Characterizing extremal digraphs for identifying codes and extremal cases of Bondy's theorem on induced subsets[*]

Florent Foucaud   Reza Naserasr   Aline Parreau

February 15, 2012

Abstract

An identifying code of a (di)graph G is a dominating subset C of the vertices of G such that all distinct vertices of G have distinct (in)neighbourhoods within C. In this paper, we classify all finite digraphs which only admit their whole vertex set as an identifying code. We also classify all such infinite oriented graphs. Furthermore, by relating this concept to a well-known theorem of A. Bondy on set systems, we classify the extremal cases for this theorem.

1 Introduction

Identifying codes are dominating sets having the property that any two vertices of the graph have distinct neighbourhoods within the identifying code. As a consequence, they can be used to uniquely identify or locate the vertices of a graph. Identifying codes have been widely studied since the introduction of the concept in [11]. The theory is applied to problems such as fault-diagnosis in multiprocessor systems [11] or emergency sensor networks in facilities [14]. The concept of identifying codes is an extension of previous works on locating-dominating sets, studied in e.g. [8, 15]. Identifying codes have first been studied in undirected graphs, but the concept has naturally been extended to directed and oriented graphs [6, 7, 16].

In this paper, we study classes of directed graphs which only admit large minimum identifying codes, extending earlier works on such problems for undirected graphs (see e.g. [8, 9, 10]). We also relate the problem of identifying codes in directed graphs to a set theoretic problem first studied by A. Bondy in [4].

Let us now give some notations and definitions. In the following and unless otherwise stated, we will denote by D, a directed graph with vertex set $V(D)$ and arc set $A(D)$. A directed graph will conveniently be called a *digraph*. An arc pointing from vertex u towards vertex v will be denoted $\overrightarrow{uv}$. The arcs $\overrightarrow{uv}$ and $\overrightarrow{vu}$ are called *symmetric*. A digraph without any symmetric arcs will be called an *oriented graph*. For a vertex u of D, we denote by $N^-(u)$, the set of outgoing neighbours of u, and by $N^+(u)$, the set of incoming neighbours of u. By $d^+(u)$ and $d^-(u)$, respectively, we denote the in-degree and the out-degree of the vertex u. Let $B_1^+(u)$ and $B_1^-(u)$ denote the incoming and outgoing balls of radius 1 centered at u (they include u). A vertex of D without any incoming arc is called a *source* in D. The *underlying graph* of D is the graph obtained from D by ignoring the directions of the arcs. We say that D is *connected* if its underlying graph is connected.

In a digraph D, a *dominating set* is a subset C of $V(D)$ such that for all $u \in V(D)$, $B_1^+(u) \cap C \neq \emptyset$. A *separating code* of D is a subset C of vertices such that each pair of vertices u, v of D is *separated*, that is: $B_1^+(u) \cap C \neq B_1^+(v) \cap C$. If C is both a dominating set and a separating code, it is called an *identifying code* of D.

Two vertices u and v are called *twins* in a digraph D if $B_1^+(u) = B_1^+(v)$. It is easy to observe that a digraph admits a separating code and, therefore, an identifying code, if and only if it has no pair of twins ($C = V(D)$ is one such code in this case). Such digraphs are called *twin-free*. Note that if two vertices u and v are twins, necessarily the symmetric arcs $\overrightarrow{uv}$ and $\overrightarrow{vu}$ must exist. As a consequence, oriented graphs are always twin-free.

In this paper, we will also consider separating codes in undirected bipartite graphs. Given a bipartite graph $G = (S \cup T, E)$ with S and T being two parts, an *S-separating code* of G is a subset C of vertices in T such that $N(u) \cap C \neq N(v) \cap C$ for all distinct vertices u and v of S. If moreover, every vertex of S is also dominated by some vertex of C, then C is called an *S-discriminating code* of G. When G admits an S-discriminating code then we say that G is *S-identifiable*. The concept of S-separating codes has also been studied under the name of *test collection problem*, see e.g. [13]. The concept of S-discriminating codes was introduced in [5]. We note that to find an identifying code, respectively a separating code, of a graph or a digraph one could form its incidence bipartite graph and search for an S-discriminating or an S-separating code in this incidence bipartite graph.

One of the most natural questions in the study of separating or identifying codes is to find the smallest such set for a given graph. We denote by $\overrightarrow{\gamma^S}(D)$ and by $\overrightarrow{\gamma^{ID}}(D)$, the minimum sizes of a separating code and of an identifying code of D, respectively. It is observed that $\overrightarrow{\gamma^S}(D) \leq \overrightarrow{\gamma^{ID}}(D) \leq \overrightarrow{\gamma^S}(D) + 1$. Determining the value of $\overrightarrow{\gamma^{ID}}(D)$ was proven to be NP-hard in [7], even for strongly connected bipartite oriented graphs and for bipartite oriented graphs without directed cycles.

