Distance-based vertex identification in graphs: the outer multiset dimension

Reynaldo Gil-Pons[a], Yunior Ramírez-Cruz[b], Rolando Trujillo-Rasua[c,*], Ismael G. Yero[d]

[a] *Center for Pattern Recognition and Data Mining*
Patricio Lumumba s/n, Santiago de Cuba, 90500, Cuba
[b] *Interdisciplinary Centre for Security, Reliability and Trust*
University of Luxembourg
6 av. de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg
[c] *School of Information Technology*
Deakin University
221 Burwood Hwy, Burwood, VIC 3125, Australia
[d] *Departamento de Matemáticas, Escuela Politécnica Superior de Algeciras*
Universidad de Cádiz
Av. Ramón Puyol s/n, 11202 Algeciras, Spain

Abstract

Given a graph G and a subset of vertices $S = \{w_1, \ldots, w_t\} \subseteq V(G)$, the multiset representation of a vertex $u \in V(G)$ with respect to S is the multiset $m(u|S) = \{\!|d_G(u, w_1), \ldots, d_G(u, w_t)|\!\}$. A subset of vertices S such that $m(u|S) = m(v|S) \iff u = v$ for every $u, v \in V(G) \setminus S$ is said to be a multiset resolving set, and the cardinality of the smallest such set is the outer multiset dimension. We study the general behaviour of the outer multiset dimension, and determine its exact value for several graph families. We also show that computing the outer multiset dimension of arbitrary graphs is NP-hard, and provide methods for efficiently handling particular cases.

Keywords: graph, resolvability, resolving set, multiset resolving set, metric dimension, outer multiset dimension

---

*Corresponding author.

Email addresses: `rey@cerpamid.co.cu` (Reynaldo Gil-Pons), `yunior.ramirez@uni.lu` (Yunior Ramírez-Cruz), `rolando.trujillo@deakin.edu.au` (Rolando Trujillo-Rasua), `ismael.gonzalez@uca.es` (Ismael G. Yero)

1. Introduction

The characterisation of vertices in a graph by means of unique features, known as *distinguishability* or *resolvability*, has found applications in computer networks where nodes ought to be localised based on their properties rather than on identifiers [6], or to determine the social role of an actor in society in comparison to other peers with similar structural properties [7]. In fact, simple structural properties of vertices, such as their degree or the subgraphs induced by their neighbours, have been successfully used to re-identify (supposedly) anonymous users in social graphs [8, 16, 17].

This article focuses on vertex characterisations that are defined in relation to a subset of vertices of the graph. The earliest of such characterisations is known as *metric representation*, introduced independently by Slater [14] in 1975 and Harary and Melter [3] in 1976. Formally, given an ordered set of vertices $S = \{w_1, \ldots, w_t\} \subseteq V$ in a graph $G = (V, E)$, the metric representation of a vertex $u \in V$ with respect to S is the t-vector $\mathrm{r}(u|S) = (d_G(u, w_1), \ldots, d_G(u, w_t))$, where the metric $d_G(u, v)$ is computed as the length of a shortest $u - v$ path in G. An ordered subset S satisfying that every two distinct vertices u and v in the graph have different metric representation, i.e. $\mathrm{r}(u|S) \neq \mathrm{r}(v|S)$, is said to be a *resolving set*. The minimum cardinality amongst the resolving sets in a graph G is known as the *metric dimension* of G, and denoted as $\dim(G)$. The metric dimension of graphs has been extensively studied in literature since the 70s. Issues that are relevant to the present day, such as privacy in online social networks, are still benefiting from such research effort [9, 10, 11, 15].

The assumption that resolvability requires an order to exist (or be imposed) on a set S for obtaining metric representations remained unchallenged until 2017, when Simanjuntak, Vetrík, and Mulia introduced the notion of *multiset representation* [13] by looking at the multiset of distances rather than at the standard vector of distances.

For a vertex $u \in V$ and a vertex set $S \subseteq V$, the multiset representation of u with respect to S, denoted $\mathrm{m}(u|S)$, is defined by

$$\mathrm{m}(u|S) = \{\!\!\{ d_G(u, w_1), \ldots, d_G(u, w_t) \}\!\!\},$$

where $\{\!\!\{ . \}\!\!\}$ denotes a multiset.

Using this definition, the notions of resolvability in terms of the metric representation were straightforwardly extended to consider resolvability in

terms of the multiset representation [5, 13]. Our main observation in this article is that these straightforward extensions are in fact an oversimplification of the problem of distinguishing vertices in a graph based on the multiset representation. We argue that this problem has two flavours, one of which has been neglected in literature.
Contributions. This article makes the following contributions.

- We generalise the metric dimension of graphs to accommodate different characterisations of their vertices, such as the metric and multiset representations (Section 2). We show that the metric dimension problem with respect to the multiset representation admits two interpretations: one that can be found in the literature [5, 13] and is known as the *multiset dimension*, and another one that we call the *outer multiset dimension*. The latter is well-defined, whereas the multiset dimension [5, 13] is undefined for an infinite number of graphs. We also show that the outer multiset dimension finds applications on measuring the re-identification risk of users in a social graph. To the best of our knowledge, the multiset dimension has no obvious practical application.

- We characterise several graph families for which the outer multiset dimension can be easily determined, or bounded by the metric dimension (Section 3).

- We prove that the problem of computing the outer multiset dimension in a graph is NP-Hard (Section 4).

- We provide a polynomial computational procedure to calculate the outer multiset dimension of full 2-ary trees (Section 5), and a parallelisable algorithm for the general case of full δ-ary trees.

2. A generalisation of the metric dimension

We consider a simple and connected graph $G = (V, E)$ where V is a set of vertices and E a set of edges. The distance $d_G(v, u)$ between two vertices v and u in G is the number of edges in a shortest path connecting them. If there is no ambiguity, we will simply write $d(v, u)$.

