}.$$

Assuming $\mathbf{x} = (a_1, a_2, \ldots, a_n) = a_1 a_2 \cdots a_n$ and $\mathbf{y} = (b_1, b_2, \ldots, b_m) = b_1 b_2 \cdots b_m$, we can define the *concatenation* $(\mathbf{x}, \mathbf{y})$ of the words $\mathbf{x}$ and $\mathbf{y}$ as follows:

$$(\mathbf{x}, \mathbf{y}) = (a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m) = a_1 a_2 \cdots a_n b_1 b_2 \cdots b_m.$$

The *sum* of the words $\mathbf{x} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{y} = (b_1, b_2, \ldots, b_n)$ is defined as

$$\mathbf{x} + \mathbf{y} = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

where the sum $a_i + b_i$ $(1 \leqslant i \leqslant n)$ is performed modulo 2.

If $C_1 \subseteq \mathbb{F}^n$ and $C_2 \subseteq \mathbb{F}^m$ are codes, then their *direct sum*

$$C_1 \oplus C_2 = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \in \mathbb{F}^n, \mathbf{y} \in \mathbb{F}^m\} \subseteq \mathbb{F}^{n+m}$$

is defined as a code in $\mathbb{F}^{n+m}$. The following well-known lemma, which considers the intersections of balls of radius 1 in $\mathbb{F}^n$, is often quite useful.

Lemma 3.1. For $\mathbf{x} \in \mathbb{F}^n$ and $\mathbf{y} \in \mathbb{F}^n$:

$$|B_1(\mathbf{x}) \cap B_1(\mathbf{y})| = \begin{cases} n + 1, & \text{if } d(\mathbf{x}, \mathbf{y}) = 0. \\ 2, & \text{if } 1 \leqslant d(\mathbf{x}, \mathbf{y}) \leqslant 2. \\ 0, & \text{if } 2 < d(\mathbf{x}, \mathbf{y}). \end{cases}$$

Furthermore, the intersection of three distinct Hamming balls of radius 1 contains at most one word.

Note that $(r, \leqslant \ell)$-IdCs do not exist for all lengths n. For $\ell \geqslant 1$, the existence of an $(r, \leqslant \ell)$-IdC is discussed in the following theorem.

Theorem 3.2. ([48]*)
Let $r(n, K)$ denote the smallest radius r such that there exists an r-covering code in $\mathbb{F}^n$ with cardinality K. If $1 \leqslant \ell < 2^n$ and $r \geqslant r(n, \ell)$, then there exists no $(r, \leqslant \ell)$-IdC in $\mathbb{F}^n$.

Extensive coverage of the values $r(n, K)$ can be found in [19]*. For example, if $\ell = 1$, then an r-IdC in $\mathbb{F}^n$ exists if and only if $n \geqslant r + 1$. Based on these results, we also obtain that there exists no $(1, \leqslant 2)$-IdC of length smaller than 4 and no $(2, \leqslant 2)$-IdC of length smaller than 6.

In the rest of this section, we first focus on complexity issues for Hamming spaces. Then, in Subsection 3.2, various lower bounds on r-IdCs are considered. In Subsection 3.3, we proceed with the constructions of r-IdCs leading to the best known upper bounds. In Subsection 3.4, we discuss some natural conjectures concerning identification in binary Hamming spaces and consider the progress made towards them. Finally, in Subsections 3.5 and 3.6, we study codes identifying sets of vertices, and briefly discuss some other variants of (usual) $(r, \leqslant 1)$-IdCs.

3.1. Complexity issues for binary Hamming spaces

As stated in Section 2, the problem of deciding whether there exists an r-IdC in a general graph with at most k codewords is NP-complete. The problem remains algorithmically difficult when we restrict ourselves to binary Hamming spaces. More precisely, the following theorem is proved in [53]* thanks to a polynomial transformation from 3-SAT.

Theorem 3.3. ([53]*)
The following problem is NP-complete:

> Name: Not r-IdC
> Instance: An integer n, a list of binary codewords of length n, forming a code C; an integer r.
> Question: Are there distinct words $\mathbf{x}$ and $\mathbf{y}$ belonging to $\mathbb{F}^n$ with $B_r(\mathbf{x}) \cap C = \emptyset$ or with $B_r(\mathbf{x}) \cap C = B_r(\mathbf{y}) \cap C$?

Observe that, for the binary Hamming spaces, it is not necessary to encode the graph by its adjacency matrix, as usually done for an arbitrary graph. Indeed, the specification of the length n suffices to characterize $\mathbb{F}^n$. In fact, even the problem of deciding whether a given code is r-identifying in $\mathbb{F}^n$ is co-NP-complete.

3.2. Lower bounds

In this subsection, we discuss various lower bounds concerning r-IdCs in $\mathbb{F}^n$. The first one of these lower bounds presented in the following theorem is based on the simple fact that if $C \subseteq \mathbb{F}^n$ is an r-IdC, then there are at most $|C|$ words r-covered by exactly one codeword and all the other words of $\mathbb{F}^n$ have to be r-covered by at least two codewords.

Theorem 3.4. ([62])
We have $Id_r(\mathbb{F}^n) \geqslant \dfrac{2^{n+1}}{V(n,r)+1}$.

This lower bound can be further refined by observing that if $C \subseteq \mathbb{F}^n$ is an r-IdC, then the number of words r-covered by exactly k codewords is restricted by the binomial coefficient $\binom{|C|}{k}$. This result has been originally presented in [62] and then later revisited in [29], where the case $r \geqslant n/2$ has also been efficiently handled.

Theorem 3.5. ([29], [62])
Let $C \subseteq \mathbb{F}^n$ be an r-IdC as well as s and s' be the largest integers such that

$$\sum_{i=1}^{s} \binom{|C|}{i} \leqslant 2^n \text{ and } \sum_{i=0}^{s'} \binom{|C|}{i} \leqslant 2^n.$$

(i) If $r < n/2$, then the cardinality of C satisfies

$$|C| \cdot V(n,r) \geqslant \sum_{i=1}^{s} i \binom{|C|}{i} + (s+1)\left(2^n - \sum_{i=1}^{s} \binom{|C|}{i}\right).$$

(ii) If $n/2 \leqslant r \leqslant n-1$, then the cardinality of C satisfies

$$|C| \cdot V(n, n-r-1) \geqslant \sum_{i=1}^{s'} i \binom{|C|}{i} + (s'+1)\left(2^n - \sum_{i=0}^{s'} \binom{|C|}{i}\right).$$

Although the previous lower bounds are rather simple, they are still viable when the length n is small compared to the radius r as we will later see in Subsection 3.3. Notice also that the lower bounds do not hold only for the binary Hamming spaces, but they can be generalized for all graphs for which the cardinality of a ball of radius r does not depend on the chosen centre.

For the radius $r = 1$, the previous lower bounds can be improved by further restricting the number of words covered by exactly two codewords: if $C \subseteq \mathbb{F}^n$ is a 1-IdC and $\mathbf{x} \in \mathbb{F}^n$ such that $\mathbf{x}$ is covered by exactly $k \ (\geqslant 2)$ codewords, then there exist at most $\binom{k}{2}$ vertices $\mathbf{y}$ satisfying $|I_1(C; \mathbf{y})| = 2$ and $I_1(C; \mathbf{y}) \subseteq I_1(C; \mathbf{x})$ (see also Lemma 3.1). The result was originally presented in [62] and later a more convenient proof was provided in [8]*. Besides a few sporadic lengths n, the following lower bound is the best currently known one.

Theorem 3.6. ([8]*, [62])
We have $Id_1(\mathbb{F}^n) \geqslant \dfrac{n2^n}{V(n, 2)}$.

For the radius $r > 1$, the first efforts to improve the lower bounds of Theorems 3.4 and 3.5 were made in [31]. This approach was further refined in [29].

3.3. Upper bounds

Concerning the upper bounds, we first focus on the ones related to the radius $r = 1$. The exact values of $Id_1(\mathbb{F}^n)$ are known for $n \in \{2, 3, \ldots, 7\}$ as explained in the following.

- By Theorem 3.6, we have $Id_1(\mathbb{F}^2) \geqslant 3$, $Id_1(\mathbb{F}^3) \geqslant 4$ and $Id_1(\mathbb{F}^5) \geqslant 10$. Moreover, the lower bounds can be attained by the constructions given in [62].

- In [8]*, the lower bounds $Id_1(\mathbb{F}^4) \geqslant 6$ and $Id_1(\mathbb{F}^7) \geqslant 31$ of Theorem 3.6 are respectively improved to 7 and 32. Moreover, 1-IdCs attaining the improved lower bounds are also given.

- In [8]*, a 1-IdC in $\mathbb{F}^6$ with 19 codewords is presented. By Theorem 3.6, it is known that $Id_1(\mathbb{F}^6) \geqslant 18$ and later, in [31], it is shown, based on extensive computer searches, that $Id_1(\mathbb{F}^6) = 19$.

For $n \geqslant 8$, the smallest cardinalities of 1-IdCs remain unknown: the best known lower bounds follow by Theorem 3.6 and the best constructions have been given in [9]. For an efficient method of constructing 1-IdCs, we require the following definition of μ-fold coverings.

Definition 3. Let μ be an integer. If $C \subseteq V$ is a code such that $|I_r(C; u)| \geqslant \mu$ for all $u \in V$, then we say that C is a μ-*fold r-covering*. Furthermore, if $|I_r(C; u)| = \mu$ for all $u \in V$, then we say that C is a *perfect μ-fold r-covering*.

New 1-identifying codes can now be constructed based on 1-IdCs with the additional property that the code is also 2-fold 1-covering. The method is based on the classical $\pi(\mathbf{u})$-construction discussed, for example, in [19, Section 3.4]*. For all $\mathbf{u} \in \mathbb{F}^n$, the *parity-check bit* $\pi(\mathbf{u})$ is defined as

$$\pi(\mathbf{u}) = \begin{cases} 0 & \text{if } w(\mathbf{u}) \text{ is even} \\ 1 & \text{if } w(\mathbf{u}) \text{ is odd.} \end{cases}$$

Theorem 3.7. ([30])
If $C \subseteq \mathbb{F}^n$ is a 2-fold 1-covering and a 1-IdC in $\mathbb{F}^n$, then

$$C' = \{(\pi(\mathbf{u}), \mathbf{u}, \mathbf{u} + \mathbf{c}) \mid \mathbf{u} \in \mathbb{F}^n, \mathbf{c} \in C\} \ (\subseteq \mathbb{F}^{2n+1})$$

is a 2-fold 1-covering and a 1-IdC in $\mathbb{F}^{2n+1}$.

For the radius $r > 1$, we have the rather simple result from [7]* stating that

$$Id_r(\mathbb{F}^{r+1}) = 2^{r+1} - 1. \tag{1}$$

For larger radii, other exact values of $Id_r(\mathbb{F}^n)$ are rather scarcely known. In what follows, we focus more closely on the exact values for the radii $r = 2$ and $r = 3$.

- In [7]*, it has been shown that $Id_2(\mathbb{F}^4) = 6$; besides giving a 1-IdC with 6 codewords, the authors also improve the lower bound of Theorem 3.5 from 5 to 6. In the same paper, the constructions with 6 and 8 codewords are presented to attain the lower bounds $Id_2(\mathbb{F}^5) \geqslant 6$ and $Id_2(\mathbb{F}^6) \geqslant 8$ of Theorem 3.5, respectively.