The theory of identifying codes is studied mostly for undirected graphs, which can be viewed as symmetric digraphs. The following is one of the first general bounds:

Theorem 1 ([2, 10]). *For any symmetric finite twin-free digraph D having at least one arc, we have $\overrightarrow{\gamma^{ID}}(D) \leq |V(D)| - 1$.*

Undirected graphs achieving this bound have been characterized in [9]. For the directed analogue of this classification problem, the class of digraphs satisfying $\overrightarrow{\gamma^{ID}}(D) = |V(D)|$ is already rich. In this paper we classify all such finite digraphs. This is done in Section 2. A simple corollary of our work is an extension of Theorem 1 which says that every digraph with at least one symmetric arc satisfies: $\overrightarrow{\gamma^{ID}}(D) \leq |V(D)| - 1$. In Section 3 we classify all infinite oriented graphs in which the whole vertex set is the only identifying code. In Section 4, we discuss the relationship between this result and the following theorem of A. Bondy on "induced subsets":

Theorem 2 (Bondy's theorem [4]). *Let $\mathcal{A} = \{A_1, A_2, \cdots, A_n\}$ be a collection of n distinct subsets of an n-set X. Then there exists an element x of X such that the sets $A_1 - x, A_2 - x, \cdots, A_n - x$ are all distinct.*

Here $A_i - x$ could be the empty set. Though, to be distinct, at most one of them could be empty. In Section 4 we classify all the set systems in which for any possible choice of x in Bondy's theorem we will have $A_i - x = \emptyset$ for some i.

Since its publication in 1972, this theorem has received a lot of attention and various proofs of different nature have been provided for it, we refer to [1, 3, 4, 12, 17] for some interesting proofs.

2 Finite extremal digraphs

In this section, we classify all finite digraphs in which the whole vertex set is the only identifying code. We begin with the following definitions of operations on digraphs, illustrated in Figure 1.

Given two digraphs D_1 and D_2 on disjoint sets of vertices, we define the *disjoint union* of D_1 and D_2, denoted $D_1 \oplus D_2$, to be the digraph whose vertex set is $V(D_1) \cup V(D_2)$ and whose arc set is $A(D_1) \cup A(D_2)$.

Given a digraph D and a vertex x not in $V(D)$ we define $x \overrightarrow{\lhd}(D)$ to be the digraph with vertex set $V(D) \cup \{x\}$ whose arcs are the arcs of D together with each arc $\overrightarrow{xv}$ for every $v \in V(D)$.

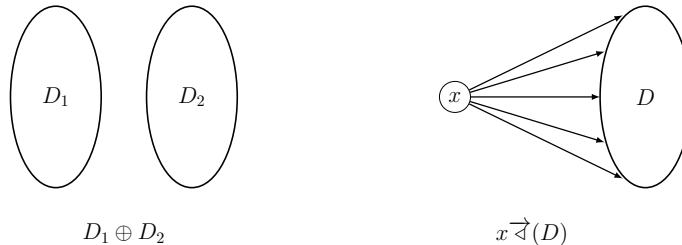$$D_1 \oplus D_2 \qquad\qquad\qquad x\overrightarrow{\lhd}(D)$$

Figure 1: Illustrations of the two operations

Definition 3. *Let $(K_1, \oplus, \overrightarrow{\lhd})$ be the closure of the one-vertex graph K_1 with respect to the operations $\oplus$ and $\overrightarrow{\lhd}$. That is, the class of all graphs that can be built from K_1 by repeated applications of $\oplus$ and $\overrightarrow{\lhd}$.*

The *transitive closure* of a digraph D is a digraph obtained from D by adding the arc $\overrightarrow{xy}$ whenever there is a directed path from x to y. A *rooted oriented tree* is an oriented tree with a specific vertex v (called the root) such that for every other vertex u the path connecting u to v is a directed path from v to u. In such a tree, given an arc $\overrightarrow{xy}$, we say that x is the father of y and that y is a child of x. The vertices with a directed path to x are called *ancestors* of x and vertices with a directed path from x are *descendents* of x. A *rooted oriented forest* is a disjoint union of rooted oriented trees.

By these definitions it is easy to check that:

Observation 4. *Every element of $(K_1, \oplus, \overrightarrow{\lhd})$ is the transitive closure of a rooted oriented forest.*

For a digraph D in $(K_1, \oplus, \overrightarrow{\lhd})$ let us denote by $\overrightarrow{F}(D)$ the rooted oriented forest whose transitive closure is D.