The metric dimension of graphs has traditionally been studied based on the so-called metric representation, which is the vector of distances from a vertex to an ordered subset of vertices of the graph. To accommodate other types of relations between vertices, we generalise the metric dimension by considering any equivalence relation $\sim \subseteq V \times V$ over the set of vertices of the graph. That is, we consider a relation $\sim$ that is reflexive, symmetric, and transitive. We use $[u]_\sim$ to denote the equivalence class of the vertex $u \in V$ with respect to the relation $\sim$, while $V/\sim$ denotes the partition of V composed of the equivalence classes induced by $\sim$.

Definition 2.1 (Resolving and outer resolving set)**.** *A subset S of vertices in a graph $G = (V, E)$ is said to be resolving (resp. outer resolving) with respect to $\sim$ if all equivalence classes in $V/\sim$ (resp. $(V - S)/\sim$) have cardinality one.*

While standard resolving sets distinguish all vertices in a graph, outer resolving sets only look at those vertices that are not in S, hence the name. We remark that there exist applications working under the assumption that S is given, implying that vertices in S do not need to be distinguishable. For example, in an active re-identification attack on a social graph [1, 15], attackers first retrieve a set of attacker nodes by using a pattern recognition algorithm, then they re-identify other users in the network based on their metric representations with respect to the set of attacker nodes.

We use $\sim_S$ to denote the relation on the set of vertices of a graph defined by $u \sim_S v \iff \mathrm{r}(u|S) = \mathrm{r}(v|S)$, where $\mathrm{r}(v|S)$ is the vector of distances from v to vertices in S, and $\cong_S$ to denote the relation $u \cong_S v \iff \mathrm{m}(u|S) = \mathrm{m}(v|S)$, where $\mathrm{m}(v|S)$ is the multiset of distances from v to vertices in S. These two relations are interconnected in the following way.

Proposition 2.2. *For every non-trivial graph G, the following facts hold:*

i. Every resolving set of G with respect to $\cong_S$ is an outer resolving set.

ii. Every outer resolving set of G with respect to $\cong_S$ is an outer resolving set of G with respect to $\sim_S$.

iii. Every outer resolving set of G with respect to $\sim_S$ is a resolving set of G, and vice versa.

Proof. Let $S \subseteq V(G)$ be a resolving set of G with respect to $\cong_S$. Then, every pair of distinct vertices $u, v \in V(G)$ satisfy $\mathrm{m}(u|S) \neq \mathrm{m}(v|S)$. Thus, it trivially follows that the same property holds for every pair of distinct vertices $u, v \in V(G) \setminus S$. This completes the proof of (i).

The second property follows straightforwardly from the fact that $\mathrm{m}(u|S) \neq \mathrm{m}(v|S) \implies \mathrm{r}(u|S) \neq \mathrm{r}(v|S)$, and (iii) is a well-known property of resolving sets based on the metric representation. $\square$

Figure 1 depicts the relations between resolvability notions enunciated in Proposition 2.2 in the form of a hierarchy. In the figure, every arrow from resolvability notion A to resolvability notion B indicates that a set S which is resolving as defined by A is also resolving as defined by B. We use the following shorthand notation in Figure 1 and in the remainder of this article.

- *resolving set* to denote a resolving set with respect to $\sim_S$.
- *multiset resolving set* to denote a resolving set with respect to $\cong_S$.
- *outer resolving set* to denote an outer resolving set with respect to $\sim_S$.
- *outer multiset resolving set* to denote an outer resolving set with respect to $\cong_S$.

$$\begin{tikzcd}
 & \textit{multiset resolvability} \arrow[d] & \\
 & \textit{outer multiset resolvability} \arrow[dl] \arrow[dr] & \\
\textit{outer resolvability} \arrow[rr, leftrightarrow] & & \textit{resolvability}
\end{tikzcd}$$

Figure 1: Hierarchy of resolvability notions.

Definition 2.3 (Metric dimension and outer metric dimension)**.** *The metric dimension (resp. outer metric dimension) of a simple connected graph $G = (V, E)$ with respect to a structural relation $\sim$ is the minimum cardinality amongst a resolving (resp. outer resolving) set in G with respect to $\sim$. If no resolving (resp. outer resolving) set exists, we say that the metric dimension (resp. outer metric dimension) is undefined.*

An example of a metric dimension definition that is undefined for some graphs is given by Simanjuntak et al. [13]. They use the multiset representation to distinguish vertices. It is easy to prove that a complete graph has no multiset resolving set, which leads to indefinition. Conversely, the outer metric dimension with respect to the multiset representation is always defined, given that for every graph $G = (E, V)$, V is an outer multiset resolving set.

Overall, we highlight the fact that, while the outer metric dimension and the standard metric dimension with respect to the metric representation are equivalent (see Figure 1), the use of the multiset representation renders the outer metric dimension different from the standard metric dimension. In fact, the outer multiset dimension is defined for any graph, whereas the multiset dimension is not. Furthermore, recent privacy attacks and countermeasures on social networks [1, 12, 15] rely on the notion of *outer* resolving set, rather than on the original notion of resolving set. The remainder of this article is thus dedicated to the study of the *outer multiset dimension*, that is, the outer metric dimension with respect to $\cong_S$.

3. Basic results on the outer multiset dimension

In this section we characterise several graph families for which the outer multiset dimension can be easily determined, or bounded by the metric dimension otherwise. We start by providing notation that we use throughout the paper.

Notation. Let $G = (V, E)$ be a graph of order $n = |V(G)|$. We will say that G is *non-trivial* if $n \geq 2$. K_n, N_n, P_n and C_n stand for the complete, empty, path and cycle graphs, respectively, of order n. Moreover, we will use the notation $u \leftrightarrow_G v$ (negated as $u \not\leftrightarrow_G v$) to indicate that u and v are *adjacent* in G, that is $(u, v) \in E$. For a vertex v of G, $N_G(v)$ denotes the set of neighbours of v in G, that is $N_G(v) = \{u \in V(G) : \ u \leftrightarrow v\}$. The set $N_G(v)$ is called the *open neighbourhood* of the vertex v in G and $N_G[v] = N_G(v) \cup \{v\}$ is called the *closed neighbourhood* of v in G. The degree of a vertex v of G will be denoted by $\delta_G(v)$. If there is no ambiguity, we will drop the subscripts and simply write $u \leftrightarrow v$, $u \not\leftrightarrow v$, $N(v)$, etc. Two different vertices u, v are called *true twins* if $N[u] = N[v]$. Likewise, u, v are called *false twins* if $N(u) = N(v)$. In general, u, v are called *twins* if they are either true twins or false twins. Moreover, a vertex u is called a *twin* if there exists $v \neq u$ such that u and v are twins. Note that the property of being twins induces an equivalence relation on the vertex set of any graph.