- In [30], a 2-IdC in $\mathbb{F}^7$ with 14 codewords is given. Furthermore, it is shown by an exhaustive computer search that smaller 2-IdCs in $\mathbb{F}^7$ do not exist.

- In [39]*, a 3-IdC in $\mathbb{F}^5$ with 10 codewords is given. Furthermore, in [29], the lower bound $Id_3(\mathbb{F}^5) \geqslant 9$ of Theorem 3.5 is improved to 10 by a careful analysis of the proof of the theorem in that particular case.

- In [9], 3-IdCs in $\mathbb{F}^6$ and $\mathbb{F}^7$ with 7 and 8 codewords have been presented to attain the lower bounds $Id_3(\mathbb{F}^6) \geqslant 7$ and $Id_3(\mathbb{F}^7) \geqslant 8$, respectively.

For larger radii, the best known lower bounds mostly follow by Theorem 3.5 (and its refinement in [29]), when n is large enough. The best known constructions are mainly due to [9], although progress has also been made in [28] and [31]. In addition, a construction with the asymptotically smallest possible cardinality is presented in the following theorem from [52]*.

Theorem 3.8. ([52]*)
Let $0 \leqslant \rho < 1$ and let $H(x) = -x \log_2 x - (1 - x) \log_2(1 - x)$ denote the binary entropy function. Then: $\lim\limits_{n \to \infty} \dfrac{\log_2 Id_{\lfloor \rho n \rfloor}(\mathbb{F}^n)}{n} = 1 - H(\rho).$

3.4. Conjectures in binary Hamming spaces

In this subsection, we consider three conjectures concerning identifying codes in binary Hamming spaces. In what follows, we first introduce the following (rather plausible) conjectures.

(i) In [72], it has been shown that $Id_1(\mathbb{F}^n) \leqslant Id_1(\mathbb{F}^{n+1})$ for all $n \geqslant 2$. It is also stated as an open question whether the same monotonicity result also holds for larger radii $r > 1$. However, in [70], it has been noted that, for example, $Id_5(\mathbb{F}^6) = 63$ (by (1)) and $31 \leqslant Id_5(\mathbb{F}^7) \leqslant 32$ (by [9]). Hence, the monotonicity does not hold for all n, but it still seems plausible that there exists an integer $n(r)$ such that $Id_r(\mathbb{F}^n) \leqslant Id_r(\mathbb{F}^{n+1})$ for all $n \geqslant n(r)$. However, no progress towards this conjecture has been made.

(ii) In [8]* and [62], it has been conjectured that $Id_{r+t}(\mathbb{F}^{n+m}) \leqslant Id_r(\mathbb{F}^n)Id_t(\mathbb{F}^m)$. Although there is definite evidence that the conjecture should hold, no proof has been presented; even the significantly simpler statement $Id_1(\mathbb{F}^{n+1}) \leqslant 2Id_1(\mathbb{F}^n)$ still remains open.

(iii) In the seminal paper [62], the monotonicity of the size of an optimal r-IdC in $\mathbb{F}^n$ with respect to the radius r is discussed. In [27], it has been shown that there exists an integer n_r such that $Id_{r+1}(\mathbb{F}^n) \leqslant Id_r(\mathbb{F}^n)$ when $n \geqslant n_r$. However, the requirement for n_r is rather crude and improvements on that seem plausible.

In what follows, we concentrate more closely on Conjecture (ii). We first present some results concerning the weaker formulation $Id_1(\mathbb{F}^{n+1}) \leqslant 2Id_1(\mathbb{F}^n)$ of the conjecture. In [6]* and [8]*, first attempts towards this goal have been presented.

Theorem 3.9. ([8]*)
Let C be a 1-IdC in $\mathbb{F}^n$.

(1) The direct sum $C \oplus \mathbb{F}$ is a 1-IdC in $\mathbb{F}^{n+1}$ if and only if for all codewords $\mathbf{c} \in C$ there exists another codeword $\mathbf{c}' \in C$ such that $d(\mathbf{c}, \mathbf{c}') = 1$.

(2) The direct sum $C \oplus \mathbb{F}^2$ is a 1-IdC in $\mathbb{F}^{n+2}$. In particular, we have $Id_1(\mathbb{F}^{n+2}) \leqslant 4Id_1(\mathbb{F}^n)$.

In the following theorem (by [28]), it is shown that $Id_1(\mathbb{F}^{n+1}) \leqslant (2+\varepsilon_n)Id_1(\mathbb{F}^n)$, where $\varepsilon_n \to 0$ as $n \to \infty$.

Theorem 3.10. ([28])
Assume that $n \geqslant 2$. Then we have $Id_1(\mathbb{F}^{n+1}) \leqslant \left(2 + \dfrac{1}{n+1}\right) Id_1(\mathbb{F}^n)$.

Next we focus on the general formulation $Id_{r+t}(\mathbb{F}^{n+m}) \leqslant Id_r(\mathbb{F}^n)Id_t(\mathbb{F}^m)$ of the conjecture. In [28], various weaker versions of the conjecture have been proved. In particular, it has been shown that $Id_{r+1}(\mathbb{F}^{n+m}) \leqslant 4Id_r(\mathbb{F}^n)Id_1(\mathbb{F}^m)$. Furthermore, in [27], it has been proved that the conjecture holds when m is relatively small in comparison to n.

Theorem 3.11. ([27])
Let r and t be integers such that $r \geqslant 2$ and $t \geqslant 2$, and let $n \geqslant 2(t + r)$. If

$$2t \leq a \leq \frac{2n}{9(t+r)} \left(\frac{2^{r+3}(t-1)!(r-1)!}{9(9(t+r))^r} \right)^{\frac{1}{t}},$$

then we have

$$Id_{r+t}(\mathbb{F}^n) \leqslant Id_t(\mathbb{F}^a)Id_r(\mathbb{F}^{n-a}).$$

The following theorem, which has been presented in [30], is also closely related to the generalized formulation of Conjecture (ii).

Theorem 3.12. ([30])
Let r and n_i be positive integers. Then we have

$$Id_r(\mathbb{F}^{n_1+n_2+\cdots+n_r}) \leqslant \prod_{i=1}^{r} Id_1(\mathbb{F}^{n_i}).$$

3.5.   Identifying sets of vertices

In this subsection, we study $(r, \leqslant \ell)$-IdCs in binary Hamming spaces.

First efforts on identifying sets of vertices in $\mathbb{F}^n$ were already made in the seminal paper [62]. In what follows, we first focus on the case with the radius $r = 1$. The following theorem states that a $(1, \leqslant \ell)$-IdC is always a $(2\ell - 1)$-fold 1-covering. The theorem was presented in the case $\ell = 2$ in [48]* and later generalized to the case $\ell > 2$ in [68]. Its proof is based on an insightful application of Lemma 3.1.

Theorem 3.13. ([48]*, [68])
If C is a $(1, \leqslant \ell)$-IdC, then C is also a $(2\ell - 1)$-fold 1-covering. This further implies that

$$Id_{(1,\leq\ell)}(\mathbb{F}^n) \geqslant \left\lceil (2\ell - 1)\frac{2^n}{n+1} \right\rceil. \tag{2}$$

Inequality (2) of the previous theorem follows from [19, Chapter 14]*. The converse of the previous theorem also holds for $\ell \geqslant 3$ as shown in [63] (based again on Lemma 3.1).

Theorem 3.14. ([63])
Let ℓ be an integer such that $\ell \geqslant 3$. A code $C \subseteq \mathbb{F}^n$ is $(1, \leqslant \ell)$-identifying if and only if it is a $(2\ell - 1)$-fold 1-covering.

Determining the cardinalities of μ-fold t-coverings in $\mathbb{F}^n$ is a well-studied problem. Combining the results of [19, Chapter 14]* with the previous theorem, the following corollary is obtained.

Corollary 3.15. ([63])
Let ℓ be an integer such that $\ell \geqslant 3$. Then $Id_{(1,\leq\ell)}(\mathbb{F}^n) = (2\ell - 1)\dfrac{2^n}{n+1}$ if and only if there exist integers $i \geqslant 0$ and $\mu_0 > 0$ such that μ_0 divides $2\ell - 1$, and $2\ell - 1 \leqslant 2^i \mu_0$, and $n = \mu_0 2^i - 1$.

Notice that the requirement $\ell \geqslant 3$ is essential for Theorem 3.14. Indeed, in [77]*, it has been shown that there exists an infinite family of 3-fold 1-covering codes which are not $(1, \leqslant 2)$-identifying. However, in the same paper, optimal constructions for $(1, \leqslant 2)$-IdCs have been presented based on the so-called $\pi(\mathbf{u})$-construction already applied in Theorem 3.7.

Theorem 3.16. ([77]*)
If $C \subseteq \mathbb{F}^n$ is a $(1, \leqslant 2)$-IdC, then

$$C' = \{(\pi(\mathbf{u}), \mathbf{u}, \mathbf{u} + \mathbf{c}) \mid \mathbf{u} \in \mathbb{F}^n, \mathbf{c} \in C\} \; (\subseteq \mathbb{F}^{2n+1}),$$

where $\pi(\mathbf{u})$ denotes the parity-check bit of $\mathbf{u}$, is a $(1, \leqslant 2)$-IdC in $\mathbb{F}^{2n+1}$. This further implies that $Id_{(1,\leq 2)}(\mathbb{F}^{2n+1}) \leqslant 2^n Id_{(1,\leq 2)}(\mathbb{F}^n)$.

In [77]*, it has been proved that there exist $(1, \leqslant 2)$-IdCs $C_1 \subseteq \mathbb{F}^5$ and $C_2 \subseteq \mathbb{F}^7$ with 16 and 48 codewords, respectively. Hence, the codes C_1 and C_2 attain the lower bound of Theorem 3.13 implying that C_1 and C_2 are also perfect 3-fold 1-coverings. Applying the previous theorem to the codes C_1 and C_2, we obtain two infinite families of $(1, \leqslant 2)$-IdCs which are perfect 3-fold 1-coverings attaining the lower bound of Theorem 3.13. Hence, an optimal $(1, \leqslant 2)$-IdC is obtained for lengths $n \geqslant 5$ for which a perfect 3-fold 1-covering also exists (see [19, Theorem 14.2.4]*). More precisely, an optimal $(1, \leqslant 2)$-IdC in $\mathbb{F}^n$ is obtained for the lengths $n = 3 \cdot 2^k - 1$ and $n = 2^{k+2} - 1$, where k is a positive integer.

To find $(1, \leqslant \ell)$-IdCs for the lengths other than the ones with perfect $(2\ell - 1)$-fold 1-coverings, the direct sum construction of the following theorem may be used; for $\ell = 2$, the result is due to [48]*, and for $\ell > 2$, the result is presented in [63].

Theorem 3.17. ([48]*, [63])
Let ℓ be an integer such that $\ell \geqslant 2$. If $C \subseteq \mathbb{F}^n$ is a $(1, \leqslant \ell)$-IdC, then so is the direct sum $C \oplus \mathbb{F} \subseteq \mathbb{F}^{n+1}$. In particular, we have $Id_{(1,\leq \ell)}(\mathbb{F}^{n+1}) \leqslant 2 Id_{(1,\leq \ell)}(\mathbb{F}^n)$.