Proposition 5. *For every digraph D in $(K_1, \oplus, \overrightarrow{\lhd})$ we have $\overrightarrow{\gamma^{ID}}(D) = |V(D)|$. Furthermore, if a vertex x is a source in D, then $V(D) - x$ is a separating code of D. Otherwise, the vertex x and its father in $\overrightarrow{F}(D)$ are the only ones not being separated from each other by the set $V(D) - x$.*

Proof. Let C be an identifying code of D. Except for its roots, each vertex of the forest $\overrightarrow{F}(D)$ must be in C in order to be separated from its father. But the sources need also to be in C in order to be dominated. □ □

The next proposition is the directed analogue of Proposition 3 in [9], and since the proof is quite the same we omit it.

Proposition 6. *Let D be a finite twin-free digraph, and let S be a subset of vertices of D such that $D - S$ is twin-free. Then $\overrightarrow{\gamma^{ID}}(D) \leq \overrightarrow{\gamma^{ID}}(D - S) + |S|$.*

Let D be a twin-free digraph with vertex set $\{x_1, x_2, \cdots, x_n\}$ and let $\mathcal{A} = \{B_1^+(x_1), B_1^+(x_2), \cdots, B_1^+(x_n)\}$. Then $(\mathcal{A}, V(D))$ form a set system satisfying the conditions of Bondy's theorem. Therefore we have:

Proposition 7. *Let D be a finite twin-free digraph. Then $\overrightarrow{\gamma^{s}}(D) \leq |V(D)| - 1$.*

The following corollary of the previous proposition will also be needed.

Proposition 8. *In a finite twin-free digraph D, $\overrightarrow{\gamma^{ID}}(D) = |V(D)|$ if and only if $\overrightarrow{\gamma^{s}}(D) = |V(D)| - 1$ and for every minimum separating code of D, there is a vertex which is not dominated.*

Proof. Suppose $\overrightarrow{\gamma^{ID}}(D) = |V(D)|$. By Proposition 7 we know that $\overrightarrow{\gamma^{s}}(D) \leq |V(D)| - 1$. On the other hand, $\overrightarrow{\gamma^{ID}}(D) \leq \overrightarrow{\gamma^{s}}(D) + 1$ thus $\overrightarrow{\gamma^{s}}(D) = |V(D)| - 1$. Now, if all the vertices where dominated by some minimum separating code of D, that code would also be identifying, a contradiction.

Conversely, if $\overrightarrow{\gamma^{s}}(D) = |V(D)| - 1$ and for any minimum separating code of D, there is a vertex which is not dominated, we are forced to take all the vertices in any identifying code in order to get the domination property. $\qquad\square\qquad\qquad\qquad\square$

The following theorem shows that the family of Definition 3 is exactly the class of finite digraphs in which the whole vertex set is the only identifying code.

Theorem 9. *Let D be a finite twin-free digraph. If $\overrightarrow{\gamma^{ID}}(D) = |V(D)|$ then $D \in (K_1, \oplus, \overrightarrow{\lhd})$.*

Proof. Assume D is the smallest digraph for which $\overrightarrow{\gamma^{ID}}(D) = |V(D)|$ but $D \notin (K_1, \oplus, \overrightarrow{\lhd})$. We consider two cases:

Case a. Assume that there exists a vertex x of D such that x has no outneighbours. Then $D - x$ is a twin-free graph. By Proposition 6 we have $\overrightarrow{\gamma^{ID}}(D - x) = |V(D - x)|$ and, therefore, by the minimality of D, $D - x \in (K_1, \oplus, \overrightarrow{\lhd})$. Thus $\overrightarrow{F}(D - x)$ is well-defined. Since $V(D) - x$ is not an identifying code of D, in $V(D) - x$ either there is a vertex y which is not separated from x, or x is not dominated.

If x is not dominated, then x is an isolated vertex and, therefore, D is the disjoint union of two members of $(K_1, \oplus, \overrightarrow{\lhd})$, hence D is also in the family. So there is a vertex y which is not separated from x. Therefore, $N^+(x) = N^+(y) \cup \{y\}$. Then D is the transitive closure of the oriented tree built from $\overrightarrow{F}(D - x)$ by adding x as a child of y, a contradiction.

Case b. Every vertex has at least one outneighbour. By Proposition 7 we know that there exists a vertex x such that $D - x$ is twin-free and as in the previous case $\overrightarrow{F}(D - x)$ is well-defined.

