Structural Properties of Twin-Free Graphs

Irène Charon
GET - Télécom Paris & CNRS - LTCI UMR 5141
46, rue Barrault, 75634 Paris Cedex 13 - France
charon@infres.enst.fr

Iiro Honkala
Department of Mathematics - University of Turku
FIN-20014 Turku - Finland
honkala@utu.fi

Olivier Hudry
GET - Télécom Paris & CNRS - LTCI UMR 5141
46, rue Barrault, 75634 Paris Cedex 13 - France
hudry@infres.enst.fr

Antoine Lobstein
CNRS - LTCI UMR 5141 & GET - Télécom Paris
46, rue Barrault, 75634 Paris Cedex 13 - France
lobstein@infres.enst.fr

Submitted: Jan 31, 2006; Accepted: Dec 21, 2006; Published: Jan 29, 2007
Mathematics Subject Classification: 05C75, 05C35
Key Words: Graph Theory, Identifying Codes, Trees, Paths

Abstract

Consider a connected undirected graph $G = (V, E)$, a subset of vertices $C \subseteq V$, and an integer $r \geq 1$; for any vertex $v \in V$, let $B_r(v)$ denote the ball of radius r centered at v, i.e., the set of all vertices linked to v by a path of at most r edges. If for all vertices $v \in V$, the sets $B_r(v) \cap C$ are all nonempty and different, then we call C an r-identifying code. A graph admits at least one r-identifying code if and only if it is r-twin-free, that is, the sets $B_r(v)$, $v \in V$, are all different.

We study some structural problems in r-twin-free graphs, such as the existence of the path with $2r + 1$ vertices as a subgraph, or the consequences of deleting one vertex.

1 Introduction

Given a connected, undirected, finite graph $G = (V, E)$ and an integer $r \geq 1$, we define $B_r(v)$, the *ball* of radius r centered at $v \in V$, by

$$B_r(v) = \{x \in V : d(x, v) \leq r\},$$

where $d(x, v)$ denotes the number of edges in any shortest path between v and x.

Whenever $d(x, v) \leq r$, we say that x and v *r-cover* each other (or simply *cover* if there is no ambiguity). A set $X \subseteq V$ covers a set $Y \subseteq V$ if every vertex in Y is covered by at least one vertex in X.

Two vertices $v_1, v_2 \in V$ such that $B_r(v_1) = B_r(v_2)$ are called *r-twins* or *twins*. If G has no r-twins, that is, if

$$\forall\, v_1, v_2 \in V\ (v_1 \neq v_2),\ B_r(v_1) \neq B_r(v_2), \tag{1}$$

then we say that G is *r-twin-free* or *twin-free*.

A graph with one vertex is trivially twin-free, and generally we consider graphs with at least two vertices.

Twin-free graphs are of interest because they are strongly connected with identifying codes, which we now define.

A *code* C is a nonempty set of vertices, and its elements 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 which r-cover 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*, or *separated*, by code C.

A code C is called *r-identifying*, or *identifying*, if the sets $K_{C,r}(v), v \in V$, are all nonempty and distinct [11]. In other words, all vertices must be covered and pairwise separated by C.

Remark 1. For given $G = (V, E)$ and integer r, the graph G admits at least one r-identifying code if and only if it is r-twin-free. Indeed, if for all $v_1, v_2 \in V$, $B_r(v_1)$ and $B_r(v_2)$ are different, then $C = V$ is r-identifying. Conversely, if for some $v_1, v_2 \in V$, $B_r(v_1) = B_r(v_2)$, then for any code $C \subseteq V$, we have $K_{C,r}(v_1) = K_{C,r}(v_2)$. This is why r-twin-free graphs are also called *r-identifiable*. For instance, there is no r-identifying code in a complete graph (or clique) with at least two vertices.

Remark 2. If G is not connected, we simply consider each of its connected components, and apply the above definitions.

We recall that an *induced subgraph* of $G = (V, E)$ is a graph $G_1 = (V_1, E_1)$ where $V_1 \subseteq V$ and E_1 is the set of edges in E which have both ends in V_1, whereas a *subgraph* is a graph $G_2 = (V_2, E_2)$ where $V_2 \subseteq V$ and E_2 is included in the set of edges in E which have both ends in V_2.

For $X \subseteq V$, we denote by G_X the induced subgraph with vertex set $V \setminus X$, and for $x \in V$, we set $G_x = G_{\{x\}}$.

In the following, n will denote the number of vertices of G. For any integer $q > 0$, P_q will denote the path on q vertices, and the length of P_q will be equal to $q-1$, its number of edges. Moreover, if v_1, v_2, …, v_q denote the vertices in P_q, we shall assume that these vertices are numbered in such a way that the edges in P_q are $\{v_i, v_{i+1}\}$ for $1 \leq i < q$. The cycle of length q, $\mathcal{C}_q$, with q vertices and q edges, consists of P_q to which we add the edge $\{v_q, v_1\}$.

The motivations for identifying codes come, for instance, from fault diagnosis in multiprocessor systems. Such a system can be modeled as a graph where vertices are processors and edges are links between processors. Assume that at most one of the processors is malfunctioning and we wish to test the system and locate the faulty processor. For this purpose, some processors (constituting the code) will be selected and assigned the task of testing their neighbourhoods (i.e., the vertices at distance at most r). Whenever a selected processor (i.e., a codeword) detects a fault, it sends an alarm signal, saying that one element in its neighbourhood is malfunctioning, and we require that we can uniquely tell the location of the malfunctioning processor based only on the information which ones of the codewords gave the alarm.