Note that the theorem proves, for $\ell \geqslant 2$, the issue similar to the conjecture $Id_1(\mathbb{F}^{n+1}) \leqslant 2 Id_1(\mathbb{F}^n)$ mentioned in Subsection 3.4. The generalized conjecture $Id_{r+t}(\mathbb{F}^{n+m}) \leqslant Id_r(\mathbb{F}^n) Id_t(\mathbb{F}^m)$ can also be approached for $\ell \geqslant 2$. In fact, the conjecture has been partially solved in the case $\ell \geqslant r + 3$ in [28].

Theorem 3.18. ([28])
If $\ell \geqslant r + 3$, then $Id_{(r+1,\leq \ell)}(\mathbb{F}^{n+m}) \leqslant Id_{(1,\leq \ell)}(\mathbb{F}^n) Id_{(r,\leq \ell)}(\mathbb{F}^m)$.

Next, we consider lower bounds on $(r, \leqslant \ell)$-IdCs when the radius $r \geqslant 2$. Firstly, a lower bound similar to the one of Theorem 3.5 has been provided in [29, Theorem 3.4]. It gives the best known lower bounds on $(r, \leqslant \ell)$-IdCs when $r \geqslant 2$ and n is relatively small compared to r and ℓ. For larger lengths n, the best known lower bounds are given by the following theorem due to [29] and [31].

Theorem 3.19. ([29], [31])
For $r \geqslant 2$ and $\ell \geqslant 2$, we have

$$Id_{(r,\leq \ell)}(\mathbb{F}^n) \geqslant \left\lceil \frac{(2\ell - 2)2^n}{\binom{n}{r}} \right\rceil \quad \text{and} \quad Id_{(r,\leq \ell)}(\mathbb{F}^n) \geqslant \left\lceil \frac{(2\ell - 1)2^n}{\binom{n}{r} + \binom{n}{r-1}} \right\rceil.$$

In what follows, we continue with the constructions of $(r, \leqslant \ell)$-IdCs when the radius $r \geqslant 2$. As discussed earlier, the cardinalities $Id_{(r,\leq\ell)}(\mathbb{F}^n)$ are rather well-known for $r = 1$ and $\ell \geqslant 2$. Hence, the approach of the following theorem by [31], which generalizes Theorem 3.12, is quite effective for constructing $(r, \leqslant \ell)$-IdCs.

Theorem 3.20. ([31])
For $r \geqslant 1$ and $\ell \geqslant 2$, we have $Id_{(r,\leq\ell)}(\mathbb{F}^{n_1+n_2+\cdots+n_r}) \leqslant \prod_{i=1}^{r} Id_{(1,\leq\ell)}(\mathbb{F}^{n_i})$.

Another approach for constructing $(r, \leqslant \ell)$-IdCs is based on a generalization of the direct sum construction of Theorem 3.17. This approach is studied in the following theorem, which has been presented in [29].

Theorem 3.21. ([29])
Let ℓ be an integer such that $\ell \geqslant 2$.

(1) If $1 \leqslant r \leqslant 2$ and C is an $(r, \leqslant \ell)$-IdC in $\mathbb{F}^n$, then $C \oplus \mathbb{F}^r$ is an $(r, \leqslant \ell)$-IdC in $\mathbb{F}^{n+r}$. In particular, we have $Id_{(r,\leq\ell)}(\mathbb{F}^{n+r}) \leqslant 2^r Id_{(r,\leq\ell)}(\mathbb{F}^n)$.

(2) If $r \geqslant 3$ and C is an $(r, \leqslant \ell)$-IdC in $\mathbb{F}^n$, then $C \oplus \mathbb{F}^{r+1}$ is an $(r, \leqslant \ell)$-IdC in $\mathbb{F}^{n+r+1}$. In particular, we have $Id_{(r,\leq\ell)}(\mathbb{F}^{n+r+1}) \leqslant 2^{r+1} Id_{(r,\leq\ell)}(\mathbb{F}^n)$.

Most of the best known $(r, \leqslant \ell)$-IdCs are due to Theorems 3.20 and 3.21. However, some incremental improvements have been made in [25]. In addition, an asymptotic result similar to Theorem 3.8 for $\ell = 1$ has been proved in [57].

Theorem 3.22. ([57])
Let $\ell \geqslant 1$ be fixed and let $\rho \in [0, 1/2)$. If $r/n \to \rho$ as $n \to \infty$, then

$$\lim_{n \to \infty} \frac{\log_2 Id_{(r,\leq\ell)}(\mathbb{F}^n)}{n} = 1 - H(\rho),$$

where $H(x) = -x \log_2 x - (1-x) \log_2(1-x)$ is the binary entropy function.

3.6. Other variants

Finally, we briefly discuss some variants of $(r, \leqslant \ell)$-IdCs. We first consider a situation where, with respect to the illustration proposed in Section 1, a malfunctioning processor may incorrectly report whether a problem occurs in its ball. Formally, that leads to the following definition (by [49]*), which is in view of Section 7 presented for general graphs $G = (V, E)$.

Definition 4. ([49]*)
Let $C \subseteq V$ be a code in $G = (V, E)$, and r and ℓ be nonnegative integers. For $X \subseteq V$, we define $\mathcal{I}_r(X) = \{U \mid I_r(X) \setminus (X \cap C) \subseteq U \subseteq I_r(X)\}$. If for all distinct $X_1, X_2 \subseteq V$ with $|X_1| \leqslant \ell$ and $|X_2| \leqslant \ell$ we have $\mathcal{I}_r(X_1) \cap \mathcal{I}_r(X_2) \neq \emptyset$, then we say that C is a *strongly $(r, \leqslant \ell)$-IdC*.

For strongly $(1, \leqslant \ell)$-IdCs in binary Hamming spaces, similar results as the ones described in the case of (usual) $(1, \leqslant \ell)$-IdCs have been derived in [49]*, [64] and [69]. Their proofs are also based on a careful analysis of the μ-fold 1-covering properties of the codes. Besides strongly identifying codes, various robust identifying codes, which are able to withstand additions and/or deletions of edges in the underlying graph, have been considered for binary Hamming spaces in [45]*, [47]*, [65] and [67].

Recall that $(r, \leqslant \ell)$-IdCs can locate up to ℓ faulty processors in the considered network. However, if more than ℓ faulty processors are attempted to be located using an $(r, \leqslant \ell)$-IdC, then errors may occur and we might even be unaware that the set of faulty processors indicated by the code is incorrect. This issue has been addressed in [46]*, where a new class of so-called $(r, \leqslant \ell)^+$-*IdCs* has been introduced. These codes can *locate* up to ℓ faulty processors and *detect* if more than ℓ faulty processors exist in the network. In particular, for the binary Hamming spaces $\mathbb{F}^n$, it has been shown in [46]* that $C \subseteq \mathbb{F}^n$ is a $(1, \leqslant \ell)^+$-IdC if and only if C is a $(2\ell + 1)$-fold 1-covering.

4. Infinite grids

We survey now results related to infinite grids. The four infinite 2-dimensional grids G_S (square), G_T (triangular), G_K (king), and G_H (hexagonal) have $\mathbb{Z} \times \mathbb{Z}$ as their vertex sets. They are partially represented in Figure 3 and are formally defined in the next subsections.
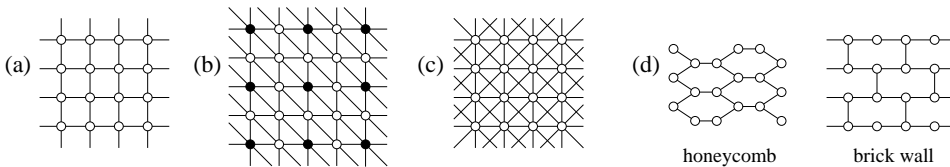

Figure 3.   Partial representations of the four grids: (a) the square grid G_S; (b) the triangular grid G_T: black vertices are codewords (cf. Theorem 4.3); (c) the king grid G_K; (d) the hexagonal grid G_H (with two possible representations: as a honeycomb or as a brick wall).

Because these grids are infinite graphs, we consider here the *density* of a code C instead of the cardinality of C. Let G denote one of these grids. For any integer q, let Q_q denote the set of vertices $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ with $|x| \leqslant q$ and $|y| \leqslant q$. The density of a code C in G is defined by $\partial(C, G) = \limsup_{q \to \infty} \frac{|C \cap Q_q|}{|Q_q|}$. Then we define the *minimum density* $\partial_r^{Id}(G)$ of G as the minimum of $\partial(C, G)$ over the set of the r-IdCs C of G. Thus, the problem considered in this section consists in computing the values of $\partial_r^{Id}(G)$ for $r \geqslant 1$ and for G belonging to $\{G_S, G_T, G_K, G_H\}$.
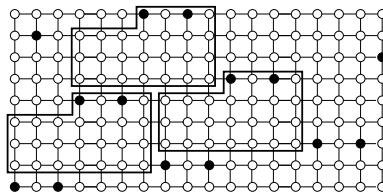

Figure 4.   A periodic 5-IdC in the square grid G_S, of density $2/25$; codewords are in black.

In the following subsections, we give lower and upper bounds for $\partial_r^{Id}(G)$, for $r \geqslant 1$ and $G \in \{G_S, G_T, G_K, G_H\}$. Several constructions of codes are obtained by heuristics searching for small subcodes inside tiles that will be repeated periodically, as done in [15] (see Figure 4 for such an example dealing with the square grid and $r = 5$). These constructions provide upper bounds for $\partial_r^{Id}(G)$.

4.1. The square grid

The square grid, G_S, has vertex set $V_S = \mathbb{Z} \times \mathbb{Z}$ and edge set $E_S = \{\{u, v\} : u - v \in \{(1, 0), (0, 1)\}\}$. The exact minimum density is known for $r = 1$, as stated by Theorem 4.1. The value 7/20 as an upper bound for $\partial_1^{Id}(G_S)$ comes from [18]* and from [4] as a lower bound. Figure 5 displays a construction providing the upper bound.

Theorem 4.1. ([4], [18]*)
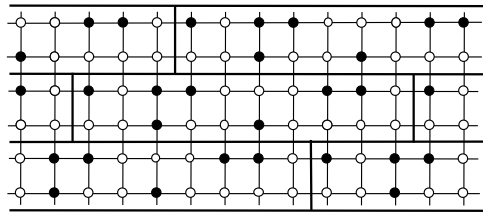For $r = 1$, $\partial_1^{Id}(G_S) = \frac{7}{20}$.

Figure 5.   A periodic 1-IdC in the square grid G_S, of density $7/20$; codewords are in black.

For $r \geqslant 2$, Theorem 4.2 specifies bounds for $\partial_r^{Id}(G_S)$. The lower bound for $r = 2$ comes from [59]; the general lower bound for $r \geqslant 3$ comes from [10]*; all the upper bounds come from [51]*.

Theorem 4.2. (a) For $r = 2$, $\frac{6}{35} \approx 0.17143 \leqslant \partial_2^{Id}(G_S) \leqslant \frac{5}{29} \approx 0.17241$.