Since every vertex, in particular, every leaf t of $\overrightarrow{F}(D - x)$ has an outneighbour in D, we must have $\overrightarrow{tx} \in A(D)$. Thus $d^+(x) \geq 1$ and, therefore, $V(D) - x$ is a dominating set. But since it is not an identifying code there is a vertex $y \neq x$ which is not separated from x by $V(D) - x$, i.e., $N^+(x) = N^+(y) \cup \{y\}$.

We now claim that y is the only leaf of $\overrightarrow{F}(D - x)$. That is because if $t \neq y$ is a leaf then $t \in N^+(x)$ so $t \in N^+(y)$ which is a contradiction. But there has to be at least one leaf in $\overrightarrow{F}(D - x)$ thus y is the only leaf in $\overrightarrow{F}(D - x)$ and, therefore, $\overrightarrow{F}(D - x)$ is a path. Now since $d^-(x) > 0$, there is a vertex in $N^-(x)$. We have $y \notin N^-(x)$ since otherwise we would have $N^+(x) = N^+(y)$.

First, assume that there exists a vertex $t \in N^-(x)$ such that the father of t in $\overrightarrow{F}(D - x)$ is not in $N^-(x)$. We claim that $C = V(D) - t$ is an identifying code. Indeed, x is the only vertex

4

dominated by all vertices of C. Vertex t and its father are separated by x. Finally, each other pair of vertices from $V(D) - x$ is separated by the one which is a descendant of the other in $\overrightarrow{F}(D - x)$.

Now, assume that there is no vertex in $N^-(x)$ with its father in $\overrightarrow{F}(D - x)$ not belonging to $N^-(x)$. Let r be the root of $\overrightarrow{F}(D - x)$. In particular, $r \in N^-(x)$. We claim that $C = V(D) - r$ is an identifying code. Indeed, x is the only vertex dominated by all vertices of C. Each pair of vertices from $V(D) - x$ is separated by the one which is a descendant of the other in $\overrightarrow{F}(D - x)$. Finally, r is the only vertex which is dominated only by x. $\qquad\square\qquad\qquad\square$

Noticing that digraphs in $(K_1, \oplus, \overrightarrow{\lhd})$ have no symmetric arcs, we have the following extension of Theorem 1:

Corollary 10. *Let D be a finite twin-free digraph having some symmetric arcs. Then $\overrightarrow{\gamma^{ID}}(D) \leq |V(D)| - 1$.*

3 Extremal infinite oriented graphs

In this section, we consider the case of infinite digraphs in which the whole vertex set is the only identifying code. To avoid set theoretic problems we only consider infinite graphs on a countable set of vertices. Thus in this section the considered directed or oriented graphs have a finite or countable set of vertices.

By the characterization of the family of simple infinite graphs needing their whole vertex set to be identifed in [9], we already have a rich family of extremal symmetric digraphs with respect to identifying codes. The family of all such directed graphs seems to be too rich to characterize. In this section we provide such characterization for the class of all oriented graphs. We begin with the following definitions:

A connected (and possibly infinite) oriented graph D is a *finite-source transitive tree* (fst-tree for short) if:

(1) for each vertex x of D, $B_1^+(x)$ induces the transitive closure of a finite directed path, P_x, with x as its end vertex, and

(2) for each pair x, y of vertices, there is a vertex $z \in V(D)$ such that $P_x \cap P_y = P_z$.

Note that in an fst-tree D, since each path P_x is finite, point (2) of the definiton implies that for any pair x, y of vertices, P_x and P_y begin with the same vertex. Hence all paths P_x begin with the same vertex, which is the unique source of D.

An oriented graph D is an *infinite-source transitive tree* (ist-tree for short) if:

(1) for all vertices x of D, $B_1^+(x)$ induces the transitive closure of an infinite path (we denote it by P_x), and

(2) for any pair x, y of vertices of D, there is a vertex $z \in V(D)$ such that $P_x \cap P_y = P_z$.

Note that an ist-tree has no source vertex but one can imagine infinity as its source.

Finally we say that an oriented graph D is a *source transitive tree* if it is either an fst-tree or an ist-tree, see Figure 2 for examples. For each pair x, y of vertices of a source transitive tree D, paths P_x and P_y share the vertices of a path P_z which includes the "beginning" of both P_x and P_y. Hence, all arcs of D can be oriented in the same direction. Moreover, there cannot be any cycle in the union of all paths P_x. This implies that each source transitive tree D is the transitive