Finally, we will use the notation $\dim_{\text{ms}}(G)$ for the outer multiset dimension of a graph G, and $\dim(G)$ for the standard metric dimension.

Remark 3.1. *For every non-trivial graph G of order n, the following facts hold:*

i. $1 \leq \dim_{\text{ms}}(G) \leq n-1$.

ii. $\dim_{\text{ms}}(G) \geq \dim(G)$.

Proof. The fact that $\dim_{\text{ms}}(G) \geq 1$ follows directly from the definition of outer multiset dimension, whereas $\dim_{\text{ms}}(G) \leq n-1$ follows trivially from the fact that every vertex v is the sole vertex in $V(G) \setminus (V(G) \setminus \{v\})$, and thus it has a unique multiset representation w.r.t. $V(G) \setminus \{v\}$, which is thus a multiset resolving set. The fact that $\dim_{\text{ms}}(G) \geq \dim(G)$ follows directly from item (ii) of Proposition 2.2. □ □

Once established the global bounds of the outer multiset dimension, we now focus on the extreme cases of these inequalities.

Remark 3.2. *A graph G satisfies* $\dim_{\text{ms}}(G) = 1$ *if and only if it is a path graph.*

Proof. Let G be a path graph. It is clear that the set $\{v\}$, where v is an extreme vertex of G, is a multiset resolving set of G, so $\dim_{\text{ms}}(G) \leq 1$. By item (ii) of Remark 3.1, $\dim_{\text{ms}}(G) \geq \dim(G) \geq 1$, so the equality holds. On the other hand, if G is not a path graph, then item (ii) of Remark 3.1 also leads to $\dim_{\text{ms}}(G) \geq \dim(G) \geq 2$, as the standard metric dimension of a graph is known to be 1 if and only if it is a path graph [2]. □

According to Remark 3.2, the cases where $\dim_{\text{ms}}(G) = \dim(G) = 1$ coincide. However, this is not the case for the upper bound of Remark 3.1 (i). Indeed, while it is easy to see that, for any positive integer n, the complete graph K_n satisfies $\dim_{\text{ms}}(K_n) = \dim(K_n) = n-1$, we have the fact that this is the sole family of graphs for which $\dim(K_n) = n-1$, whereas there exist graphs G such that $\dim_{\text{ms}}(G) = n-1 > \dim(G)$, as exemplified by the next results.

Example 3.3. *The cycle graphs C_4 and C_5 satisfy* $\dim_{\text{ms}}(C_4) = 3 > 2 = \dim(C_4)$ *and* $\dim_{\text{ms}}(C_5) = 4 > 2 = \dim(C_5)$.

Remark 3.4. *Every complete k-partite graph $G \cong K_{r_1,r_2,\ldots,r_k}$ such that $r_1 = r_2 = \ldots = r_k \geq 2$ and $\sum_{i=1}^{k} r_i = n$ satisfies* $\dim_{\mathrm{ms}}(G) = n - 1$.

Proof. Let $G \cong K_{r_1,r_2,\ldots,r_k}$ be a complete k-partite graph such that $r_1 = r_2 = \ldots = r_k \geq 2$. Let $u, v \in V(G)$ be two arbitrary vertices of G and let $S \subseteq V(G) \setminus \{u, v\}$. If $u \not\leftrightarrow v$, then $\mathrm{m}(u \mid S) = \mathrm{m}(v \mid S)$, as they are false twins in G. Consequently, S is not a multiset resolving set of G. We now treat the case where $u \leftrightarrow v$, for which we differentiate the following subcases:

- $S = V(G) \setminus \{u, v\}$. In this case, $\mathrm{m}(u \mid S) = \mathrm{m}(v \mid S) = \bigcup_{i=1}^{r-1} \{\!\{2\}\!\} \cup \bigcup_{i=1}^{r-1} \{\!\{1\}\!\} \cup \bigcup_{i=1}^{k-2} \bigcup_{j=1}^{r} \{\!\{1\}\!\}$, and so S is not a multiset resolving set of G.
- $S \subset V(G) \setminus \{u, v\}$. Here, if there exists some $x \in V(G) \setminus (S \cup \{u, v\})$ such that $x \not\leftrightarrow u$ ($x \not\leftrightarrow v$), then $\mathrm{m}(u \mid S) = \mathrm{m}(x \mid S)$ ($\mathrm{m}(v \mid S) = \mathrm{m}(x \mid S)$), as x and u (x and v) are false twins in G. Thus, S is not a multiset resolving set of G. Finally, if every $x \in V(G) \setminus (S \cup \{u, v\})$ satisfies $u \leftrightarrow x \leftrightarrow v$, then we have that $\mathrm{m}(u \mid S) = \mathrm{m}(v \mid S) = \bigcup_{i=1}^{r-1} \{\!\{2\}\!\} \cup \bigcup_{i=1}^{r-1} \{\!\{1\}\!\} \cup \bigcup_{i=1}^{t_1} \{\!\{1\}\!\} \cup \ldots \cup \bigcup_{i=1}^{t_{k-2}} \{\!\{1\}\!\}$, with $t_i \leq r$ for $i \in \{1, \ldots, k-2\}$, which entails that S is not a multiset resolving set of G.

Summing up the cases above, we have that no set $S \subseteq V(G)$ such that $|S| \leq n - 2$ is a multiset resolving set of G, and so $\dim_{\mathrm{ms}}(G) \geq n - 1$. The equality follows from item (i) of Remark 3.1. The proof is thus completed. $\square$

Example 3.3 shows two cases where the outer multiset dimension of a cycle graph is strictly larger than its standard metric dimension. With the exception of C_3, which satisfies $\dim_{\mathrm{ms}}(C_3) = \dim(C_3) = 2$, the strict inequality holds for every other cycle graph, as shown by the following result.