(b) For $r \geqslant 3$, $\frac{3}{8r+4} \leqslant \partial_r^{Id}(G_S) \leqslant \begin{cases} \frac{2}{5r} & : r \text{ even} \\ \frac{2r}{5r^2-2r+1} & : r \text{ odd}. \end{cases}$

When r increases, these bounds are close to $\frac{3}{8r} = \frac{0.375}{r}$ and $\frac{2}{5r} = \frac{0.4}{r}$. The previous upper bounds have been improved in [15], using heuristics, for $r \in \{3, 4, 5, 6\}$ (with respectively $1/8 = 0.125$, $8/85 \approx 0.094$, $2/25 = 0.08$ and $3/46 \approx 0.065$).

4.2. The triangular grid

The triangular grid, or square grid with one diagonal, G_T, has vertex set $V_T = \mathbb{Z} \times \mathbb{Z}$ and edge set $E_T = \{\{u, v\} : u - v \in \{(1, 0), (0, 1), (1, -1)\}\}$.
    As for the square grid, the exact minimum density is known for $r = 1$:

Theorem 4.3. ([62])
For $r = 1$, $\partial_1^{Id}(G_T) = \frac{1}{4}$.

This minimum value is provided by the construction displayed in Figure 3(b), which has the property that every codeword is dominated by exactly one codeword (itself) and every non-codeword is dominated by exactly two codewords (either horizontally or vertically or diagonally). This proves that the code is indeed 1-identifying, and moreover that it is best possible. Anyway, it is shown in [21]* that this property cannot stand for $r > 1$. We have instead the following results for the triangular grid.

Theorem 4.4. (a) ([10]*) For $r \geqslant 2$, $\frac{2}{6r+3} \leqslant \partial_r^{Id}(G_T) \leqslant \begin{cases} \frac{1}{2r+2} & : r \in \{1, 2, 3\} \bmod 4 \\ \frac{1}{2r+4} & : r = 0 \bmod 4. \end{cases}$

(b) ([15]) We have $\partial_3^{Id}(G_T) \leqslant \frac{2}{17} \approx 0.11765$ and $\partial_5^{Id}(G_T) \leqslant \frac{1}{13} \approx 0.07692$.

4.3. The king grid

The king grid, or square grid with two diagonals, G_K, has vertex set $V_K = \mathbb{Z} \times \mathbb{Z}$ and edge set $E_K = \{\{u, v\} : u - v \in \{(1, 0), (0, 1), (1, -1), (1, 1)\}\}$. Its name comes from the fact that, on an infinite chessboard, the r-ball of a vertex v is the set of vertices that a king, starting from v, can reach in at most r moves.

It is remarkable that the best density is known for all $r \geqslant 1$, thanks to a sharp analysis of the neighbourhoods of the vertices. In the following theorem, the value 2/9 as the lower bound on $\partial_1^{Id}(G_K)$ comes from [22]* and from [15] as the upper bound.

Theorem 4.5. (a) ([15], [22]*) For $r = 1$, $\partial_1^{Id}(G_K) = \frac{2}{9} \approx 0.22222$.

(b) ([11]*) For $r > 1$, $\partial_r^{Id}(G_K) = \frac{1}{4r}$.

Figure 6 displays a 1-IdC with minimum density 2/9 on the left and a 3-IdC with minimum density 1/12 on the right. In fact, for $r \geqslant 2$, the periodic pattern of the r-IdC is similar to the one displayed here for $r = 3$: a rectangle with two rows and $2r$ columns, with only one codeword, located in the bottom left corner; the patterns are concatenated horizontally and then the resulting strip is replicated vertically with a shift of two columns. It is shown in [11]* that this construction is optimal.
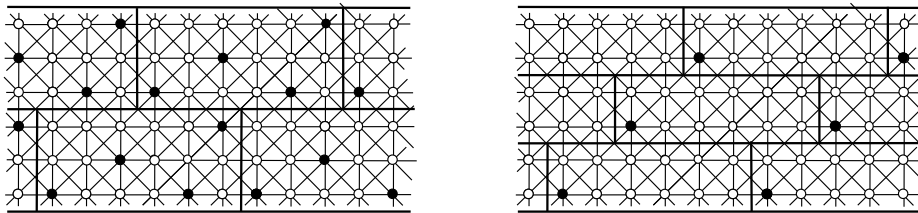

Figure 6. A periodic 1-IdC (left) and a periodic 3-IdC (right) in the king grid G_K, of density 2/9 and 1/12 respectively; codewords are in black.

4.4. The hexagonal grid

The hexagonal grid, G_H, has vertex set $V_H = \mathbb{Z} \times \mathbb{Z}$ and edge set $E_H = \{\{u, v\} : u = (i, j)$ and $u - v \in \{(0, (-1)^{i+j+1}), (1, 0)\}\}$. It is the grid with the sparsest and weakest results.

For $r = 1$, the following upper bound is from [81], improving the one provided by [20]* (equal to $3/7 \approx 0.42857$), the lower bound is from [23].

Theorem 4.6. ([23], [81])
For $r = 1$, $\frac{5}{12} \approx 0.41667 \leqslant \partial_1^{Id}(G_H) \leqslant \frac{53}{126} \approx 0.42063$.

In addition, it should be mentioned that an improvement $23/55 \approx 0.41818$ to the lower bound has been claimed in the manuscript [85].

The minimum density is known for $r = 2$, as specified by the next theorem. In this one, the lower bounds in (a) and (b) come from [61] and [71], respectively, and both upper bounds from [15]; the general lower bounds in (c) and (d) come from [10]*, and the upper bounds from [84].

Theorem 4.7. ([10]*, [15], [61], [71], [84])
(a) For $r = 2$, $\partial_2^{Id}(G_H) = \frac{4}{19} \approx 0.21053$.
   (b) For $r = 3$, $\frac{3}{25} = 0.12 \leqslant \partial_3^{Id}(G_H) \leqslant \frac{1}{6} \approx 0.16667$.
   (c) For r even with $r \geqslant 4$, $\frac{2}{5r+3} \leqslant \partial_r^{Id}(G_H) \leqslant \frac{5r+3}{6r(r+1)}$.
   (d) For r odd with $r \geqslant 5$, $\frac{2}{5r+2} \leqslant \partial_r^{Id}(G_H) \leqslant \frac{5r^2+10r-3}{(6r-2)(r+1)^2}$.

When r increases, these bounds are close to $\frac{2}{5r}$ and $\frac{5}{6r} \approx \frac{0.83333}{r}$. Better upper bounds, close to $1/(r+4)$ for r even or to $1/(r+5)$ for r odd, are known for several values of r with $r \leqslant 30$, obtained in [15] by the use of heuristics.

5. Structural properties of twin-free graphs

We shall now review some works related to the study of structural problems. First (Subsection 5.1), we consider some classic parameters of graphs, and we investigate the following question: what are the extremal values that these parameters can achieve when we restrict ourselves to r-twin-free graphs? In Subsection 5.2, we study the question when it is possible to remove a vertex from an r-twin-free graph while keeping the property of being r-twin-free, leading to the notion of *terminal graph*. Then, we study the effects of removing a vertex (Subsection 5.3) or an edge (Subsection 5.4) on the r-Id numbers.

For $S \subset V$, we denote by $G \setminus S$ the graph obtained from G by deleting the vertices of S. Similarly, for an edge e of G, we note by $G \setminus \{e\}$ the graph obtained from G by removing e.

5.1. Extremal values of usual parameters in twin-free graphs

In this subsection, we investigate the extremal values that different usual parameters of graphs can achieve in a connected r-twin-free graph. In accordance with [2]*, we consider the following parameters: the size ε (i.e., the number of edges), the minimum degree δ, the radius ρ, the diameter D, the size α of a maximum independent set (see [70] for references dealing with other parameters).

More precisely, if π stands for such a parameter, we search for the smallest value, $f_\pi(r, n)$, that π can reach over the set of connected and r-twin-free graphs with exactly n vertices ($n \geqslant 2r + 1$):

$$f_\pi(r, n) = \min\{\pi(G) : G \text{ connected, } r\text{-twin-free with } n \text{ vertices}\}.$$

We also consider the largest value that π can reach, leading to the quantity $F_\pi(r, n)$:

$$F_\pi(r, n) = \max\{\pi(G) : G \text{ connected, } r\text{-twin-free with } n \text{ vertices}\}.$$

• Number of edges ε

For any connected graph with n vertices, the size is at least $n - 1$. The path P_n on n vertices meets this bound. Thus, for any $r \geqslant 1$ and $n \geqslant 2r + 1$, $f_\varepsilon(r, n) = n - 1$.

The maximum number of edges possible for r-twin-free graphs with n vertices is known for $r = 1$, and can be achieved only by the complete graph K_n minus a maximum matching [24].

Theorem 5.1. ([24]) For $r = 1$ and $n \geqslant 3$, $F_\varepsilon(1, n) = \frac{n(n-1)}{2} - \lfloor \frac{n}{2} \rfloor$.

For $r > 1$, we are close to the exact value.

Theorem 5.2. ([2]*) (a) For $r = 2$ and n large enough, $\frac{n^2}{2} - 2n \log_2 n \lesssim F_\varepsilon(2, n) \lesssim \frac{n^2}{2} - \frac{1}{2}n \log_2 n$.
(b) For $r > 2$ and n large enough with respect to r,
$$\frac{n^2}{2} - rn \log_2 n \lesssim F_\varepsilon(r, n) \lesssim \frac{n^2}{2} - 0.63(r - 0.915)n \log_2 n.$$

• Minimum degree δ

The path P_n provides the value of $f_\delta(r, n)$ for any r and any $n \geqslant 2r + 1$: the minimum of δ for a connected and r-twin-free graph is equal to 1. Case (b) below relies on the fact that the complete graph minus a maximum matching is 1-twin-free (cf. Theorem 5.1). On the other hand, the minimum degree cannot be too large with respect to the ratio n/r, otherwise there will be r-twins; Theorem 5.3 specifies bounds stated thanks to this ratio.

Theorem 5.3. ([2]*) (a) For $r \geqslant 1$ and $n \geqslant 2r + 1$, $f_\delta(r, n) = 1$.
(b) For $r = 1$ and $n \geqslant 3$, $F_\delta(1, n) = n - 2$.
(c) For $r = 2$ and $n \geqslant 5$, $F_\delta(2, n) = \lfloor \frac{n-2}{2} \rfloor$.
(d) For $r \geqslant 3$, $F_\delta(r, 2r + 1) = 1$.
(e) For $r \geqslant 3$ and $n \geqslant 2r + 2$, set $k = \lfloor \frac{n-2}{r} \rfloor$. Then:

(e$_1$) $k - 1 \leqslant F_\delta(r, n)$ if k is odd, and $k \leqslant F_\delta(r, n)$ if k is even;
(e$_2$) for $3 \leqslant r \leqslant 5$, $F_\delta(r, n) \leqslant \frac{n}{\lfloor \frac{r}{2} \rfloor + 1} - 1$;
(e$_3$) for $r \geqslant 6$, $F_\delta(r, n) \leqslant \min\left\{ \frac{n}{\lfloor \frac{r}{2} \rfloor + 1} - 1, \frac{3n - r + 2}{2(r - 5)} \right\}$.

It is conjectured in [2]* that $F_\delta(r, n)$ is equal to $\lfloor \frac{n-2}{r} \rfloor$ for any $r \geqslant 1$ and any $n \geqslant 2r + 1$.