closure of a finite or infinite "rooted" oriented tree (even if an *ist*-tree has no properly defined root vertex, one can regard infinity as its root) which we call the *underlying tree* of D. Notice that the collection of fst-trees on a finite set of vertices is exactly the set of connected elements of $(K_1, \oplus, \overrightarrow{\lhd})$.
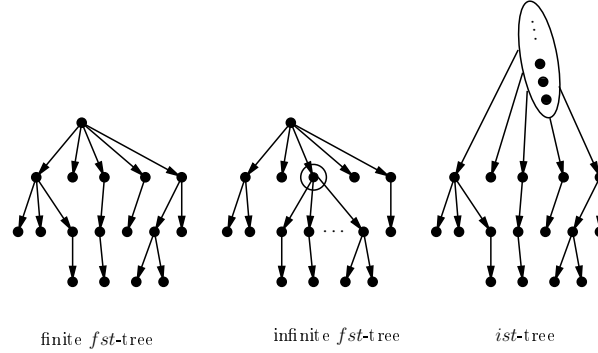

Figure 2: Underlying trees of source transitive trees

Proposition 11. *The only identifying code of a source transitive tree is its whole set of vertices.*

Proof. Let $D = (V, A)$ be a source transitive tree and x be any vertex of D. If x is a source (it can only happen if D is an fst-tree), then x must be in any identifying code of D in order to be dominated. If x is not a source, then to separate x from its father in the underlying tree of D, x itself must be in any identifying code. □ □

We are now ready to build the whole family of oriented graphs that need their whole vertex set to be identified. To this end given any oriented graph H we first build the family $\Psi(H)$ of extremal oriented graphs as follows:

For each vertex x of H if x is a source of H, then we assign an fst-tree T_x to x. If x is not a source, then we assign an ist-tree T_x to x. The choice of T_x is free but each T_x has its distinct set of vertices. For each arc $\overrightarrow{xy}$ of H we also associate a subset $V_{\overrightarrow{xy}}$ of $V(T_x)$ (the choice of $V_{\overrightarrow{xy}}$ is also free). We now build a member of $\Psi(H)$ by taking $\cup V(T_x)$ as the vertex set, arcs of T_x are also arcs of the new graph and, furthermore, for any $z \in V_{\overrightarrow{xy}}$ and any $t \in V(T_y)$, we add an arc $\overrightarrow{zt}$.

Proposition 12. *Given an oriented graph H, any digraph D in $\Psi(H)$ can only be identified by its whole vertex set.*

Proof. The sources of D are exactly the sources of the fst-trees T_x for source-vertices x of H and need to be in any identifying code in order to be dominated. If a vertex u of D is not a source, then it is in an fst or ist-tree T_x and there is, like in the proof of Proposition 11, a vertex v of T_x such that $B_1^+(u) \cap V(T_x) = (B_1^+(v) \cap V(T_x)) \cup \{u\}$ (v simply is the father of u in the underlying tree of T_x). By our construction, any incoming neighbour of u not in T_x is also an incoming neighbour of v so $B_1^+(u) = B_1^+(v) \cup \{u\}$ and u must be in any identifying code of D. □ □

Theorem 13. *Let D be an infinite twin-free oriented graph. Then a proper subset of $V(D)$ identifies all pairs of vertices of D unless $G \in \Psi(H)$ for some finite or infinite oriented graph H.*

6

Proof. Let D be an infinite oriented graph that needs its whole vertex set to be identified. Let x be a vertex of D. The set $V(D) - \{x\}$ is not an identifying code. Either x is not dominated and so, x is a source or there is a pair of vertices, say u and v, such that $B_1^+(v) = B_1^+(u) \cup \{x\}$. If $x \neq v$, we must have $\overrightarrow{uv} \in A$ and $\overrightarrow{vu} \in A$. Since D has no symmetric arc, this is not possible so, necessarily, $x = v$, $B_1^+(x) = B_1^+(u) \cup \{x\}$ and u is the only vertex such that $B_1^+(x) = B_1^+(u) \cup \{x\}$. So for any vertex x of D which is not a source there is a unique vertex we call x_{-1} such that $B_1^+(x) = B_1^+(x_{-1}) \cup \{x\}$. We may repeat this argument on x_{-i} to get x_{-i-1} for $i = 1, 2, \cdots$ as long as x_{-i} is not a source. This will result in a well defined set $\{\cdots, x_{-i}, \cdots, x_0 = x\}$ which induces a transitive closure of a finite or infinite path, which we denote by P_x (if x is a source itself, then $P_x = \{x\}$).

Assume that for two vertices x and y, $P_x \cap P_y \neq \emptyset$. Then let x_i be the first (in the order defined by the path P_x) vertex of P_x in $P_x \cap P_y$. We have $P_x \cap P_y = P_{x_i}$.