Identifying codes were introduced in [11], and they constitute now a topic of their own, studied in a large number of various papers, investigating particular graphs or families of graphs (such as certain infinite regular grids, trees, chains, cycles, or the k-cube), dealing with complexity issues, or using heuristics such as the noising methods for the construction of small codes. See, e.g., [2], [3], [4], [5], [6], [10], [13], and references therein, or [14].

In Section 2, we show that any connected r-twin-free graph contains the path P_{2r+1} as a subgraph; we conjecture that any connected r-twin-free graph contains the path P_{2r+1} as an *induced* subgraph (and we prove this for the path P_{r+2}).

In Section 3, we study the consequences of the deletion of a vertex in a connected r-twin-free graph; the results differ according to the values of r. In particular, we prove that all connected r-twin-free graphs remain r-twin-free after deleting one appropriate vertex *when* $r = 1$, and that the same is true for *all trees*, except P_{2r+1}.

Some of these results were already stated without proofs in [7].

2 The existence of P_{2r+1} in r-twin free graphs

In this section, we prove that any connected r-twin-free graph G contains P_{2r+1} as a subgraph, for all $r \geq 1$. We conjecture that G even contains P_{2r+1} as an induced subgraph, and prove it for P_{r+2}.

Theorem 1 *Let $r \geq 1$ and G be any connected r-twin-free graph with at least two vertices. Then P_{2r+1} is a subgraph of G.*

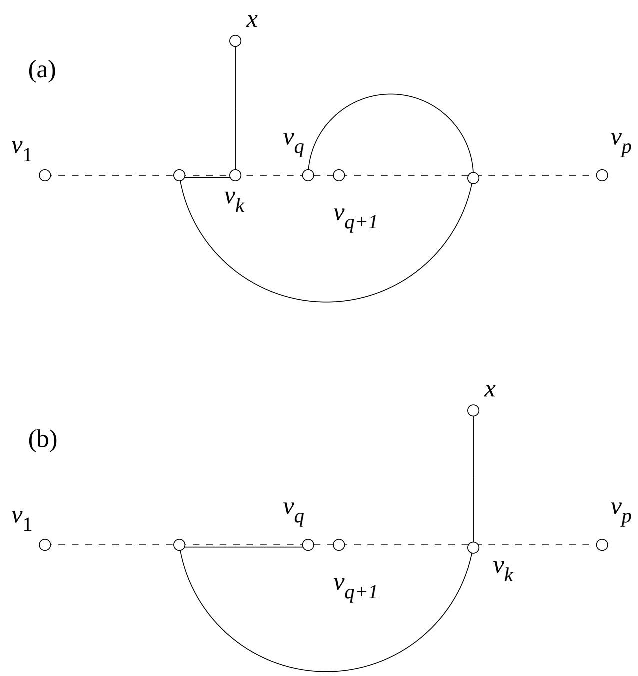

Figure 1: The path P^* is in dashed line, the path P_{r+1} is in plain line.

Proof. Let G fulfill the conditions of Theorem 1. Consider a longest path P^* of G, with p vertices, $v_1, v_2, \ldots, v_p$, and assume that $p \leq 2r$. We set $q = \lfloor p/2 \rfloor$.

Since G is r-twin-free, for any two vertices w and z, there is at least one vertex which r-covers exactly one of them; we shall say that such a vertex separates w and z.

It is easy to check that the vertices of P^* cannot separate v_q and v_{q+1}, because the length of P^* is too small. Therefore, there is a vertex $x \notin P^*$ which separates v_q and v_{q+1}. Then two cases occur.

Case (1): x covers v_q, not v_{q+1}.

Thus we have $d(x, v_q) \leq r$ and $d(x, v_{q+1}) \geq r + 1$. Since v_q and v_{q+1} are adjacent, we have $d(x, v_q) = r$ and $d(x, v_{q+1}) = r + 1$. Let P_{r+1} be a path of length r between x and v_q. The paths P^* and P_{r+1} have at least v_q in common. Let v_k be the vertex belonging simultaneously to P^* and P_{r+1}, and which is the closest to x (see Figure 1): the vertices of P_{r+1} between x (included) and v_k (excluded) do not belong to P^*. We have now two subcases, according to the value of k with respect to q.

Subcase (1.a): $1 \leq k \leq q$:

v_k is on P^* between v_1 and v_q (see Figure 1.a). Consider the path P obtained by the concatenation of the part of P_{r+1} between x and v_k and the part of P^* between v_k and v_p (note that, thanks to the definition of v_k, there is no cycle and P is indeed a path; this will also hold in all other cases). The length of P is at least $r + p - q = r + \lceil p/2 \rceil$ (there are r edges from x to v_q and $p - q$ edges from v_q to v_p); hence a contradiction: P would be longer than P^* (the length of P^* is equal to $p - 1$ and we assumed that p is less than