• Radius ρ

In any connected graph of order n, the radius is between 1 (the complete graph, the star) and $\lfloor \frac{n}{2} \rfloor$ (the path, the cycle). The study of the radius in r-twin-free graphs is easy, the results are complete.

Theorem 5.4. ([2]*) (a) For $r \geqslant 1$ and $n \geqslant 2r + 1$, $f_\rho(r, n) = r$.

(b) For $r \geqslant 1$ and $n \geqslant 2r + 1$, $F_\rho(r, n) = \lfloor \frac{n}{2} \rfloor$.

● **Diameter** D

The results on the diameter in r-twin-free graphs are also complete. Once again, Case (a) and Case (c) are related to the path P_{2r+1}, which is the only connected and r-twin-free graph on $2r + 1$ vertices, and more generally to the path P_n. Figure 7 illustrates Case (b).
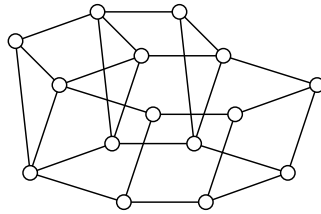

Figure 7. The case $n = 15$, $r = 3$ of Theorem 5.5, Case (b); this graph is 3-twin-free and has diameter $r + 1 = 4$.

Theorem 5.5. ([2]*) (a) For $r \geqslant 1$, $f_D(r, 2r + 1) = 2r$.

(b) For $r \geqslant 1$ and $n \geqslant 2r + 2$, $f_D(r, n) = r + 1$.

(c) For $r \geqslant 1$ and $n \geqslant 2r + 1$, $F_D(r, n) = n - 1$.

● **Size α of a maximum independent set**

In any connected graph of order n which is not the complete graph, α lies between 2 and $n - 1$ (the star). Theorem 1.2 above, from [1], contributes to Case (b) and to the lower bound of Case (c) in the following theorem.

Theorem 5.6. ([2]*) (a) For $r = 1$ and $n \geqslant 3$, $f_\alpha(1, n) = 2$.

(b) For $r \geqslant 2$, $f_\alpha(r, 2r + 1) = r + 1$.

(c) For $r \geqslant 2$ and $n \geqslant 2r + 2$, set $k = \lfloor \frac{n-2}{r} \rfloor$. Then: $r + 1 \leqslant f_\alpha(r, n) \leqslant \begin{cases} \frac{2n}{k+2} & \text{for } k \text{ even} \\ \frac{2n}{k+1} & \text{for } k \text{ odd.} \end{cases}$

The star gives Case (a) in Theorem 5.7.

Theorem 5.7. ([2]*) (a) For $r = 1$ and $n \geqslant 3$, $F_\alpha(1, n) = n - 1$.

(b) For $r \geqslant 2$ and $n \geqslant 2r + 1$, let k be the largest integer with $k + r\lceil \log_2 k \rceil \leqslant n - 1$. Then: $\max \left\{ \lceil \frac{n}{2} \rceil, k + \lceil \log_2 k \rceil \lfloor \frac{r}{2} \rfloor \right\} \leqslant F_\alpha(r, n) \leqslant n - r$.

If n is large with respect to r, then $k + \lceil \log_2 k \rceil \lfloor \frac{r}{2} \rfloor$ behaves approximately like $n - \frac{r}{2} \log_2 n$, and so behaves also the lower bound of $F_\alpha(r, n)$ in Case (b).

5.2.   Terminal graphs

We need a new definition in this section: the one of an *r-terminal graph*.

Definition 5. If $G = (V, E)$ is r-twin-free, then we say that G is r-*terminal* if for all $v \in V$, $G \setminus \{v\}$ is not r-twin-free.

In other words, G is not r-terminal if there exists a vertex $v \in V$ such that $G \setminus \{v\}$ is also r-twin-free. We are interested in the structure and number of r-terminal graphs, if they exist.

Observe that, for $r > 1$, P_{2r+1} (which is 1-twin-free, see Theorem 1.2) is r-terminal: indeed, removing any vertex from P_{2r+1} leaves a graph with r-twins. For $r = 1$, the case of P_3 is particular: removing the middle vertex yields two isolated vertices, constituting a 1-twin-free graph. This leads us to address the following questions:

- (a) Are there 1-terminal graphs?

- (b) For $r > 1$, are there r-terminal graphs other than P_{2r+1}? If so, how many are there?

As detailed below, the answer to (a) is negative. The answer to (b) is multifold: it is negative if we restrict ourselves to trees. But it is positive for $r \geqslant 3$ otherwise; the case $r = 2$ remains open.

● **The case $r = 1$**

There is no 1-terminal graph. Indeed, as observed above, for $n = 3$, P_3 is not 1-terminal. For $n \geqslant 4$, consider a vertex v such that $V \setminus \{v\}$ is a 1-IdC of G (by Theorem 1.3, such a vertex does exist). Then we can show that the graph $G \setminus \{v\}$, which may be connected or not, is still 1-twin-free and, hence, G is not 1-terminal.

The following theorem sharpens this result with respect to connectivity.

Theorem 5.8. ([12]*) Let $n \geqslant 4$ and $G = (V, E)$ be any connected 1-twin-free graph of order n. Then there exists $v \in V$ such that $G \setminus \{v\}$ is 1-twin-free and connected.

● **The case of trees**

For $r = 1$, the previous result shows that there is no 1-terminal tree. For $r > 1$, P_{2r+1} is the only r-terminal tree, as specified by the next theorem.

Theorem 5.9. ([12]*) Let $r \geqslant 1$, $n \geqslant 2r + 2$ and $T = (V, E)$ be any r-twin-free tree of order n. Then there exists a leaf $v \in V$ such that $T \setminus \{v\}$ is r-twin-free (and connected). Consequently, the only r-terminal tree for $r > 1$ is the path P_{2r+1}.

● **The case $r \geqslant 3$**

We now come back to general graphs. Then, for $r \geqslant 3$, P_{2r+1} is not the only r-terminal graph. One possible construction in order to obtain another r-terminal graph is the following: take a cycle of length $2r$ and add a neighbour to each vertex of the cycle but one; Figure 8 illustrates this construction for $r = 3$. Theorem 5.10 is more specific:
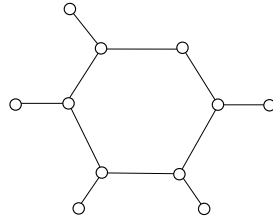
Figure 8.   A 3-terminal graph.

Theorem 5.10. ([12]*) For $r \geqslant 3$, P_{2r+1} is not the only r-terminal graph. For $r \geqslant 6$, there are infinitely many r-terminal graphs.

Observe that the construction of Theorem 5.10 does not work for $r = 2$; the problem remains open:

Open problem 1. Apart from P_5, do 2-terminal graphs exist?

Another open problem is the situation for $r \in \{3, 4, 5\}$:

Open problem 2. For $r \in \{3, 4, 5\}$, is the number of r-terminal graphs finite or infinite?

5.3.   Removing a vertex

The problem considered now is the following: given an r-twin-free graph $G = (V, E)$ and a vertex $v \in V$, and assuming that the graph $G \setminus \{v\}$ is r-twin-free, what can be said about $Id_r(G \setminus \{v\})$ with respect to $Id_r(G)$? It is easy to observe that removing a vertex may increase $Id_r(G)$. For instance, removing a vertex from the cycle C_7 on 7 vertices increases Id_2 by 1: $Id_2(C_7) = 4$ and $Id_2(P_6) = 5$ (see Figure 9).
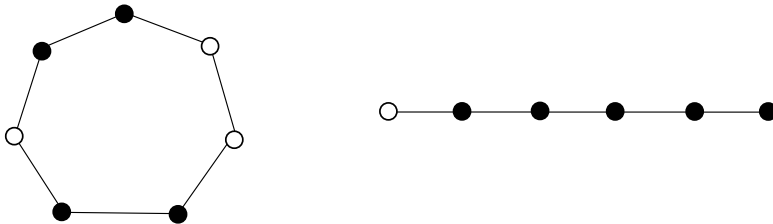
Figure 9.   Optimal 2-IdCs for C_7 and P_6; codewords are in black.

• **The case $r = 1$**
It is known from [34] that, if G and $G \setminus S$ are 1-twin-free, then we have $Id_1(G) - Id_1(G \setminus S) \leqslant |S|$. In particular, for $v \in V$, $Id_1(G) - Id_1(G \setminus \{v\}) \leqslant 1$; moreover, the two inequalities are tight. Thus, removing one vertex cannot imply a large decrease of Id_1: at most 1.

But removing a vertex may imply a large increase of Id_1, as shown in [13]*.

Theorem 5.11. ([13]*)
There exist (connected) graphs G with n vertices and a vertex v of G with $Id_1(G \setminus \{v\}) - Id_1(G) \approx 0.5n - 1.5\log_2 n$ and $Id_1(G \setminus \{v\})/Id_1(G) \approx 0.5n/\log_2 n$.

The graphs involved in this theorem are built in the following way (see Figure 10 for a partial representation of them). They are connected and bipartite. We start with k vertices c_i ($1 \leqslant i \leqslant k$), which will be, as well as the vertex v, the codewords of G. Then, for each subset $I \subseteq \{1, 2, ..., k\}$ of indices with at least two elements ($|I| \geqslant 2$), we create two vertices y_I and z_I: y_I is linked to v and to all the vertices c_i with $i \in I$, while z_I is linked to all the vertices c_i with $i \in I$ but not to v. Thus, we obtain a graph G with $n = 2(2^k - k - 1) + k + 1 = 2^{k+1} - k - 1$ vertices. As the subsets I are not singletons, $C = \{v, c_1, c_2, ..., c_k\}$ is a 1-IdC of G, with about $\log_2 n$ codewords. But, if we remove v, then y_I and z_I become almost twins for every I: the only way to separate them is to select one of them as a codeword, which, finally, requires to select about at least $n/2$ codewords.
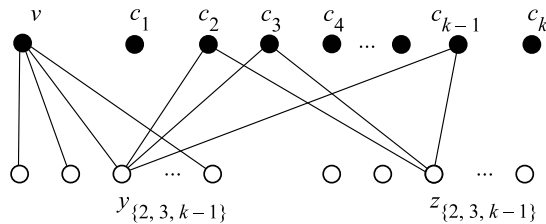

Figure 10. A partial representation (for $I = \{2, 3, k-1\}$) of the graphs involved in the proof of Theorem 5.11; codewords are in black.

M. Pelto provided a sharper result in [75] (with graphs which are not necessarily connected):

Theorem 5.12. ([75]) Let n be the order of G, S a proper subset of V and $v \in V$.

(a) If $n \geqslant 2^{|S|-1}$, then $Id_1(G \setminus S) - Id_1(G) \leqslant n - 2|S| - \left\lfloor \dfrac{n - |S|}{2^{|S|}} \right\rfloor$; the inequality is tight for n sufficiently large with respect to $|S|$.

(b) For $|S| = 1$: $Id_1(G \setminus \{v\}) - Id_1(G) \leqslant \left\lfloor \dfrac{n}{2} \right\rfloor - \varepsilon$, with $\varepsilon = 2$ for $n \in \{2, 4, 5, 6, 8\}$ and $\varepsilon = 1$ otherwise; the inequality is tight.

(c) If G is bipartite, then $Id_1(G \setminus \{v\}) - Id_1(G) \leqslant \left\lfloor \dfrac{n - \log_2(n - \log_2 n)}{2} \right\rfloor - 1$; the inequality is tight.