We now define an equivalence relation on the vertices of D: $x \equiv y$ if and only if $P_x \cap P_y \neq \emptyset$. This gives us the equivalence class of x: $T_x = \{y \in V(D) | x \equiv y\}$. The set of vertices of T_x induces either an fst-tree or an ist-tree in D. Furthermore, if $u \notin T_x$, then $\overrightarrow{uv}$ is an arc of D for either every $v \in V(T_x)$ or no $v \in V(T_x)$. In fact, if there is an arc $\overrightarrow{uy}$ in D for $y \in T_x$, then if $z \in T_x$ we have $P_y \cap P_z = P_{y_i}$. But $B_1^+(y) = B_1^+(t) \cup \{y, y_{-1}, ..., y_{i+1}\}$ so $u \in B_1^+(t)$. Now, $B_1^+(t) \subset B_1^+(z)$ so $u \in B_1^+(z)$ and $\overrightarrow{uz}$ is also an arc of D.

We construct a graph H as follows: the vertices of H are the equivalence classes T_x and there is an arc $\overrightarrow{T_x T_y}$ if there is an arc $\overrightarrow{uv}$ of D such that $u \in V(T_x)$ and $v \in V(T_y)$. It is now clear that $D \in \Psi(H)$. $\qquad\square \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We conclude this section by the following remarks:

- This proof also works for the classification of finite oriented graphs with $\overrightarrow{\gamma^{\text{ID}}}(D) = |V(D)|$. But the characterization of Theorem 9 is for all digraphs, thus it is a stronger statement.

- The oriented graphs for which the only separating set is their whole vertex set are graphs in $\Psi(H)$ as long as H has no source vertex.

4 An application to Bondy's Theorem

In this section, unless specifically mentioned, a set system is a pair $(\mathcal{A}, X)$ with X being any set of size n and $\mathcal{A}$ being a collection of n distinct subsets of X. When applying Bondy's theorem to a set system $(\mathcal{A}, X)$ where all subsets in $\mathcal{A}$ are distinct *and nonempty*, it is a natural request to be able to choose an element x of X such that no subset $A_i - x$ is the empty set. This is not always possible, just consider the set system $\mathcal{A}$ consisting of all singletons of X. Such set systems will be called *extremal*. More precisely, an extremal set system is a set system $(\mathcal{A}, X)$ in which elements of $\mathcal{A}$ are all distinct and nonempty, and where for any element x of X either there is an element $A_i \in \mathcal{A}$ with $A_i - x = \emptyset$ or there is a pair $A_i, A_j \in \mathcal{A}$ such that $A_i - x = A_j - x$. In this section we characterize all such extremal cases.

We would like to mention that almost any proof of Bondy's theorem (e.g., see [4, 3, 1, 12]) works for an extension of this theorem in which we are allowed to have more elements in X than $\mathcal{A}$. We then look for a subset $X' \subset X$ of size $|X| - |\mathcal{A}| + 1$ such that all the induced sets $A_i - X'$ are distinct. The following proposition is now an easy consequence of this general version of Bondy's theorem.

Proposition 14. *Let $(\mathcal{A}, X)$ be a set system with $|X| > |\mathcal{A}|$ where all the subsets in $\mathcal{A}$ are distinct and nonempty. There is a subset X' of X of size $|X| - |\mathcal{A}|$ such that all the subsets $A \cap (X - X')$ for $A \in \mathcal{A}$ are nonempty and distinct.*

Proof. Let X_0 be the subset of size $|X| - |\mathcal{A}| + 1$ found by extended version of Bondy's theorem i.e., all the subsets $A \cap (X - X_0)$ are distinct. Thus there is at most one subset A_0 such that $A_0 \cap (X - X_0)$ is empty. Let $x_0 \in X_0$ be an element of A_0. Then $X' = X_0 - \{x_0\}$ satisfies the proposition. □ □

To achieve our goal of characterizing the extremal set systems, we will need a few more definitions. Given a set system $(\mathcal{A}, X)$ we define its *incidence bipartite graph* $B(\mathcal{A}, X)$ to be the bipartite graph with $\mathcal{A} \cup X$ as its vertex set where $\mathcal{A}$ and X form the two parts and in which vertex $A_i \in \mathcal{A}$ is adjacent to vertex $x_j \in X$ if and only if $x_j \in A_i$. Bondy's theorem can now be restated as follows:

Theorem 15. *Let $G = (S \cup T, E)$ be an S-identifiable bipartite graph with $|S| = |T|$. Then there exists an S-separating code $C \subseteq T$ of size at most $|S| - 1$.*

As in the case of set systems one can also define the *incidence bipartite graph* of a digraph. To this end given a digraph D on a vertex set $\{x_1, x_2, \cdots, x_n\}$ we define $B(D)$ to be the bipartite graph on $S = \{x_1, x_2, \cdots, x_n\}$ and $T = \{x'_1, x'_2, \cdots, x'_n\}$ with x_i being adjacent to x'_j if either $\overrightarrow{x_i x_j} \in A(D)$ or $i = j$. The latter condition of adjacency implies that for a bipartite graph to be the incidence bipartite graph of some digraph it must admit a perfect matching. Below we prove that this is also a sufficient condition.