Remark 3.5. *Every cycle graph C_n of order $n \geq 6$ satisfies* $\dim_{\mathrm{ms}}(C_n) = 3$.

Proof. Consider an arbitrary pair of vertices $u, v \in V(C_n)$ and a pair of vertices $x, y \in V(C_n) \setminus \{u, v\}$ such that $ux \ldots yv$ is a path of C_n (note that for $n \geq 6$ at least one such pair x, y exists). We have that $\mathrm{m}(x \mid \{u, v\}) = \mathrm{m}(y \mid \{u, v\}) = \{\!\{1, d(u, v) \pm 1\}\!\}$, so no vertex subset of size 2 is a multiset resolving set of C_n. Thus, $\dim_{\mathrm{ms}}(C_n) \geq 3$.

Now, consider an arbitrary vertex $v_i \in V(C_n)$ and the set $S = \{v_{i-2}, v_i, v_{i+1}\}$, where the subscripts are taken modulo n. We differentiate the following cases for a pair of vertices $x, y \in V(C_n) \setminus S$:

1. $x = v_{i-1}$. In this case, $\mathrm{m}(x \mid S) = \{\!\!\{1, 1, 2\}\!\!\} \neq \mathrm{m}(y \mid S)$, as y is at distance 1 from at most one element in S.

2. x and y satisfy $\{d(x, v_i), d(x, v_{i-2})\} = \{d(y, v_i), d(y, v_{i-2})\}$. In this case, assuming without loss of generality that $a = d(x, v_i) < d(y, v_i)$, we have that $\mathrm{m}(x \mid S) = \{\!\!\{a, a+2, a-1\}\!\!\} \neq \{\!\!\{a, a+2, a+3\}\!\!\} = \mathrm{m}(y \mid S)$.

3. x and y satisfy $\{d(x, v_{i+1}), d(x, v_{i-2})\} = \{d(y, v_{i+1}), d(y, v_{i-2})\}$. In a manner analogous to that of the previous case, we assume without loss of generality that $b = d(x, v_{i+1}) < d(y, v_{i+1})$ and obtain that $\mathrm{m}(x \mid S) = \{\!\!\{b, b+1, b+3\}\!\!\} \neq \{\!\!\{b, b+2, b+3\}\!\!\} = \mathrm{m}(y \mid S)$.

4. In every other case, we have that $\min\{d \mid d \in \mathrm{m}(x \mid S)\} \neq \min\{d' \mid d' \in \mathrm{m}(y \mid S)$, so $\mathrm{m}(x \mid S) \neq \mathrm{m}(y \mid S)$.

Finally, summing up the cases above, we have that S is a multiset resolving set of G, and so $\dim_{\mathrm{ms}}(G) \leq |S| = 3$. This completes the proof. $\square$

Next, we characterise a large number of cases where the outer multiset dimension is strictly greater than the standard metric dimension. To that end, we first introduce some necessary notation. We represent by ${}^{n}_{r}C_{rep}$ the number of r-combinations, with repetition, from n elements. Likewise, we represent by ${}^{n}_{r}P_{rep}$ the number of r-permutations, with repetition, from n elements. Recall that ${}^{n}_{r}C_{rep} = \binom{r+n-1}{r} = \binom{r+n-1}{n-1}$, whereas ${}^{n}_{r}P_{rep} = n^r$. Finally, we recall the quantity $f(n, d)$, defined in [2] as the smallest positive integer k such that $k + d^k \geq n$. In an analogous manner, we define $f'(n, d)$ as the smallest positive integer k' such that $k' + \binom{r+d-1}{d-1} \geq n$. Since, by definition, ${}^{n}_{r}C_{rep} \leq^{n}_{r} P_{rep}$, we have that $f(n, d) \leq f'(n, d)$. With the previous definitions in mind, we introduce our next result.

Theorem 3.6. *For every graph $G = (V, E)$ of order n and diameter d such that* $\dim(G) < f'(n, d)$,

$$\dim_{\mathrm{ms}}(G) > \dim(G).$$

Proof. Let $G = (V, E)$ be a graph of order n and diameter d. It was proven in [2] that every such graph satisfies $\dim(G) \geq f(n, d)$. Indeed, no vertex subset $S \subseteq V$ such that $|S| < f(n, d)$ is a metric generator of G, because the number of different metric representations, with respect to S, for elements in $V \setminus S$ is at most $d^{|S|} < n - |S| = |V \setminus S|$. In general, if $|S| = r$, the set of all possible different metric representations for elements of $V \setminus S$ with respect to S is that of all permutations, with repetition, of r elements from $\{1, 2, \ldots, d\}$. Applying an analogous reasoning, we have that the set of all possible different multiset metric representations for elements of $V \setminus S$ with respect to S is that of all combinations, with repetition, of r elements from $\{1, 2, \ldots, d\}$. Thus, any multiset metric generator S of G must satisfy ${}^{n}_{|S|}C_{rep} \geq n - |S|$, so $\dim_{\mathrm{ms}}(G) \geq f'(n, d)$. In consequence, if $\dim(G) < f'(n, d)$, then $\dim_{\mathrm{ms}}(G) > \dim(G)$. $\square$

An example of the previous result is the wheel graph $W_{1,5} \cong \langle v \rangle + C_5$, which has diameter 2 (see Figure 2). As discussed in [2, 4], $\dim(W_{1,5}) = 2 = f(6, 2) < f'(6, 2) = 3 < 4 = \dim_{\mathrm{ms}}(W_{1,5})$.

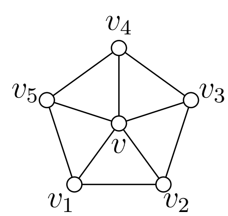

Figure 2: The wheel graph $W_{1,5} \cong \langle v \rangle + C_5$.

To conclude this section, we give a general result on the relation between outer multiset resolving sets and twin vertices, a particular case of which will be useful in further sections of this paper.