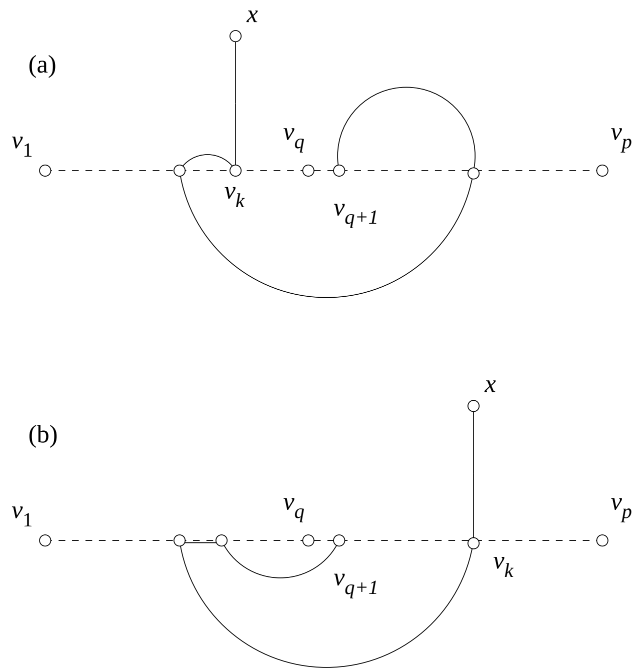

Figure 2: The path P^* is in dashed line, the path P_{r+1} is in plain line.

$2r+1$).

Subcase (1.b): $q+1 \leq k \leq p$:

Now v_k is on P^* between v_{q+1} and v_p (see Figure 1.b; it is obvious that k cannot be equal to $q+1$, but this does not matter). Consider the path P obtained by the concatenation of the part of P^* going from v_1 to v_k and the part of P_{r+1} going from v_k to x. The length of P is greater than or equal to $q+r+1$: there are q edges from v_1 to v_{q+1} and at least $r+1$ edges from v_{q+1} to x, since $d(x, v_{q+1}) = r+1$. So P is longer than P^*, a contradiction with the definition of P^*.

Case (2): x covers v_{q+1}, not v_q.

As v_q and v_{q+1} play similar roles if p is even, we may assume that p is odd; otherwise, see Case (1). Hence, $p = 2q+1$. Similarly to Case (1), we have $d(x, v_q) = r+1$ and $d(x, v_{q+1}) = r$. As above, let P_{r+1} be a path of length r between x and v_{q+1} and let v_k be the vertex belonging simultaneously to P^* and P_{r+1}, and which is the closest to x (see Figure 2). Again, we have two subcases, according to the value of k with respect to q.

Subcase (2.a): $1 \leq k \leq q$:

v_k is on P^* between v_1 and v_q (see Figure 2.a; similarly to Subcase (1.b), k cannot be equal to q, but this still does not matter). Consider the path P obtained by the concatenation of the part of P_{r+1} between x and v_k and the part of P^* between v_k and v_p. Because $d(x, v_q) = r+1$, the length of P is greater than or equal to $r+1+p-q$, which is greater than the length of P^*, hence a contradiction.

Subcase (2.b): $q+1 \leq k \leq p$:

Now v_k is on P^* between v_{q+1} and v_p (see Figure 2.b). Consider the path P obtained by the concatenation of the part of P^* going from v_1 to v_k and the part of P_{r+1} going from v_k to x. The length of P is greater than or equal to $q + r$ (there are q edges from v_1 to v_{q+1} and at least r edges from v_{q+1} to x, since $d(x, v_{q+1}) = r$). So P is again longer than P^*, a contradiction.

So we have seen that in all cases, P^* contains at least $2r + 1$ vertices. $\triangle$

As an immediate consequence, we obtain a result which was proved in [12].

Corollary 2 *[12, Prop. 4.1] Let $r \geq 1$ and G be any connected r-twin-free graph with $n > 1$ vertices. Then we have $n \geq 2r + 1$.* $\triangle$

The bound of Corollary 2 is the best possible, since, for any $r \geq 1$, the paths P_n are r-twin-free for any $n \geq 2r + 1$ (see [2] for a study of r-identifying codes on paths). We conjecture that a statement stronger than that in Theorem 1 holds:

Conjecture 3 *Let $r \geq 1$ and G be any connected r-twin-free graph with at least two vertices. Then P_{2r+1} is an induced subgraph of G.*

This conjecture is true for $r = 1$, as we now show, using the following lemma.

Lemma 4 *Let $r \geq 1$ and G be any connected r-twin-free graph with at least two vertices. Then P_{r+2} is an induced subgraph of G.*

Proof. Consider two distinct vertices a and b with $d(a, b) = 1$. Since they are not twins, there exists (without loss of generality) a vertex x ($x \neq b$) such that $x \in B_r(b)$ and $x \notin B_r(a)$. Because a and b are adjacent, we have $d(b, x) = r$ and $d(a, x) = r + 1$, which means that a and any vertices forming a shortest path between b and x constitute a path with $r + 2$ vertices and no chord. $\triangle$

Corollary 5 *Let G be any connected one-twin-free graph with at least two vertices. Then P_3 is an induced subgraph of G.* $\triangle$

3 Induced subgraphs with one vertex less

Let $r \geq 1$ and $G = (V, E)$ be a connected, r-twin-free graph with at least two vertices; we say that G is *r-terminal* if for all vertices $x \in V$, G_x is not r-twin-free, and that G is not r-terminal if there exists a vertex $x \in V$ such that G_x is r-twin-free. We denote by $\mathcal{T}_r$ the set of r-terminal graphs.

Thanks to Corollary 2, we need only to consider graphs with $n \geq 2r + 1$. Using Theorem 1, it is easy to see that if $n = 2r+1$, the only r-twin-free graph is the path P_{2r+1}, for $r \geq 1$, and the only r-terminal graph is P_{2r+1}, for $r > 1$ (the case of P_3 is particular, because removing the middle vertex yields two isolated vertices which constitute a one-twin-free graph — see Remark 2).