So, for $r = 1$, if one vertex is deleted, the identification number cannot drop by more than one, and can increase by a quantity close to $n/2$.

● **The case $r \geqslant 2$**
Now both $Id_r(G \setminus \{v\}) - Id_r(G)$ and $Id_r(G) - Id_r(G \setminus \{v\})$ can be large: $Id_r(G \setminus \{v\}) - Id_r(G)$ can increase to approximately $\frac{n}{4}$ (for even r) and $\frac{n(3r-1)}{12r}$ (for odd r), or $\frac{n(2r-2)}{2r+1}$, according to whether we want the graphs to be connected or not, and $Id_r(G) - Id_r(G \setminus \{v\})$ to approximately $\frac{n(r-1)}{r}$.

Theorem 5.13. ([13]*) (a) Let $r \geqslant 2$. There exist connected graphs G of order n and a vertex v of G with $Id_r(G \setminus \{v\}) - Id_r(G) \geqslant \dfrac{(n-1)(2r-2)}{2r+1} - 2r$.

(b) Let $r \geqslant 2$ be even. There exist connected graphs G of order n and a vertex v of G such that $G \setminus \{v\}$ is connected and $Id_r(G \setminus \{v\}) - Id_r(G) \geqslant \dfrac{n}{4} - (r+1)$.

(c) Let $r \geqslant 3$ be odd. There exist connected graphs G of order n and a vertex v of G such that $G \setminus \{v\}$ is connected and $Id_r(G \setminus \{v\}) - Id_r(G) \geqslant \dfrac{n(3r-1)}{12r} - r$.

Theorem 5.14. ([13]*) There exist connected graphs G of order n with $n = 2 \bmod r$ and a vertex v of G such that $G \setminus \{v\}$ is connected and $Id_r(G) - Id_r(G \setminus \{v\}) \geqslant \dfrac{(n-3r-1)(r-1)+1}{r}$.

The proof of this theorem is based on graphs obtained by a collection of paths P_r plus two extra vertices, including v, linked to the two extremities of each path; Figure 11 illustrates Theorem 5.14 for $n = 17$, $r = 3$.
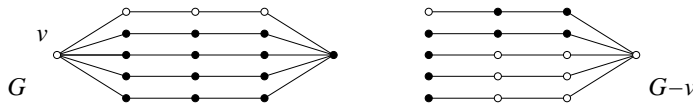

Figure 11. Illustration of Theorem 5.14 for $r = 3$: in both graphs, the black vertices form an optimal 3-IdC; $Id_3(G) - Id_3(G \setminus \{v\}) = 13 - 8 = 5$.

5.4. Removing an edge

Let G be a graph and e an edge of G. The problem considered here is similar to the previous one, but with the deletion of an edge instead of a vertex: given an r-twin-free graph $G = (V, E)$ and an edge $e \in E$, and assuming that the graph $G \setminus \{e\}$ is r-twin-free, what can be said about $Id_r(G \setminus \{e\})$ with respect to $Id_r(G)$?

● **The case $r = 1$**

Theorem 5.15. ([14]*)
If G and $G \setminus \{e\}$ are 1-twin-free, then $Id_1(G) - Id_1(G \setminus \{e\}) \in \{-2, -1, 0, 1, 2\}$ and these values can be reached by pairs of connected graphs G and $G \setminus \{e\}$.

● **The case $r \geqslant 2$**
Now the differences $Id_r(G) - Id_r(G \setminus \{e\})$ and $Id_r(G \setminus \{e\}) - Id_r(G)$ can be large, and we obtain results which slightly vary with r.

Theorem 5.16. ([14]*)
Let $k \geqslant 2$ be an arbitrary integer.

(a1) Let $r \geqslant 2$. There exist a graph G with $(r+1)k + r\lceil \log_2(k+2) \rceil + 2r$ vertices and an edge e of G with $Id_r(G) \geqslant k$ and $Id_r(G \setminus \{e\}) \leqslant r\lceil \log_2(k+2) \rceil + r + 3$.

(a2) Let $r \geqslant 5$. There exist a graph G with $(2r - 2)k + r\lceil\log_2(k + 2)\rceil + r + 3$ vertices and an edge e of G with $Id_r(G) \geqslant k$ and $Id_r(G \setminus \{e\}) \leqslant r\lceil\log_2(k + 2)\rceil + r + 1$.

(b1) For $r = 2$, there exist a graph G with $3k + 2\lceil\log_2(k + 2)\rceil + 5$ vertices and an edge e of G with $Id_2(G \setminus \{e\}) \geqslant k$ and $Id_2(G) \leqslant 2\lceil\log_2(k + 2)\rceil + 5$.

(b2) Let $r \geqslant 3$. There exist a graph G with $(2r - 2)k + r\lceil\log_2(k + 2)\rceil + r + 2$ vertices and an edge e of G with $Id_r(G \setminus \{e\}) \geqslant k$ and $Id_r(G) \leqslant r\lceil\log_2(k + 2)\rceil + r + 1$.

Closely related results can be obtained if we require that $G \setminus \{e\}$ is connected.

Note that k can be chosen arbitrarily and is linked to the order n of G and $G \setminus \{e\}$ by the relation $n = (c_1 r + c_2)k + r\lceil\log_2(k + 2)\rceil + (c_3 r + c_4)$ where the integer quadruple (c_1, c_2, c_3, c_4) takes different values in (a1), (a2), (b1) and (b2) above. This means, roughly speaking, that k is a fraction, depending on r, of n; therefore, Theorem 5.16 implies that, given $r \geqslant 2$, there is an infinite collection of graphs G of order n and two positive constants α and β with $Id_r(G) \geqslant \alpha n$ and, after deletion of a suitable edge e, $Id_r(G \setminus \{e\}) \leqslant \beta \log_2 n$ (or the other way round: $Id_r(G) \leqslant \beta \log_2 n$ and $Id_r(G \setminus \{e\}) \geqslant \alpha n$). We can see that adding or deleting one edge can lead to quite a drastic difference for the identification numbers.

We studied above what the identification numbers can become when adding or deleting vertices or edges. But we can also be interested in what an *existing* (optimal) code becomes when edges or vertices are deleted or added. This leads to the notion of *robustness*: an r-Id code C is *t-edge-robust* [43]* in G if C remains an r-IdC in all the graphs obtained from G by adding or deleting edges, with a total amount of additions and deletions at most t. This issue is dealt with, among others, in [45]* and in [66]. Different definitions exist for t-vertex-robust codes, see [43]* or [79].

6. Number of optimal codes

In this section, we are interested in graphs with many optimal codes. In his PhD thesis [74] under Iiro Honkala's supervision, M. Pelto worked on the notion of completely different optimal identifying codes in infinite grids. During his PhD defense the following question arose: how large can the number of optimal r-IdCs be? Considering once again our example of the location of a faulty processor in a multiprocessor network, this means that we want not only to use the smallest possible number of controllers, but also to have a large number of choices for their locations.

Let $\nu_r(n)$ be the maximum number of optimal (labelled) r-IdCs that a graph on n vertices may have. It is easy to observe that $\nu_r(n)$ can be exponential with respect to n. Indeed, consider a cycle C_{2r+2} on $2r + 2$ vertices. Any set of $2r + 1$ vertices is an optimal r-IdC of C_{2r+2} and there are $2r + 2$ such sets. Replicating C_{2r+2} k times provides a (disconnected) graph on $k(2r + 2)$ vertices and with $(2r + 2)^k$ optimal r-IdCs. Hence the lower bound $(2r + 2)^k \leqslant \nu_r(k(2r + 2))$, i.e., when n is divisible by $2r + 2$, $(2r + 2)^{n/(2r+2)} = 2^{n\log_2(2r+2)/(2r+2)} \leqslant \nu_r(n)$. In particular, for $r = 1$, we obtain the lower bound $2^{n/2} \leqslant \nu_1(n)$ when n is divisible by 4.

In order to obtain a larger number of optimal 1-IdCs, the notion of *separating-only code* is defined in [41]* as follows for $r = 1$:

Definition 6. Let $G = (V, E)$ be a graph. A subset C of V is said to be a separating-only code (SOC) of G if:

- $\exists\, v_0 \in V : I_1(C; v_0) = \emptyset$;

- $\forall\, v_1 \in V,\, \forall\, v_2 \in V$ with $v_1 \neq v_2 :\ I_1(C; v_1) \neq I_1(C; v_2)$.

Let $\sigma(G)$ denote the size of a minimum SOC of G and let $\mu(G)$ denote the number of optimal SOCs of G.

Remarks.
1. Some 1-twin-free graphs do not admit any SOC (for instance, the path on three vertices).
2. In a SOC C, the vertex v_0 with $I_1(C; v_0) = \emptyset$ in Definition 6 is unique (with respect to C).
3. Since adding the vertex v_0 of Definition 6 to the associated SOC leads to a 1-IdC, we have the inequalities $Id_1(G) - 1 \leqslant \sigma(G) \leqslant Id_1(G)$.

The following discussion is based on [41]*. Consider the binary Hamming space $\mathbb{F}^3$, with 8 vertices. Figure 12 displays the two non-isomorphic unlabelled optimal SOCs of $\mathbb{F}^3$, each requiring three codewords: $\sigma(\mathbb{F}^3) = 3$. There are 24 (respectively 8) labelled SOCs isomorphic to the SOC on the left (respectively right) of Figure 12: $\mu(\mathbb{F}^3) = 32$.
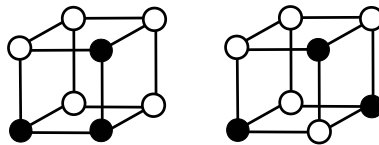

Figure 12.   The two non-isomorphic optimal SOCs of $\mathbb{F}^3$; the three black vertices are the codewords.

On the other hand, Figure 13 displays the four non-isomorphic unlabelled optimal 1-IdCs of $\mathbb{F}^3$: $Id_1(\mathbb{F}^3) = 4$. There are 6 labelled 1-IdCs isomorphic to the first unlabelled model, 24 to the second, 24 to the third and 2 to the fourth: $\nu_1(\mathbb{F}^3) = 56$.
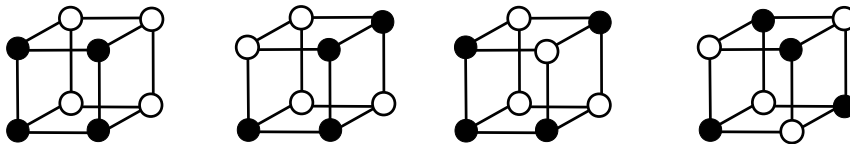

Figure 13.   The four non-isomorphic optimal 1-IdCs of $\mathbb{F}^3$; the black vertices are the codewords.