Proposition 3.7. *Let G be a non-trivial graph and let $S \subseteq V(G)$ be an outer multiset resolving set of G. Let $u, v \in V(G)$ be a pair of twin vertices. Then, $u \in S$ or $v \in S$.*

Proof. The proof follows from the fact that, as twin vertices, u and v satisfy $d(u, x) = d(v, x)$ for every $x \in V(G) \setminus \{u, v\}$, which entails that u and v have the same multiset representation according to any subset of $V(G)\setminus\{u, v\}$. $\square$

Corollary 3.8. *Let G be a non-trivial graph and let $\mathcal{T} = \{[u_1], [u_2], \ldots, [u_t]\}$ be the set of equivalence classes induced in $V(G)$ by the twin equivalence relation. Then,*

$$\dim_{\mathrm{ms}}(G) \geq \sum_{i=1}^{t} (|[u_i]| - 1) .$$

Proof. The result follows from the fact that, for every twin equivalence class, at most one element can be left out of any outer multiset resolving set. □

4. Complexity of the outer multiset dimension problem

In the previous section, we showed that algorithms able to compute the metric dimension can be used to determined or bound the outer multiset dimension. The trouble is, however, that calculating the metric dimension is NP-Hard [6]. We prove in this section that computing the outer multiset dimension of a simple connected graph is NP-hard as well. The proof is, in some way, inspired by the NP-hardness proof of the metric dimension problem given in [6]. To begin with, we formally state the decision problem associated to the computation of the outer multiset dimension:

Outer Multiset Dimension (DimMS)
INSTANCE: A graph $G = (V, E)$ and an integer k satisfying $1 \leq k \leq |V| - 1$.
QUESTION: Is $\dim_{\mathrm{ms}}(G) \leq k$?

Theorem 4.1. *The problem* DimMS *is NP-complete.*

Proof. The problem is clearly in NP. We give the NP-completeness proof by a reduction from 3-SAT. Consider an arbitrary input to 3-SAT, that is, a formula F with n variables and m clauses. Let $x_1, x_2, \ldots, x_n$ be the variables, and let $C_1, C_2, \ldots, C_m$ be the clauses of F. We next construct a connected graph G based on this formula F. To this end, we use the following gadgets.

For each variable x_i we construct a gadget as follows (see Fig. 3).

- Nodes T_i, F_i are the "true" and "false" ends of the gadget. The gadget is attached to the rest of the graph only through these nodes.
- Nodes a_i^1, a_i^2, b_i^1, b_i^2 "represent" the value of the variable x_i, that is, a_i^1 and a_i^2 will be used to represent that variable x_i is true, and b_i^1 and b_i^2 that it is false.
- Nodes d_i^1 and d_i^2 will help to differentiate between nodes in different gadgets.
- Q_i is a set of end-nodes of cardinality q_i adjacent to d_i^1. Notice that all these nodes are indistinguishable from d_i^2. Moreover, the cardinalities of these sets Q_i are pairwise distinct, which is necessary for our purposes in the proof. We further on state the explicit values of their cardinalities.

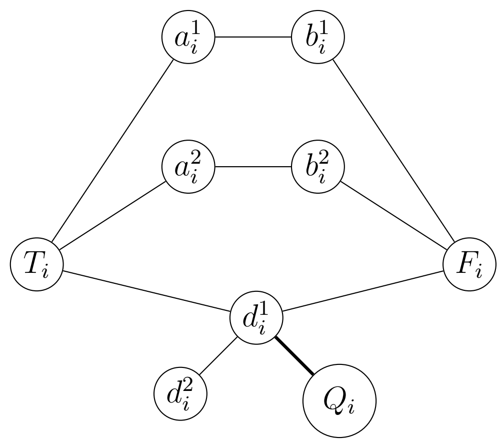

Figure 3: Gadget of a variable x_i

For each clause C_j we construct a gadget as follows (see Figure 4).

- Nodes c_j^1 and c_j^3 will be helpful in determining the truth value of C_j.
- Nodes c_j^2 and c_j^4 will help to differentiate between nodes in different gadgets.
- P_j is a set of end-nodes of cardinality p_j adjacent to c_j^2. Notice that all these nodes are indistinguishable from c_j^4. As in the case of the sets Q_i from the variable gadgets, the cardinalities of these sets P_j are also pairwise distinct.

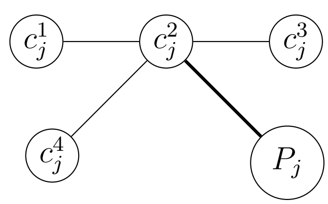

Figure 4: Gadget of clause C_j

As mentioned before, we require some conditions on the cardinalities of the sets P_i and Q_i from the variables and clauses gadgets, respectively. The values of their cardinalities (which we require in our proof) are as follows. For every $i \in \{1, \ldots, n\}$ we make $q_i = 2 \cdot i \cdot n$, and for every $j \in \{1, \ldots, m\}$ we make $p_j = 2 \cdot j \cdot n + 2n^2$. In concordance, we notice that the set of numbers p_i and q_j are pairwise distinct. Also, we clearly see that $\sum q_i + \sum p_j$ is polynomial in $n + m$.

The gadgets representing the variables and the gadgets representing the clauses are connected in the following way in order to construct our graph G.

- Nodes c_j^1, for every j, are adjacent to nodes T_i, F_i for all i.
- If a variable x_i does not appear in a clause C_j, then the nodes T_i, F_i are adjacent to c_j^3.
- If a variable x_i appears as a positive literal in a clause C_j, then the node F_i is adjacent to c_j^3.
- If a variable x_i appears as a negative literal in a clause C_j, then the node T_i is adjacent to c_j^3.

We first remark that the constructed graph G is connected, and that its order is polynomial in the number of variables and clauses of the original 3-SAT instance. We will prove now that the formula F is satisfiable if and only if the multiset dimension of G is exactly $M = \sum_{i=1}^{n} q_i + \sum_{j=1}^{m} p_j + n$.