In this section, we address the following issue: are the paths P_{2r+1} (with the exception of P_3) the only r-terminal graphs?

The answer to this question is multifold: it is positive if $r = 1$ (Corollary 7), or if we restrict ourselves, for any r, to *trees* (Corollary 11); it is negative if $r \geq 3$ (Theorem 12). The case $r = 2$ remains open.

3.1 The case $r = 1$

The following lemma can be found in [1], [9], [8].

Lemma 6 *Let $n \geq 3$ be an integer, and G be any connected one-twin-free graph with n vertices. Then there exists a one-identifying code in G with $n - 1$ vertices.*

Proof. We refer to [9], which gives an elegant proof of a more general result. $\triangle$

An easy consequence of Lemma 6 is that $\mathcal{T}_1 = \emptyset$:

Corollary 7 *Let $n \geq 3$ be an integer, and $G = (V, E)$ be any connected one-twin-free graph with n vertices. Then G is not one-terminal.*

Proof. If $n = 3$, $G = P_3$, so we can assume that $n \geq 4$. By Lemma 6, there is a one-identifying code C of size $n - 1$ in G. Consider G_x with $\{x\} = V \setminus C$ (G_x may be connected or not); obviously, C is still one-identifying in G_x, because removing x does not cut connexions of length r $(= 1)$ between pairs of vertices not containing x itself (this explains why the cases $r = 1$ and $r > 1$ are different). Therefore, G_x is one-twin-free. $\triangle$

The following theorem sharpens Corollary 7.

Theorem 8 *Let $n \geq 4$ be an integer, and $G = (V, E)$ be any connected one-twin-free graph with n vertices. Then there exists a vertex $x \in V$ such that G_x is one-twin-free and connected.*

Proof. In this proof, we use twin, twin-free and terminal for one-twin, one-twin-free and one-terminal, respectively. By Corollary 7, we know that G is not terminal. If Theorem 8 were not true, let $G = (V, E)$ be the smallest counter-example, that is, G satisfies:

(i) G is connected,
(ii) G is twin-free,
(iii) for all $x \in V$, G_x twin-free $\Rightarrow G_x$ not connected,
(iv) $n \geq 4$, and
(v) among all graphs satisfying (i)–(iv), $|V|$ is the smallest possible.

We show that such a graph cannot exist.

Let $x \in V$ be such that G_x is twin-free (such a vertex x exists because G is not terminal). By (iii), G_x consists of at least two connected components, F and H, see Figure 3.

If G is a star centered at x with at least four vertices, i.e., $G = (V, E)$ where $V = \{x, v_1, \ldots, v_{n-1}\}$, $E = \{\{x, v_i\} : 1 \leq i \leq n-1\}$, $4 \leq n$, then for any i between 1 and $n-1$, G_{v_i} is twin-free and connected, contradicting (iii). Threrefore we assume from now on that at least one connected component in G_x, say H, has at least two vertices.

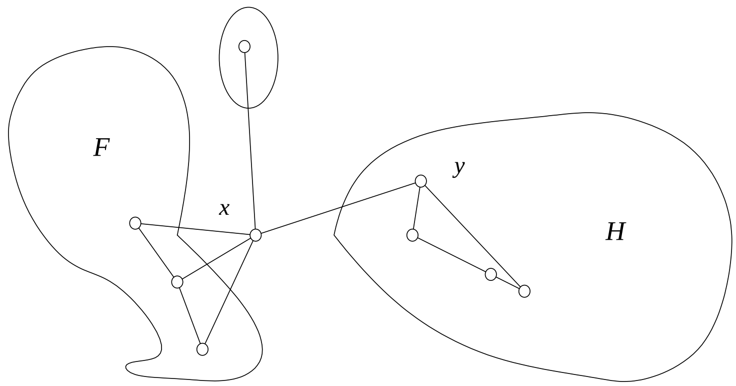

Figure 3: The vertex x and the connected components of G_x.

Step 1. We show that there are at least two edges between x and H. Assume on the contrary that there is only one, say $\{x, y\}$ (see Figure 3).

We construct the graph $G' = (V', E')$ by contracting the vertices x and y into one vertex xy: $V' = \big(V \setminus \{x, y\}\big) \cup \{xy\}$, $E' = \big(E \setminus \{\{x, y\}\} \setminus \{\{x, t\} : \{x, t\} \in E\} \setminus \{\{y, t\} : \{y, t\} \in E\}\big) \cup \{\{xy, t\} : \{x, t\} \in E \text{ or } \{y, t\} \in E\}$.

It is easy to see that G' is connected and twin-free, because G is. We now show that G' satisfies (iii). Let $z \in V'$ be such that G'_z is twin-free (such a z exists, because, by Corollary 7, G' is not terminal).

If $z = xy$, then G'_z is not connected.

If $z \neq xy$, G_z is twin-free, because G'_z is. Then, by (iii), G_z is not connected. This in turn implies that G'_z is not connected.

Therefore G' satisfies (i)–(iii) and has fewer vertices than G, a contradiction, unless $n = 4$. In this case however, we would have $G' = P_3$ and necessarily, since G is twin-free, $G = P_4$, but P_4 does not satisfy (iii).

This proves that there are at least two edges between x and H. This also shows that there are at least three vertices in H: if y and z were the only vertices in H and since they are connected, they would be twins (both in G and G_x).

Step 2. We still consider the connected component H, which by assumption is twin-free and has at least three vertices.