Now, consider three copies of $\mathbb{F}^3$ and add an extra vertex linked to the 24 vertices of these three copies (see Figure 14). We obtain the first graph, G_1, of a series of graphs G_q described below. In order to construct optimal 1-IdCs of G_1, we consider the following two constructions:

- combine optimal 1-IdCs of each $\mathbb{F}^3$;

- or select the extra vertex, choose one copy of $\mathbb{F}^3$ and select an optimal SOC in it, then combine them with optimal 1-IdCs inside the other two copies of $\mathbb{F}^3$.
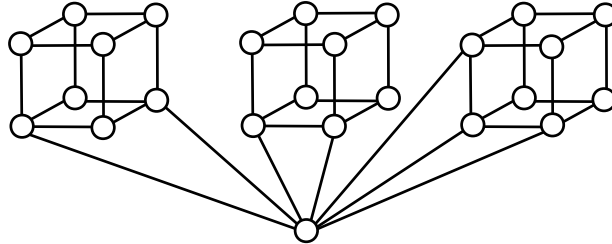

Figure 14.   The graph G_1; only some edges are represented between the vertices of the three copies of $\mathbb{F}^3$ and the extra vertex at the bottom.

In both cases, all the vertices, including the extra vertex, are covered and separated from the others. More precisely, we obtain optimal 1-IdCs of G_1. Observe that G_1 has 25 vertices with $\nu_1(G_1) = 56 \times 56 \times 56 + 3 \times 32 \times 56 \times 56 = 476\,672$. Replicating G_1 k times provides a (disconnected) graph with $n = 25k$ vertices and $476\,672^k = 2^{n \log_2(476\,672)/25} \approx 2^{0.7545n}$ optimal 1-IdCs.

In fact, this result is improved in [41]* by designing a series of connected graphs G_q defined in a similar way as G_1: for $q \geqslant 2$, G_q is obtained from G_{q-1} by replicating G_{q-1} three times and by adding an extra vertex linked to the vertices of the three copies of G_{q-1} (see Figure 15).
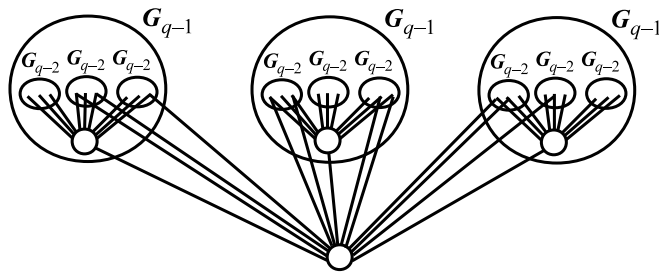

Figure 15.   The scheme of the graph G_q ($q \geqslant 3$).

Finally, based on the graphs G_q, we obtain the following theorem:

Theorem 6.1. ([41]*) For an infinite number of integers n, there exist connected graphs of order n admitting approximately $2^{0.770n}$ different optimal 1-IdCs.

For $r > 1$, other constructions, based on trees admitting many optimal r-IdCs, are proposed, leading to the next theorem, where $\frac{1+\log_2 5}{2}$ is approximately equal to 0.664.

Theorem 6.2. ([41]*) Let $r \geqslant 1$ be an integer and $\varepsilon > 0$ be a real. For an infinite number of integers n, there exist connected graphs of order n admitting $2^{\left(\frac{1+\log_2 5}{2} - \varepsilon\right)n}$ different optimal r-IdCs.

Note that a study of the set of the optimal r-identifying codes in twin-free graphs can also be found in [42]*.

7.   Variants and related concepts

We have tried to give an overview of several contributions by Iiro Honkala on identifying codes. Some variants could have been considered too.

For instance, with respect to the context exposed in Section 1, we could discuss the situation when a processor cannot control itself, i.e., when we consider the open neighbourhood $N_r(v) = B_r(v) \setminus \{v\}$ instead of the whole ball $B_r(v)$ for any vertex v, leading to the concept of *open neighbourhood* identification. This was introduced in [49]* (see also [82]). This definition is one of several definitions for so-called *fault-tolerant codes*, where different scenarii are considered for the alarms given by the codewords.

The strongly $(r, \leqslant \ell)$-IdCs (of Definition 4) are also related to open neighbourhoods. Indeed, for $\ell = 1$, the definition can be reformulated as follows: a code C is strongly $(r, \leqslant 1)$-identifying if for all $v_1 \in V$, $v_2 \in V$ with $v_1 \neq v_2$, the sets $\{B_r(v_1) \cap C, N_r(v_1) \cap C\}$ and $\{B_r(v_2) \cap C, N_r(v_2) \cap C\}$ are disjoint. See [40]*, where the best density for a strongly $(1, \leqslant 1)$-IdC in the triangular grid is proved to be $6/19 \approx 0.3158$, or [64] and [69].

A more drastic variant consists in adopting other patterns instead of the r-balls, as in [54]* (see also, e.g., [35], [60] or [80]), especially for the grids defined in $\mathbb{Z} \times \mathbb{Z}$ where the patterns may be no longer isotropic (for example, we may adopt a horizontal path centred – or not... – on v instead of $B_r(v)$).

But surely the most studied variant of r-IdCs is the one of r-*locating-dominating codes* (see [70]). They are defined as follows:

Definition 7. An r-*locating-dominating code* C of a graph $G = (V, E)$ is a subset of V fulfilling the two properties:

- $\forall\, v \in V : I_r(C; v) \neq \emptyset$;

- $\forall\, v_1 \in V \setminus C,\, \forall\, v_2 \in V \setminus C$ with $v_1 \neq v_2 : \; I_r(C; v_1) \neq I_r(C; v_2)$.

So, with respect to r-IdCs, the difference is that the r-separation property must be fulfilled only by non-codewords (thus, any r-IdC is also an r-locating-dominating code). Locating-dominating codes were introduced by P. J. Slater in 1983 [83] (for a more easily accessible source, see [76]). Since this date, they have been extensively studied (see [70] and, for updated references, the ongoing bibliography [58]), and Iiro Honkala contributed much to their study. But summarizing Iiro's contributions to that topic would require another article...

References

[1] Auger D. Induced paths in twin-free graphs, *Electron. J. Combin.* **15(1)** (2008), N17. doi:10.37236/892.

[2]* Auger D, Charon I, Honkala I, Hudry O, and Lobstein A. Edge number, minimum degree, maximum independent set, radius and diameter in twin-free graphs, *Adv. Math. Commun.* **3(1)** (2009), 97–114. Erratum 2009. **3(4)**:429–430. doi:10.3934/amc.2009.3.97.

[3] Auger D, Charon I, Hudry O, and Lobstein A. Complexity results for identifying codes in planar graphs, *Int. Trans. Oper. Res.* 2010. **17(6)**:691–710. doi:10.1111/j.1475-3995.2009.00750.x.

[4] Ben-Haim Y, and Litsyn S. Exact minimum density of codes identifying vertices in the square grid, *SIAM J. Discrete Math.* 2005. **19**:69–82. doi:10.1137/S089548010444408.

[5] Bertrand N. *Codes identifiants et codes localisateurs-dominateurs sur certains graphes*, mémoire de stage de maîtrise, ENST, Paris, France 2001, 28 pages.

[6]* Blass U, Honkala I, and Litsyn S. On the size of identifying codes, *Lecture Notes in Comput. Sci.* 1999. **1719**:142–147. doi:10.1007/3-540-46796-3_14.

[7]* Blass U, Honkala I, and Litsyn S. On binary codes for identification, *J. Combin. Des.* **8** (2000), 151–156. `https://doi.org/10.1002/(SICI)1520-6610(2000)8:2<151`.

[8]* Blass U, Honkala I, and Litsyn S. Bounds on identifying codes, *Discrete Math.*, Selected Papers in Honor of Helge Tverberg, 2001. **241**:119–128.

[9] Charon I, Cohen G, Hudry O, and Lobstein A. New identifying codes in the binary Hamming space, *European J. Combin.* 2010. **31**:491–501. doi:10.1016/j.ejc.2009.03.032.

[10]* Charon I, Honkala I, Hudry O and Lobstein A. General bounds for identifying codes in some infinite regular graphs, *Electron. J. Combin.* 2001. **8(1)**:R39.

[11]* Charon I, Honkala I, Hudry O, and Lobstein A. The minimum density of an identifying code in the king lattice, *Discrete Math.* 2004. **276**:95–109. doi:10.1016/S0012-365X(03)00306-6.

[12]* Charon I, Honkala I, Hudry O and Lobstein A, Structural properties of twin-free graphs, *Electron. J. Combin.* 2007. **14(1)**:R16.

[13]* Charon I, Honkala I, Hudry O and Lobstein A, Minimum sizes of identifying codes in graphs differing by one vertex, *Cryptogr. Commun.* 2013. **5**:119–136.

[14]* Charon I, Honkala I, Hudry O and Lobstein A. Minimum sizes of identifying codes in graphs differing by one edge, *Cryptogr. Commun.* 2014. **6**:157–170. doi:10.1007/s12095-013-0094-x.

[15] Charon I, Hudry O, and Lobstein A. Identifying codes with small radius in some infinite regular graphs, *Electron. J. Combin.* 2002. **9(1)**: R11.

[16] Charon I, Hudry O, and Lobstein A. Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard, *Theoret. Comput. Sci.* 2003. **290**:2109–2120. doi:10.1016/S0304-3975(02)00536-4.

[17] Charon I, Hudry O, and Lobstein A. Extremal cardinalities for identifying and locating-dominating codes, *Discrete Math.* 2007. **307**:356–366. doi:10.1016/j.disc.2005.09.027.

[18]* Cohen G, Gravier S, Honkala I, Lobstein A, Mollard M, Payan Ch, and Zémor G. Improved identifying codes for the grid, *Electron. J. Combin.* 1999. **6(1)**:R19.

[19]* Cohen G, Honkala I, Litsyn S, and Lobstein A. *Covering Codes*, North-Holland Publishing Co. (1997).

[20]* Cohen G, Honkala I, Lobstein A, and Zémor G. Bounds for codes identifying vertices in the hexagonal grid, *SIAM J. Discrete Math.* 2000. **13**:492–504. doi:10.1137/S0895480199360990.

[21]* Cohen G, Honkala I, Lobstein A, and Zémor G. On identifying codes, in A. Barg and S. Litsyn (eds), *Proceedings of DIMACS Workshop on Codes and Association Schemes '99, Piscataway, USA*, American Mathematical Society. 2001. **56**:97–109.

[22]* Cohen G, Honkala I, Lobstein A, and Zémor G. On codes identifying vertices in the two-dimensional square lattice with diagonals, *IEEE Trans. Comput.* 2001. **50**:174–176. doi:10.1109/12.908992.

[23] Cukierman A, and Yu G. New bounds on the minimum density of an identifying code for the infinite hexagonal grid, *Discrete Appl. Math.* 2013. **161**:2910–2924. doi:10.1016/j.dam.2013.06.002.

[24] Daniel M. *Codes identifiants*, mémoire de DEA ROCO. Université Joseph Fourier, Grenoble, France. 2003, 46 pages.

[25] Dhanalakshmi R, and Durairajan C. Constructions of r-identifying codes and $(r, \leq l)$-identifying codes, *Indian J. Pure Appl. Math.* 2019. **50**:531–547. doi:10.1007/s13226-019-0343-6.

[26] Exoo G, Junnila V, and Laihonen T. On location-domination of set of vertices in cycles and paths, *Congr. Numer.* 2010. **202**:97–112.

[27] Exoo G, Junnila V, and Laihonen T, and Ranto S. Locating vertices using codes, *Congr. Numer.* 2008. **191**:143–159.

[28] Exoo G, Junnila V, and Laihonen T, and Ranto S. Upper bounds for binary identifying codes, *Adv. in Appl. Math.* 2009. **42**:277–289. doi:10.1016/j.aam.2008.06.004.