First, let us look at some properties that must be fulfilled by a multiset resolving set S of minimum cardinality in G. First, as the nodes in $Q_i \cup \{d_i^2\}$, for every $i \in \{1, \ldots, n\}$, are indistinguishable among them, and at least $|Q_i|$ of them must be in S, we can assume without lost of generality that $Q_i \subset S$. By using a similar reasoning, also $P_j \subset S$ for every $j \in \{1, \ldots, m\}$. Moreover, for every $i \in \{1, \ldots, n\}$, at least one of the nodes a_i^1, a_i^2, b_i^1, b_i^2 must be in S, otherwise some pairs of them would have the same multiset representation, which is not possible. Thus, the cardinality of S is at least M. Clearly, if $M = |S|$, then we have already fully described a set of nodes that could represent S.

Lemma 4.2. *Consider a set S^* containing exactly M nodes given as follows. All nodes in Q_i for $i \in \{1, \ldots, n\}$, all nodes in P_j for $j \in \{1, \ldots, m\}$, and exactly one node from each set $\{a_i^1, a_i^2, b_i^1, b_i^2\}$ for $i \in \{1, \ldots, n\}$ are in S^*. Then, all pairs of nodes have different multiset representations with respect to S^*, except possibly c_j^1 and c_j^3 (for some $j \in \{1, \ldots, m\}$).*

Proof. To prove the lemma, we will explicitly compute the multiset representation of each node. For easier representation, we use a vector $(x_1, \ldots, x_n)$ to denote the multiset over positive integers such that 1 has multiplicity x_1, 2 has multiplicity x_2, and so on.

- $\mathrm{m}(c_j^4|S^*) = (0, p_j, 0, n, \cdots)$
- $\mathrm{m}(c_j^2|S^*) = (p_j, \cdots)$
- $\mathrm{m}(d_i^1|S^*) = (q_i, \cdots)$
- $\mathrm{m}(d_i^2|S^*) = (0, q_i, \cdots)$
- $(\mathrm{m}(T_i|S^*), \mathrm{m}(F_i|S^*))$ is equal to either $((1, q_i, \cdots), (0, q_i + 1, \cdots))$ or $((0, q_i + 1, \cdots), (1, q_i, \cdots))$
- $a_i^1, a_i^2, b_i^1, b_i^2$: Let's assume $b_i^2 \in S^*$. Then $\{\mathrm{m}(a_i^1|S^*), \mathrm{m}(a_i^2|S^*), \mathrm{m}(b_i^1|S^*)\} = \{(1, 0, q_i, \cdots), (0, 1, q_i, \cdots), (0, 0, q_i + 1, \cdots)\}$. An analogous result remains if the assumption that $b_i^2 \in S^*$ is dropped, based on the following observations. First, one and only one of the nodes $a_i^1, a_i^2, b_i^1, b_i^2$ is in S^*, and the distances from the other three to this one are exactly $1, 2, 3$ in some order. Second, each of these nodes have q_i nodes at distance 3.
- $\mathrm{m}(c_j^1|S^*) = (0, p_j + n, \sum_{i=1}^n q_i, \sum_{l=1}^m p_l - p_j)$
- c_j^3 : the number of nodes at distance two depends on which node belongs to S^* from each variable gadget. We distinguish three possible cases for the distance between c_j^3 and the node from S^* belonging to the gadget corresponding to a variable x_i.
  - If x_i appears in C_j as a positive literal, and $a_i^1 \in S^*$ or $a_i^2 \in S^*$, then such distance is 3.
  - If x_i appears in C_j as a negative literal, and $b_i^1 \in S^*$ or $b_i^2 \in S^*$, then such distance is 3.
  - If none of the above situations occurs, then such distance is 2.

  Therefore, the multiset representation of c_j^3 is related to the set $(0, p_j + w_j, \sum q_i + n - w_j, \sum p_l - p_j)$, where w_j is the number of nodes from gadgets representing some x_i matching the third case above.

Notice that, as the difference between any p_j and any q_i is at least $2n$, all pairs of nodes have also a different multiset representation, except possibly (c_j^1, c_j^3) that depend on the selected nodes from each variable gadget. We next particularise some of these situations.

- As $q_i \neq p_j$ for every $i \in \{1, \dots, n\}$ and every $j \in \{1, \dots, m\}$, we observe $\mathrm{m}(c_j^2|S^*) \neq \mathrm{m}(d_i^1|S^*)$, $\mathrm{m}(c_j^4|S^*) \neq \mathrm{m}(d_i^2|S^*)$.
- Since $p_j \neq q_i + 1$ for every $i \in \{1, \dots, n\}$ and every $j \in \{1, \dots, m\}$, we deduce $\mathrm{m}(c_j^4|S^*) \neq \mathrm{m}(T_i|S^*)$ and $\mathrm{m}(c_j^4|S^*) \neq \mathrm{m}(F_i|S^*)$.
- Since $p_{j_1} \neq p_{j_2} + n$ for every $j_1, j_2 \in \{1, \dots, m\}$, we get $\mathrm{m}(c_{j_1}^4|S^*) \neq \mathrm{m}(c_{j_2}^1|S^*)$.

Remaining cases trivially follow, and are left to the reader, and so the proof of the lemma is complete. □

We will now show a way to transform the set S^* into values for the variables x_i that will lead to a satisfiable assignment for F. If $S^* \cap \{a_i^1, a_i^2\} \neq \emptyset$ for some variable x_i, then we set the variable $x_i = \texttt{true}$ (with respect to S^*). Otherwise ($S^* \cap \{a_i^1, a_i^2\} = \emptyset$ or equivalently $S^* \cap \{b_i^1, b_i^2\} \neq \emptyset$), we set $x_i = \texttt{false}$. Hence, the clause C_j is `true` or `false` in the natural way, according to the values previously given to its variables.

Lemma 4.3. *Let S^* be a set of nodes as defined in the premise of Lemma 4.2. Then c_j^1 and c_j^3 have different multiset representations with respect to S^* if and only if the clause C_j is* `true`.

Proof. Notice that the distance between c_j^3 and the node in S^* from the gadget corresponding to x_i is 3 if and only if the clause C_j is `true` (see Lemma 4.2). Thus, $w_j = 0$ (as defined in Lemma 4.2) when the clause C_j is `false`, and only in this case $\mathrm{m}(c_j^3|S^*) = \mathrm{m}(c_j^1|S^*)$. □

By using the lemmas above, we conclude the NP-completeness reduction, through the following two lemmas.