If H has exactly three vertices, then $H = P_3$ and it is easy to see from Figure 4 that, no matter how the vertices in H are linked to x in G, it is possible to find a vertex u in H such that G_u is twin-free and connected, again contradicting (iii).

If H has at least four vertices, then, since H has fewer vertices than G, H cannot satisfy simultaneously (i)–(iii). But H is connected and twin-free, by assumption. So H does not satisfy (iii): there is a vertex u in H such that H_u is connected and twin-free. It is not difficult now to see that G_u is connected and twin-free, again contradicting (iii). Indeed, G_u is connected because x is connected to a vertex other than u in H, as seen in Step 1; and G_u is twin-free, because H_u and G_x are twin-free and obviously x is not a twin in G_u.

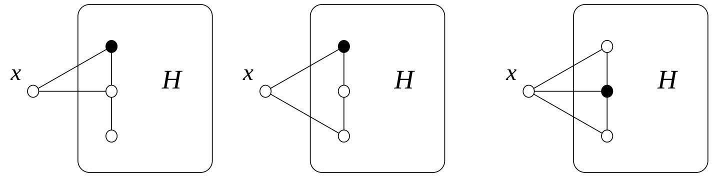

Figure 4: The case $H = P_3$. Vertices in black can be removed.

In all cases, we see that a graph G satisfying (i)–(iv) does not exist, which proves Theorem 8. △

3.2 Trees

We now consider the case of trees. We first give an easy lemma.

Lemma 9 *Let $r \geq 1$ and $n \geq 2r+2$ be integers, and $T = (V, E)$ be any (connected) r-twin-free tree with n vertices. If $\beta \in V$ has degree at least three, and if $C_1, C_2, \ldots, C_p$ are the connected components of T_β, then at least $p-1$ components C_i have, as an induced subgraph, a path with at least r vertices which has one end adjacent to β.*

Proof. Let $c_{i,1} \in C_i$ be at distance one from β. Since the p sets $B_r(c_{i,1})$ are different and nonempty, at least $p-1$ of the vertices $c_{i,1}$ have a vertex at distance at least $r-1$ in C_i. △

In other words, we have just proved that at least $p-1$ components contain (at least) one leaf f (that is, a vertex with degree one) with $d(f, \beta) \geq r$.

Theorem 10 *Let $r \geq 1$ and $n \geq 2r+2$ be integers, and $T = (V, E)$ be any (connected) r-twin-free tree with n vertices. Then there exists a leaf $x \in V$ such that T_x is r-twin-free (and connected).*

Proof. If T is a path, removing one of its ends gives a path with at least $2r+1$ vertices, which is still r-twin-free. So we assume that there is at least one vertex with degree at least three.

In the rest of this proof, we shall say that a vertex separates two vertices v_1 and v_2 if it covers exactly one of them, which means that v_1 and v_2 are not twins.

For any leaf $\alpha \in V$, let β_α be its closest vertex with degree at least three; among all leaves, we choose a leaf which has the smallest possible $d(\alpha, \beta_\alpha)$. We call this leaf x, we set $d = d(x, \beta_x)$ (so the distance between any leaf and any vertex with degree at least three is at least d), and we claim that T_x is r-twin-free. Since all the vertices that are r-covered by x are also r-covered by the (only) vertex at distance one from x, all we have to show is that, if x r-separates two vertices, v_1 and v_2, belonging to $V \setminus \{x\}$, then there is another vertex, z, that also separates them. For the same reason, without loss of generality, we can assume that $d(x, v_1) \leq r$ and $d(x, v_2)$ is *equal* to $r+1$ (otherwise,

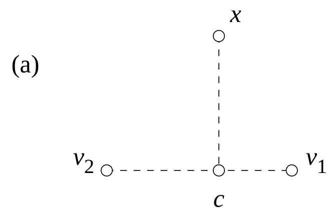

Figure 5: How to find a vertex x such that T_x is twin-free (1).

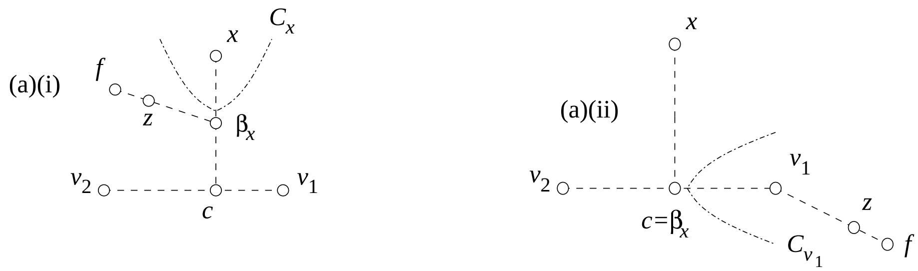

Figure 6: How to find a vertex x such that T_x is twin-free (2).

the vertex at distance one from x would separate v_1 and v_2). We can also assume that $d(v_1, v_2) \leq r$: otherwise, $v_1 \notin B_r(v_2)$ and v_1 and v_2 are not twins.

Consider the two paths between x and v_1 on the one hand, x and v_2 on the other hand. Two cases occur, depending on whether they both contain v_1 or not (see Figure 5).

Case (a): the two paths diverge before reaching v_1. Let c be the vertex where the two paths diverge; it is clear that β_x is between x and c. Two subcases occur: (i) $\beta_x \neq c$, (ii) $\beta_x = c$ (see Figure 6).