[29] Exoo G, Junnila V, and Laihonen T, and Ranto R. Improved bounds on identifying codes in binary Hamming spaces, *European J. Combin.* 2010. **31**:813–827. doi:10.1016/j.ejc.2009.09.002.

[30] Exoo G, Laihonen T, and Ranto S. Improved upper bounds on binary identifying codes, *IEEE Trans. Inform. Theory*. 2007. **53**:4255–4260. doi:10.1109/TIT.2007.907434.

[31] Exoo G, Laihonen T, and Ranto S. New bounds on binary identifying codes, *Discrete Appl. Math.* 2008. **156**:2250–2263. doi:10.1016/j.dam.2007.09.017.

[32] Foucaud F. *Algorithmic and combinatorial aspects of identifying codes in graphs*, PhD thesis, Université Bordeaux 1, France 2012.

[33] Foucaud F. Decision and approximation complexity for identifying codes and locating-dominating sets in restricted graph classes. *J. Discrete Alg.* 2015. **31**:48–68. doi:10.1016/j.jda.2014.08.004.

[34] Foucaud F, Guerrini E, Kovše M, Naserasr R, Parreau A, and Valicov P. Extremal graphs for the identifying code problem, *European J. Combin.* 2011. **32**:628-638. doi:10.1016/j.ejc.2011.01.002.

[35] Foucaud F, and Kovše M. Identifying path covers in graphs, *J. Discrete Alg.* 2013. **23**:21–34. doi:10.1016/j.jda.2013.07.006.

[36] Foucaud F, Laihonen T, and Parreau A. An improved lower bound for $(1, \leqslant 2)$-identifying codes in the king grid, *Adv. Math. Commun.* 2014. **8**:35–52. doi:10.3934/amc.2014.8.35.

[37] Gravier S, and Moncel J. Construction of codes identifying sets of vertices, *Electron. J. Combin.* 2005. **12(1)**:R13. doi:10.37236/1910.

[38] Gravier S, and Moncel J. On graphs having a $V \setminus \{x\}$ set as an identifying code, *Discrete Math.* 2007. **307**:432–434. doi:10.1016/j.disc.2005.09.035.

[39]* Honkala I. On the identifying radius of codes, *Proceedings of the Seventh Nordic Combinatorial Conference, Turku,* TUCS Gen. Publ. 1999. **15**:39–43.

[40]* Honkala I. An optimal strongly identifying code in the infinite triangular grid, *Electron. J. Combin.* 2010. **17(1)**:R91. doi:10.37236/363.

[41]* Honkala I, Hudry O, and Lobstein A. On the number of optimal identifying codes in a twin-free graph, *Discrete Appl. Math.* 2015. **180**:111–119. doi:10.1016/j.dam.2014.08.020.

[42]* Honkala I, Hudry O, and Lobstein A. On the ensemble of optimal identifying codes in a twin-free graph, *Cryptogr. Commun.* 2016. **8**:139–153. doi:10.1007/s12095-015-0148-3.

[43]* Honkala I, Karpovsky MG, and Levitin LB. On robust and dynamic identifying codes, *IEEE Trans. Inform. Theory* 2006. **52**: 599–612. doi:10.1109/TIT.2005.862097.

[44]* Honkala I, and Laihonen T. On identifying codes in the triangular and square grids, *SIAM J. Comput.* 2004. **33**:304–312. doi:10.1137/S009753970343311.

[45]* Honkala I, and Laihonen T. On identifying codes that are robust against edge changes, *Inform. and Comput.* 2007. **205**:1078–1095. doi:10.1016/j.ic.2007.01.003.

[46]* Honkala I, and Laihonen T. On a new class of identifying codes in graphs, *Inform. Process. Lett.* 2007. **102**:92–98. doi:10.1016/j.ipl.2006.11.007.

[47]* Honkala I, and Laihonen T. On vertex-robust identifying codes of level three, *Ars Combin.* 2010. **94**:115–127.

[48]* Honkala I, Laihonen T, and Ranto S. On codes identifying sets of vertices in Hamming spaces, *Des. Codes Cryptogr.* 2001. **24**:193–204. doi:10.1023/A:1011256721935.

[49]* Honkala I, Laihonen T, and Ranto S. On strongly identifying codes, *Discrete Math.* 2002. **254**:191–205. doi:10.1016/S0012-365X(01)00357-0.

[50]* Honkala I, Laihonen T, and Ranto S. On locating-dominating codes in binary Hamming spaces, *Discrete Math. Theor. Comput. Sci.* 2004. **6**:265–282. doi:10.46298/dmtcs.322.

[51]* Honkala I, and Lobstein A. On the density of identifying codes in the square lattice, *J. Combin. Theory*, Ser. B 2002. **85**:297–306. doi:10.1006/jctb.2001.2106.

[52]* Honkala I, and Lobstein A. On identifying codes in binary Hamming spaces, *J. Combin. Theory*, Ser. A 2002.**99**:232–243. doi:10.1006/jcta.2002.3263.

[53]* Honkala I, and Lobstein A, On the complexity of the identification problem in Hamming spaces, *Acta Inform.* 2002. **38**:839–845. doi:10.1007/s00236-002-0093-4.

[54]* Honkala I, and Lobstein A. On identification in Z^2 using translates of given patterns, *J. UCS* 2003. **9(10)**:1204–1219. doi:10.3217/jucs-009-10-1204.

[55] Hudry O, and Lobstein A. More results on the complexity of identifying problems in graphs, *Theor. Comput. Sci.* 2016. **626**:1–12. doi:10.1016/j.tcs.2016.01.021.

[56] Hudry O, and Lobstein A. Unique (optimal) solutions: complexity results for identifying and locating-dominating codes, *Theor. Comput. Sci.* 2019. **767**:83–102. doi:10.1016/j.tcs.2018.09.034.

[57] Janson S, and Laihonen T. On the size of identifying codes in binary hypercubes, *J. Combin. Theory*, Ser. A 2009. **116**:1087–1096. doi:10.1016/j.jcta.2009.02.004.

[58] Jean D. Watching systems, identifying, locating-dominating and discriminating codes in graphs, `https://dragazo.github.io/bibdom/main.pdf`

[59] Junnila V. New lower bound for 2-identifying code in the square grid, *Discrete Appl. Math.* 2013. **161**:2042–2051. doi:10.1016/j.dam.2013.02.032.

[60] Junnila V, and Laihonen T. Identification in $\mathbb{Z}^2$ using Euclidean balls, *Discrete Appl. Math.* 2011. **159**:335–343. doi:10.1016/j.dam.2010.12.008.

[61] Junnila V, and Laihonen T. Optimal lower bound for 2-identifying codes in the hexagonal grid, *Electron. J. Combin.* 2012. **19(2)**:P38. doi:10.37236/2414.

[62] Karpovsky MG, Chakrabarty K, and Levitin LB. On a new class of codes for identifying vertices in graphs, *IEEE Trans. Inform. Theory* 1998. **44**:599–611.

[63] Laihonen T. Sequences of optimal identifying codes, *IEEE Trans. Inform. Theory* 2002. **48**:774–776. doi:10.1109/18.986043.

[64] Laihonen T. Optimal codes for strong identification, *European J. Combin.* 2002. **23**:307–313. doi:10.1006/eujc.2002.0571.

[65] Laihonen T. On optimal edge-robust and vertex-robust $(1, \leq \ell)$-identifying codes, *SIAM J. Discrete Math.* 2005. **18**(4):825–834. doi:10.1137/S0895480104440754.

[66] Laihonen T. Optimal t-edge-robust r-identifying codes in the king lattice, *Graphs Combin.* 2006. **22**:487–496. doi:10.1007/s00373-006-0682-z.

[67] Laihonen T. On edge-robust $(1, \leq \ell)$-identifying codes in binary Hamming spaces, *Int. J. Pure Appl. Math.* 2007. **36**:87–102. doi:10.48550/arXiv.cs/0703066.

[68] Laihonen T, and Ranto S. Codes identifying sets of vertices, in *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (Melbourne, 2001)*, *Lecture Notes in Comput. Sci.* 2001. (2227):82–91. doi:10.1007/3-540-45624-4_9.

[69] Laihonen T, and Ranto S. Families of optimal codes for strong identification, *Discrete Appl. Math.* 2002. **121**(1-3):203–213. doi:10.1016/S0166-218X(01)00248-7.

[70] Lobstein A, Hudry O, and Charon I. Locating-domination and identification, in T. Hayes, S. Hedetniemi, M. Henning (eds), *Topics in Domination in Graphs*, Springer 2020. pp. 251–299. `https://hal.science/hal-02916929`.

[71] Martin R, and Stanton B. Lower bounds for identifying codes in some infinite grids, *Electron. J. Combin.* 2010. **17(1)**:R122. doi:10.48550/arXiv.1004.3281.

[72] Moncel J. Monotonicity of the minimum cardinality of an identifying code in the hypercube, *Discrete Appl. Math.* 2006. **154**:898–899. doi:10.1016/j.dam.2005.05.030.

[73] Pelto M. New bounds for $(r, \leqslant 2)$-identifying codes in the infinite king grid, *Cryptogr. Commun.* 2010. **2**:41–47. doi:10.1007/s12095-009-0015-1.

[74] Pelto M. On identifying and locating-dominating codes in the infinite king grid, PhD thesis, University of Turku, Finland, 2012.

[75] Pelto M. Maximum difference about the size of optimal identifying codes in graphs differing by one vertex, *Discrete Math. Theor. Comput. Sci.* 2015. **17(1)**:339–356. doi:10.46298/dmtcs.2107.

[76] Rall DF, and Slater PJ. On location-domination numbers for certain classes of graphs, *Congr. Numer.* 1984. **45**:97-106.

[77]* Ranto S, Honkala I, and Laihonen T. Two families of optimal identifying codes in binary Hamming spaces, *IEEE Trans. Inform. Theory* 2002. **48**:1200–1203.

[78] Rao NSV. Computational complexity issues in operative diagnosis of graph-based systems, *IEEE Trans. Comput.* 1993. **42**:447–457. doi:10.1109/12.214691.

[79] Ray S, Ungrangsi R, Pellegrini F De, Trachtenberg A, and Starobinski D. Robust location detection in emergency sensor networks, *Proceedings of INFOCOM 2003, San Francisco, USA* (2003):1044–1053. doi:10.1109/INFCOM.2003.1208941.

[80] Rosendahl P. On the identification of vertices using cycles, *Electron. J. Combin.* 2003. **10(1)**:R7. doi:10.37236/1700.

[81] Salo V, and Törmä I. Finding codes on infinite grids automatically, to appear in *Fund. Inform.* 2024. (see also arXiv:2303.00557).

[82] Seo SJ, and Slater PJ. Open neighborhood locating-dominating sets, *Australas. J. Combin.* 2010. **46**:109–119. ID:15716275.

[83] Slater PJ. Domination and location in graphs, Research Report No. 93, National University of Singapore 1983.

[84] Stanton B. Improved bounds for r-identifying codes of the hex grid, *SIAM J. Discrete Math.* 2011. **25**:159–169. doi:10.1137/100791610.

[85] Stolee D. Automated Discharging Arguments for Density Problems in Grids, arXiv:1409.5922, ID:15143804.