Lemma 4.4. *If F is satisfiable, then the outer multiset dimension of G is M.*

Proof. Recall that $\dim_{\mathrm{ms}}(G) \geq M$. It remains to prove that if F is satisfiable then $\dim_{\mathrm{ms}}(G) \leq M$. Let us construct a set S in the following way. If x_i is `true`, then $a_i^1 \in S$. Otherwise (x_i is `false`), $b_i^1 \in S$. Also, we add to S all nodes in the sets P_j and Q_i. Hence, according to Lemmas 4.2 and 4.3, S is a multiset resolving set, and its cardinality is exactly M. □

Lemma 4.5. *If the outer multiset dimension of G is M, then F is satisfiable.*

Proof. Let S be a set of nodes of cardinality equal to the multiset dimension of G. Hence, as explained before, without lost of generality all nodes in the sets P_j, Q_i, and exactly one node of a_i^1, a_i^2, b_i^1, b_i^2, must belong to S, and no other node is in S. If $a_i^1 \in S$ or $a_i^2 \in S$, then let x_i be `true`. Otherwise, let x_i be `false`. Since S is a multiset resolving set, according to Lemmas 4.2 and 4.3, all clauses C_j of F must be `true`, unless the nodes c_j^1 and c_j^3 would have the same multiset representation, which is not possible. If all clauses of F are `true`, then F is satisfiable, as claimed. □

The last two lemmas together complete the reduction from 3-SAT to the problem of deciding whether the outer multiset dimension of a graph G is equal to a given positive integer. The latter problem can in turn be trivially reduced to DimMS. This completes the proof. □

5. Particular cases involving trees

Given that, in general, computing the outer multiset dimension of a graph is NP-hard, it remains an open question for which families of graphs the outer multiset dimension can be efficiently computed. The goal of this section is to provide a computational procedure and a closed formula to compute the outer multiset dimension of full δ-ary trees. A *full δ-ary tree* is a rooted tree whose root has degree δ, all its leaves are at the same distance from the root, and its descendants are either leaves or vertices of degree $\delta + 1$. We expect the results obtained in this section to pave the way for the study of the outer multiset dimension of general trees.
Notation. Given a multiset M and an element x, we denote the multiplicity of x in M as $M[x]$. We use $\epsilon_G(x)$ to denote the eccentricity of the vertex x in a graph G, which is defined as the largest distance between x and any other vertex in the graph. We will simply write $\epsilon(x)$ if the considered graph is clear from the context. Given a tree T rooted in w, we use T_x to denote the subtree induced by x and all descendants of x, *i.e.* those vertices having a shortest path to w that contains x. Finally, an outer multiset basis is said to be an outer multiset resolving set of minimum cardinality.

We start by enunciating a simple lemma that characterises multiset resolving sets in full δ-ary trees.

Lemma 5.1. *Let T be a full δ-ary tree rooted in w with $\delta > 1$. A set of vertices $S \subseteq V(T)$ is an outer multiset resolving set if and only if $\forall_{u,v \in V(T)\setminus S} \colon d(u,w) = d(v,w) \implies \mathrm{m}(u|S) \neq \mathrm{m}(v|S)$.*

Proof. Necessity follows from the definition of outer multiset resolving sets. To prove sufficiency we need to prove that

$$\forall_{u,v \in V(T)\setminus S} \colon d(u,w) \neq d(v,w) \implies \mathrm{m}(u|S) \neq \mathrm{m}(v|S).$$

Take two vertices $x, y \in V(T) \setminus S$ such that $d(x,w) < d(y,w)$. Because T is a full δ-ary tree, we obtain that $d(x,w) < d(y,w) \iff \epsilon_T(x) < \epsilon_T(y)$. Also, there must exist two leaf vertices y_1, y_2 in T which are siblings and satisfy $d(y_1,y) = d(y_2,y) = \epsilon_T(y)$. Considering that y_1 and y_2 are false twins, we obtain that $y_1, y_2 \notin S \implies \mathrm{m}(y_1|S) = \mathrm{m}(y_2|S)$. Therefore, given that $d(y_1,w) = d(y_2,w)$, it follows that $y_1 \in S$ or $y_2 \in S$. We assume, without loss of generality, that $y_1 \in S$. On the one hand, we have that $d(y_1,y) \in \mathrm{m}(y|S)$. On the other hand, because $d(y_1,y) = \epsilon_T(y) > \epsilon_T(x)$, we obtain that $d(y_1,y) \notin \mathrm{m}(x|S)$, implying that $\mathrm{m}(x|S) \neq \mathrm{m}(y|S)$. $\square$

Based on the result above, we provide conditions under which an outer multiset basis can be constructed in a recursive manner. Recall that an outer multiset basis is an outer multiset resolving set of minimum cardinality.

Lemma 5.2. *Given a natural number $\ell > 1$, let $T_1, \ldots, T_\delta$ be δ full δ-ary trees of depth ℓ with pairwise disjoint vertex sets. Let $w_1, \ldots, w_\delta$ be the roots of $T_1, \ldots, T_\delta$, respectively, and let T be the full δ-ary tree rooted in w defined by the set of vertices $V(T) = V(T_1) \cup \cdots \cup V(T_\delta) \cup \{w\}$ and edges $E(T) = E(T_1) \cup \cdots \cup E(T_\delta) \cup \{(w,w_1), \ldots, (w,w_\delta)\}$. Let $S_1, \ldots, S_\delta$ be outer multiset bases of $T_1, \ldots, T_\delta$, respectively. Then*

$$\forall_{i \neq j \in \{1,\ldots,\delta\}} \, \mathrm{m}_{T_i}(w_i|S_i)[\epsilon_{T_i}(w_i)] \neq \mathrm{m}_{Tj}(w_j|S_j)[\epsilon_{T_j}(w_j)] \implies$$
$$S_1 \cup \ldots \cup S_\delta \text{ is an outer multiset basis of } T.$$

Proof. Consider two vertices x and y in T such that $d_T(x,w) = d_T(y,w)$. We will prove that $\mathrm{m}_T(x|S) \neq \mathrm{m}_T(y|S)$, which gives that S is an outer multiset resolving set via application of Lemma 5.1. Our proof is split in two cases, depending on whether x and y are within the same sub-branch or not.