(i) $\beta_x \neq c$: from now on, for a vertex $y \in V \setminus \{\beta_x\}$, let C_y be the connected component containing y in T_{β_x}. By the definition of β_x and d, there is a connected component in T_{β_x}, different from C_x and C_c, which contains a leaf f with $d(f, \beta_x) \geq d$. This shows that on this path, the vertex z at distance d from β_x plays the same role as x with respect to v_1 and v_2: z r-covers v_1, not v_2.

(ii) $\beta_x = c$: then in C_{v_1}, there is at least one leaf, f, which is linked to β_x by a path going through v_1, and we know that $d(f, \beta_x) \geq d$. On this path, the vertex z at distance d from β_x plays a role similar to x: it r-covers v_1, not v_2.

Case (b): the path going to v_2 goes through v_1. There are four subcases: (i) β_x is between x and v_1, (ii) β_x is between v_1 and v_2 with $d(\beta_x, v_1) \neq d(\beta_x, v_2)$, (iii) β_x is between v_1 and v_2 with $d(\beta_x, v_1) = d(\beta_x, v_2)$, (iv) β_x is on the other side of v_2 (see Figure 7).

(i) β_x is between x and v_1 (and β_x can be equal to v_1):
this case is actually the same as Case (a)(i), with $v_1 = c$: let f be a leaf in a connected component in T_{β_x} which is neither C_x nor C_{v_2}. Again, $d(f, \beta_x) \geq d$, and between β_x and f, there is a vertex z at distance d from β_x, which plays the same role as x.

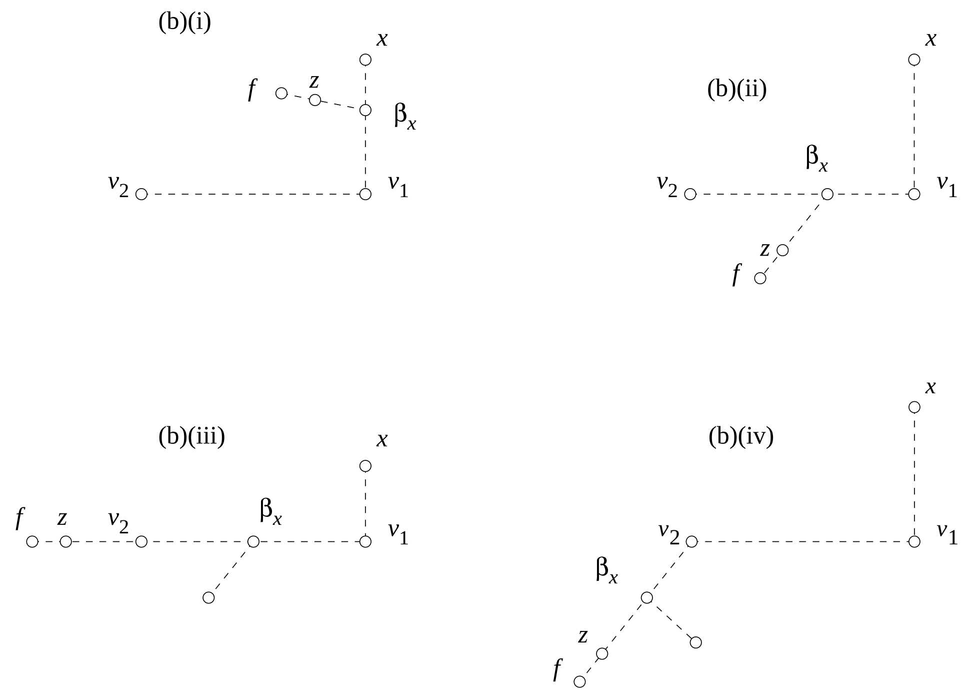

Figure 7: How to find a vertex x such that T_x is twin-free (3).

(ii) β_x is between v_1 and v_2, with $d(\beta_x, v_1) \neq d(\beta_x, v_2)$ (and β_x can be equal to v_2, not to v_1):
among the connected components in T_{β_x} which are not C_{v_1} (if $v_2 = \beta_x$), and are neither C_{v_1} nor C_{v_2} (if $v_2 \neq \beta_x$), let f be a leaf with the property that the distance between f and β_x is the largest possible. If $d < r$, then by Lemma 9, $d(f, \beta_x) \geq r$; if $d \geq r$, then $d(f, \beta_x) \geq r$, because $d(f, \beta_x) \geq d$. Then, since $d(v_1, v_2) \leq r$, it is easy to find, between f and β_x, a vertex z which r-covers v_1, not v_2 (respectively, v_2, not v_1), if $d(\beta_x, v_1) < d(\beta_x, v_2)$ (respectively, $d(\beta_x, v_1) > d(\beta_x, v_2)$), i.e., a vertex other than x which also separates v_1 and v_2.

(iii) β_x is between v_1 and v_2, with $d(\beta_x, v_1) = d(\beta_x, v_2) = \delta$:
let f be a leaf on the other side of v_2; again, $d(f, \beta_x) \geq d = \delta + d(x, v_1) > \delta$. Let z be the vertex between f and v_2 at distance d from β_x. Then $d(x, v_1) = d(z, v_2)$ and $d(x, v_2) = d(z, v_1)$, which means that z also separates v_1 and v_2.

(iv) β_x is on the other side of v_2 (and is not equal to v_2):
by Lemma 9, there is a leaf f in a connected component in T_{β_x} which is not equal to T_{v_2}, which is at distance at least r from β_x. Between f and v_2, there is a vertex z at distance r from v_2, which therefore separates v_2 and v_1.