Lemma 16. *A bipartite graph $G = (S \cup T, E)$ is the incidence bipartite graph of some digraph D if and only if $|S| = |T|$ and G admits a perfect matching $\varphi : S \to T$.*

Proof. It follows from the definition of the incidence bipartite graph of a digraph that G must have both parts of the same size and admits a perfect matching which matches the two copies of each vertex.

Now, given a bipartite graph G with parts S and T of equal size together with a perfect matching φ, one can construct a digraph D with vertex set S in the following way. For each pair x, y of vertices if $x\varphi(y) \in E(G)$, then $\overrightarrow{xy}$ is an arc of D. The constructed digraph has G as its incidence bipartite graph. □ □

An example of the correspondence between a bipartite graph with a perfect matching (thick edges) and a digraph is given in Figure 3.

We are now ready to achieve our goal of classifying the extremal set systems in Bondy's theorem when $|\mathcal{A}| = |X|$.

Theorem 17. *A set system $(\mathcal{A}, X)$ with $|\mathcal{A}| = |X|$ is extremal if and only if its incidence bipartite graph $B(\mathcal{A}, X)$ is the incidence bipartite graph of a digraph in $(K_1, \oplus, \overrightarrow{\vartriangleleft})$.*

Proof. If $B(\mathcal{A}, X)$ is the incidence bipartite graph of a member D of $(K_1, \oplus, \overrightarrow{\vartriangleleft})$, then by Theorem 9 we have $\overrightarrow{\gamma^{\mathrm{iD}}}(D) = |V(D)|$ thus by Proposition 8 with any separating code of size $|V(D)| - 1$ there must be a vertex which is not dominated. Any separating code $V(D) - x$ of D corresponds to the choice of x in Bondy's theorem, and leaves some vertex in D undominated — that is, there exists $A_i \in \mathcal{A}$ such that $A_i - x = \emptyset$. Hence $(\mathcal{A}, X)$ is extremal.

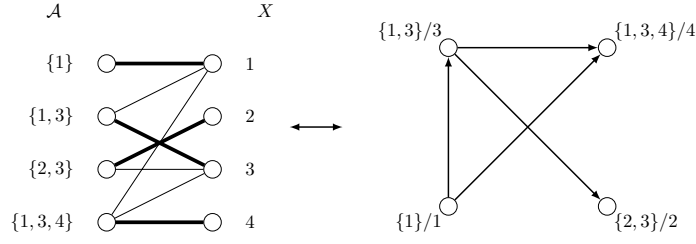For the other direction, we distinguish two cases.

8

Figure 3: Digraph of a bipartite graph

Case a. $B(\mathcal{A}, X)$ admits a perfect matching. Then the directed graph D built from $B(\mathcal{A}, X)$ using this matching as in the method of Proposition 8 is such that $\overrightarrow{\gamma^{\text{ID}}}(D) = |V(D)|$. Thus by Theorem 9, $D \in (K_1, \oplus, \overrightarrow{\lhd})$ and we are done.

Case b. $B = B(\mathcal{A}, X)$ does not contain a perfect matching. Then by Hall's marriage theorem (see e.g. [3]), there is a subset X' of X such that $|N_B(X')| < |X'|$. For $A_i \in \mathcal{A}$ we define $A'_i = A_i \cap X'$ and $\mathcal{A}' = \{A'_1 \cdots A'_{|\mathcal{A}|}\} - \emptyset$. Consider the set system $(\mathcal{A}', X')$ ($|\mathcal{A}'| < |X'|$). By Proposition 14 there is an element x_0 in X' such that $A'_i - x_0$ are all nonempty and distinct as long as they induce distinct elements in $\mathcal{A}'$. Now it is easy to check that $X - x_0$ induces nonempty and distinct elements on $\mathcal{A}$. $\qquad\square \qquad\qquad\qquad\qquad\qquad \square$

Using the previous theorem we can describe the extremal set systems $(\mathcal{A}, X)$ with $|\mathcal{A}| = |X|$ purely in the terminology of sets:

Corollary 18. *A set system $(\mathcal{A}, X)$ with $|\mathcal{A}| = |X|$ is extremal if and only if:*

- $\bigcup A_i = X$

- *for any subset A_i with at least two elements, there is an element x of A_i such that $A_i - x \in \mathcal{A}$*

Proof. If D is a digraph of $(K_1, \oplus, \overrightarrow{\lhd})$, it is easy to see that the set system corresponding to its incidence bipartite graph $B(D)$ has the properties.