In all cases, there is a vertex z, other than x, which separates v_1 and v_2. Therefore, G_x is twin-free. Moreover, since x is a leaf, G_x is connected. $\triangle$

We therefore have the following corollary, the first part of which is already contained in Corollary 7.

Corollary 11 *There is no one-terminal tree. For a given $r > 1$, the only r-terminal tree is the path P_{2r+1}.* $\triangle$

3.3 The case $r \geq 3$

We now consider general graphs, for $r \geq 3$.

Theorem 12 *For each integer $r \geq 3$, there is a graph G, $G \neq P_{2r+1}$, which is r-terminal.*

Proof. We search for a connected r-twin-free graph $G = (V, E)$, with $|V| \geq 2r+2$, $r \geq 3$, such that for all $x \in V$, G_x is not r-twin-free.

In this proof, calculations are carried modulo $2r$. Take a cycle of length $2r$ with vertices c_i ($i = 0, 1, \ldots, 2r-1$), and add one vertex s_i (which we shall call the *spike* of c_i) together with the edge $\{c_i, s_i\}$, for every value of i except one.

The resulting graph G is clearly r-twin-free. A cycle point c_i r-covers all the vertices except the spike s_{i+r} diagonally across — and of course there is one cycle point which r-covers all the vertices in the graph. In particular, each point in the cycle r-covers all the vertices in the cycle. Each spike s_i r-covers all the cycle points except c_{i+r} (the one diagonally across); so, indeed, the graph is r-twin-free.

It is also r-terminal. If we remove one spike, then there is another cycle point that r-covers all the vertices in the graph. If we consider one cycle point, c_i, then (since $r \geq 3$), both c_{i+1} and c_{i+2} — or both c_{i-1} and c_{i-2} — have spikes; say, s_{i+1}, s_{i+2} are in the graph G. If now we remove c_i, then c_{i+1} and s_{i+2} trivially r-cover the same vertices because $r \geq 3$. $\triangle$

Other constructions can be thought of. We give, without proof, one example which gives smaller graphs than the ones in the proof of Theorem 12. In this example, calculations are carried modulo $2r+1$. We consider a cycle of length $2r+1$ ($r \geq 3$) with vertices c_i ($i = 0, 1, \ldots, 2r$), and add the spike s_i for every value of i except

- $3k$ and $2r+1-3k$, for $0 \leq k \leq m$, if $r = 3m$ or $3m+1$,
- $2+3k$ and $2r+1-(2+3k)$, for $0 \leq k \leq m$, if $r = 3m+2$,

see Figure 8 for $r \in \{3, 4, 5, 6\}$.

3.4 Large terminal graphs and open problems

The above example as well as the construction of Theorem 12 do not work for $r = 2$. Other constructions have been tested and failed, and the problem remains open: apart from P_5, do two-terminal graphs exist?

When $r \geq 6$, the set $\mathcal{T}_r$ is infinite, as shows the following theorem.

Theorem 13 *For each integer $r \geq 6$, there are infinitely many r-terminal graphs.*

Proof. Assume first that $r \geq 7$.

For each $i = 1, 2, \ldots, m$ ($m \geq 3$) let $\mathcal{C}_i$ be a $2r$-cycle with vertices $c_i(j)$ (indices j modulo $2r$). Connect the cycles together by adding an edge from $c_i(r)$ to $c_{i+1}(0)$ for all

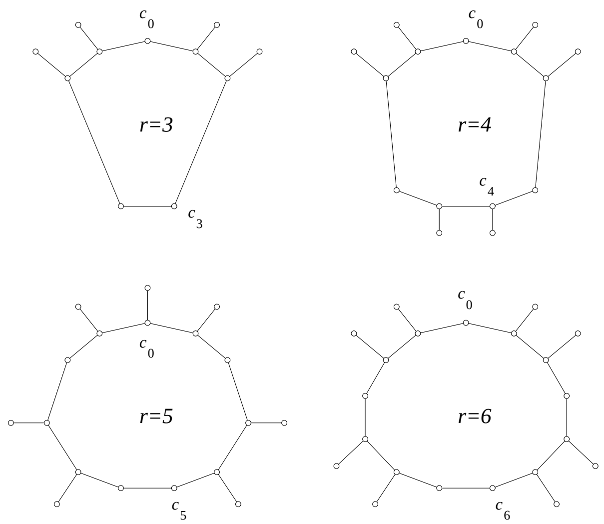

Figure 8: Different r-terminal graphs.

$i = 1, 2, \ldots, m-1$ and from $c_m(r)$ to $c_1(0)$; in fact, it is convenient to consider indices i modulo m.

For all i add spikes $s_i(j)$ to all the vertices $c_i(j)$ with $j = 1, -2, 3, -4, \ldots, (-1)^r(r-1)$. There are $m(r-1)$ such vertices, see Figure 9.

We claim that this graph G with $n = m(3r-1)$ vertices is r-terminal.

We first prove that G is r-twin-free. Assume that we know $B_r(v)$ for an unknown vertex v. We show that we can deduce v.

Clearly, $\mathcal{C}_i \subseteq B_r(v)$ if and only if $v \in \mathcal{C}_i$. Assume that this is the case. For $j = 0, 1, \ldots, r-1$, the vertices $c_i(\pm j)$ both r-cover $2r-1-2j$ vertices of $\mathcal{C}_{i-1}$, and $c_i(r)$ does not cover any vertices of $\mathcal{C}_{i-1}$, and, moreover, for $j = \pm 1, \pm 2, \ldots, \pm(r-1)$, exactly one of $c_i(j+r)$ or $c_i(-j-r)$ (the points diagonally opposite to $c_i(j)$ and $c_i(-j)$) has a spike attached to it and this spike r-separates $c_i(j)$ and $c_i(-j)$, so we can identify v.

If v does not belong to any $\mathcal{C}_i$, then it is one of the spikes. Given $B_r(v)$, we find out which is the cycle $\mathcal{C}_i$ of which $B_r(v)$ contains $2r-1$ vertices (there is only one). If the vertex of $\mathcal{C}_i$ which is missing from $B_r(v)$ is $c_i(j)$, then we know that v is the spike diagonally opposite, i.e., $v = s_i(j+r)$.

It suffices now to prove that G is r-terminal. If we delete one of the spikes, say $s_i(j)$, then by the construction $s_i(-j)$ is not a spike either, and $c_i(j+r)$ and $c_i(-j-r)$ are r-twins.

Assume finally that we delete a vertex $c_i(j) \in \mathcal{C}_i$ for some i and j.

If $j = 0$, then $c_i(-1)$ and $s_i(-2)$ are twins. The case $j = r$ is symmetrical.

Assume that $j \in S := \{\pm 1, \ldots, \pm(r-1)\}$. Because $r \geq 7$, we know that $j-1, j-2, j-3 \in S$ or $j+1, j+2, j+3 \in S$; say the latter. By the construction, $c_i(j+2)$ has a spike or $c_i(j+1)$ and $c_i(j+3)$ both have spikes. If $c_i(j+2)$ has a spike, then $c_i(j+1)$ and $s_i(j+2)$ are twins; if $c_i(j+1)$ and $c_i(j+3)$ both have spikes, then $c_i(j+2)$ and $s_i(j+3)$ are twins.

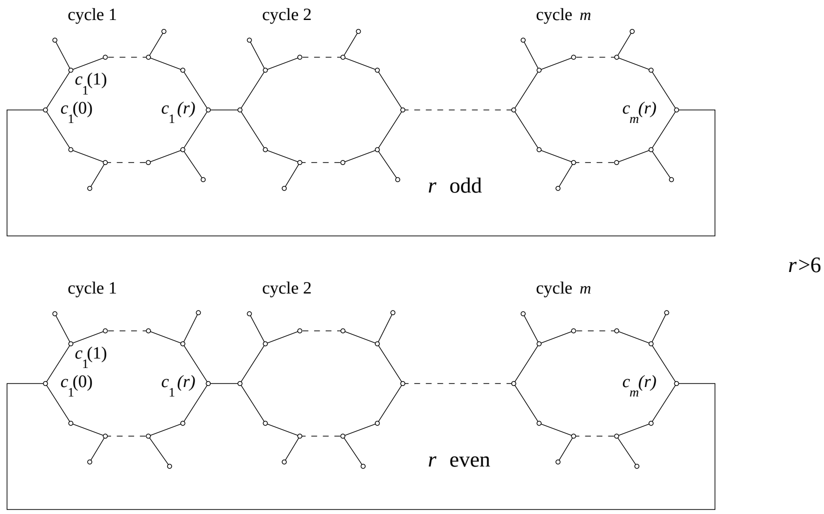

Figure 9: The r-terminal graph constructed in the proof of Theorem 13.

Assume finally that $r = 6$. We can use the same argument, if we add spikes to $c_i(j)$ with $j = 1, -2, -3, 4, -5$ (instead of what we did in the general case). $\triangle$

Therefore, another open problem is the situation for $r = 3, 4, 5$: there exist r-terminal graphs, but are they in finite or infinite number?

Observe that, whether they are in finite or infinite number, if we could prove that, for a given $r > 1$, all r-terminal graphs contain the path P_{2r+1} as an induced subgraph (which is the case for all r-terminal graphs described in Section 3.3), then Conjecture 3 would hold: simply consider a graph G which is not r-terminal, and delete vertices in G until you get a graph G' which is r-terminal; if G' contains P_{2r+1} as an induced subgraph, so does the initial graph G.

References

[1] N. Bertrand, Codes identifiants et codes localisateurs-dominateurs sur certains graphes, Mémoire de stage de maîtrise (under O. Hudry's supervision), ENST, Paris (France), June 2001.

[2] N. Bertrand, I. Charon, O. Hudry, A. Lobstein, Identifying and locating-dominating codes on chains and cycles, *Europ. J. Combin.*, **25/7** (2004), 969–987.

[3] N. Bertrand, I. Charon, O. Hudry, A. Lobstein, 1-identifying codes on trees, *Austral. J. Combin.*, **31** (2005), 21–35.

[4] U. Blass, I. Honkala, S. Litsyn, Bounds on identifying codes, *Discrete Math.* **241** (2001), 119–128.

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

[6] I. Charon, O. Hudry, A. Lobstein, Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard, *Theoret. Comput. Sci.* **290(3)** (2003), 2109–2120.

[7] I. Charon, O. Hudry, A. Lobstein, On the structure of identifiable graphs, *Electron. Notes Discrete Math.* **22** (2005), 491–495.

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

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

[10] I. Honkala, T. Laihonen, S. Ranto, On strongly identifying codes, *Discrete Math.* **254** (2002), 191–205.

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

[12] J. Moncel, Codes identifiants dans les graphes, Thèse de Doctorat de l'Université de Grenoble, France, 2005.

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

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