Now, let $(\mathcal{A}, X)$ be a set system having the properties of the corollary. Assume there is an element x such that all $A_i - x$ are distinct and nonempty. Take A_i to be the smallest subset containing x, it exists because $\bigcup A_i = X$. Then $A_i \neq \{x\}$ and so there is an element y such that $A_i - y = A_j$. Then necessarily $x \neq y$ and A_j is a smaller set than A_i containing x. This is a contradiction. $\qquad\square \qquad\qquad\qquad\qquad\qquad \square$

We conclude this section by relating this work to a similar problem studied by Charon et al. in [5]. We have previously seen that Bondy's theorem can be stated in the language of separating codes in bipartite graphs. Furthermore, the extremal case we have studied, where a minimum separating code does always give an undominated vertex, is equivalent to the one where the only minimum discriminating code consists in the whole vertex set (in fact Proposition 8 holds also for separating and discriminating codes in bipartite graphs). Therefore Theorem 17 can be stated in the language of discriminating codes:

Corollary 19. *Given a bipartite graph $G = (S \cup T, E)$ with $|S| = |T|$ which is S-identifable, a minimum discriminating code $C \subseteq T$ of G has size $|S|$ if and only if G is the incidence bipartite graph $B(D)$ of some digraph D in $(K_1, \oplus, \overrightarrow{\lhd})$.*

References

[1] L. Babai and P. Frankl. *Linear algebra methods in combinatorics, preliminary version 2.* Department of Computer Science, The University of Chicago, 1992.

[2] N. Bertrand. Codes identifiants et codes localisateurs-dominateurs sur certains graphes, Master thesis, ENST, Paris, France, June 2001.

[3] B. Bollobás. *Combinatorics: Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability.* Cambridge University Press, 1986.

[4] J. A. Bondy. Induced subsets. *Journal of Combinatorial Theory Series B* 12(2):201–202, 1972.

[5] I. Charon, G. Cohen, O. Hudry and A. Lobstein. Discriminating codes in bipartite graphs: bounds, extremal cardinalities, complexity. *Advances in Mathematics of Communications* 4(2):403–420, 2008.

[6] I. Charon, S. Gravier, O. Hudry, A. Lobstein, M. Mollard and J. Moncel. A linear algorithm for minimum 1-identifying codes in oriented trees. *Discrete Applied Mathematics* 154(8):1246–1253, 2006.

[7] I. Charon, O. Hudry and A. Lobstein. Identifying and locating-dominating codes: NP-completeness results for directed graphs. *IEEE Transactions on Information Theory* 48(8):2192–2200, 2002.

[8] I. Charon, O. Hudry and A. Lobstein. Extremal cardinalities for identifying and locating-dominating codes in graphs. *Discrete Mathematics* 307(3-5):356–366, 2007.

[9] F. Foucaud, E. Guerrini, M. Kovše, R. Naserasr, A. Parreau and P. Valicov. Extremal graphs for the identifying code problem. *European Journal of Combinatorics* 32(4):628–638, 2011.

[10] S. Gravier and J. Moncel. On graphs having a $V \setminus \{x\}$ set as an identifying code. *Discrete Mathematics* 307(3-5):432–434, 2007.

[11] M. G. Karpovsky, K. Chakrabarty and L. B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Transactions on Information Theory* 44:599–611, 1998.

[12] L. Lovász, *Combinatorial problems and exercises.* North-Holland, Amsterdam, 1979.

[13] B. M. E. Moret and H. D. Shapiro. On minimizing a set of tests. *SIAM Journal of Scientifical and Statistical Computation* 6(4):983–1003, 1985.

[14] S. Ray, R. Ungrangsi, F. De Pellegrini, A. Trachtenberg and D. Starobinski. Robust location detection in emergency sensor networks. *Proceedings of the IEEE INFOCOM*, pages 1044–1053, April 2003.

[15] P. J. Slater and D. F. Rall. On location-domination numbers for certain classes of graphs. *Congressus Numerantium*, 45:97–106, 1984.

[16] R. D. Skaggs. Identifying vertices in graphs and digraphs. PhD thesis, University of South Africa, South Africa, February 2007, available online at http://hdl.handle.net/10500/2226.

[17] A. Winter. Another algebraic proof of Bondy's theorem on induced subsets. *Journal of Combinatorial Theory Series A* 89(1):145–147, 2000.