First, assume that $x \in V(T_i)$ and $y \in V(T_j)$ for some $i \neq j \in \{1, \ldots, \delta\}$. For every leaf vertex z in T, but not in T_i, we obtain that $d_T(x, z) = \epsilon_T(x)$. Because $\epsilon_T(x) > \epsilon_{T_i}(x)$, we get $\mathrm{m}_T(x|S)[\epsilon_T(x)] = \sum_{k \in \{1,\ldots,\delta\} \setminus \{i\}} \mathrm{m}_{T_k}(w_k|S)[\epsilon_{T_k}(w_k)]$. Analogously, we obtain that $\mathrm{m}_T(y|S)[\epsilon_T(y)] = \sum_{k \in \{1,\ldots,\delta\} \setminus \{j\}} \mathrm{m}_{T_k}(w_k|S)[\epsilon_{T_k}(w_k)]$. Therefore,

$$\mathrm{m}_T(x|S)[\epsilon_T(x)] - \mathrm{m}_T(y|S)[\epsilon_T(y)] = \mathrm{m}_{T_j}(w_j|S)[\epsilon_{T_j}(w_j)] - \mathrm{m}_{T_i}(w_i|S)[\epsilon_{T_i}(w_i)].$$

By considering the fact that $\mathrm{m}_{T_j}(w_j|S)[\epsilon_{T_j}(w_j)] \neq \mathrm{m}_{T_i}(w_i|S)[\epsilon_{T_i}(w_i)]$, we obtain that $\mathrm{m}_T(x|S)[\epsilon_T(x)] \neq \mathrm{m}_T(y|S)[\epsilon_T(y)]$, which implies that $\mathrm{m}_T(x|S) \neq \mathrm{m}_T(y|S)$.

For the second case assume that $x \in V(T_i)$ and $y \in V(T_i)$ for some $i \in \{1, \ldots, \delta\}$. This implies that $\mathrm{m}_T(x|S_i) \neq \mathrm{m}_T(y|S_i)$, because S_i is a multiset resolving set in T_i. Moreover, for every vertex $z \in V(T) \setminus V(T_i)$ it holds that $d_T(x, z) = d_T(y, z)$, which gives the expected result: $\mathrm{m}_T(x|S) \neq \mathrm{m}_T(y|S)$.

Up to here we have proved that S is an outer multiset resolving set. To prove that S is a basis, we only need to show that for any outer multiset resolving set S' in T, it is satisfied that the sets $S' \cap V(T_1), \ldots, S' \cap V(T_\delta)$ are outer multiset resolving sets in $T_1, \ldots, T_\delta$, respectively. Given that $S_1, \ldots, S_\delta$ are outer multiset bases, this would mean that S is an outer multiset resolving set of minimum cardinality.

We proceed by contrapositive. Let $S'_1 = S' \cap V(T_1), \ldots, S'_\delta = S' \cap V(T_2)$. Assume that S'_i is not an outer multiset resolving set in T_i for some $i \in \{1, \ldots, \delta\}$. Then, there must exist vertices x and y such that $\mathrm{m}_{T_i}(x|S'_i) = \mathrm{m}_{T_i}(y|S'_i)$. As in a previous reasoning, since x and y are both within T_i, it follows that $\forall_{z \in V(T) \setminus V(T_i)} d_T(x, z) = d_T(y, z)$. Hence, $\mathrm{m}_T(x|S') = \mathrm{m}_T(y|S')$, which is a contradiction. $\square$

Lemma 5.2 provides a sufficient condition for obtaining an outer multiset basis of a full δ-ary tree T by joining bases of the first level branches of T. This is useful for the development of a computational procedure that finds the outer multiset dimension of an arbitrary full δ-ary tree. Despite this fact, here we are interested in finding a closed formula for the outer multiset dimension of full δ-ary trees. The next result will prove itself a key element towards such a goal.

Theorem 5.3. *Let T^δ_ℓ be a full δ-ary tree of depth ℓ. Let n be the smallest positive integer such that there exist $\delta + 1$ outer multiset bases $S_1, \ldots, S_{\delta+1}$ in T^δ_n satisfying that* $\forall_{i \neq j \in \{1,\ldots,\delta+1\}} \mathrm{m}_{T^\delta_n}(w|S_i)[\epsilon_{T^\delta_n}(w)] \neq \mathrm{m}_{T^\delta_n}(w|S_j)[\epsilon_{T^\delta_n}(w)]$, *where w is the root of T_n^δ. Then, for every $\ell \geq n$, the outer multiset dimension of T_ℓ^δ is given by $\delta^{\ell-n} \times \dim_{\mathrm{ms}}(T_n^\delta)$.*

Proof. We proceed by induction.
Hypothesis. For some $\ell \geq n$, the following two conditions hold:

1. There exists $\delta+1$ outer multiset bases $S_1, \ldots, S_{\delta+1}$ in T_ℓ^δ satisfying that $\forall_{i \neq j \in \{1,\ldots,\delta+1\}} \mathrm{m}_{T_\ell^\delta}(w|S_i)[\epsilon_{T_\ell^\delta}(w)] \neq \mathrm{m}_{T_\ell^\delta}(w|S_j)[\epsilon_{T_\ell^\delta}(w)]$

2. The outer multiset dimension of T_ℓ^δ is given by $\delta^{\ell-n} \times \dim_{\mathrm{ms}}(T_n^\delta)$.

Clearly, these two conditions hold for $\ell = n$ (base case). The remainder of this proof will be dedicated to finding $\delta+1$ outer multiset bases $R_1, \ldots, R_{\delta+1}$ of $T_{\ell+1}^\delta$ that satisfy condition (1). The second condition will follow straightforwardly from the size of the bases $R_1, \ldots, R_{\delta+1}$.

Let w' be the root of $T_{\ell+1}^\delta$ and w the root of T_ℓ^δ. Let $w_1, \ldots, w_\delta$ be the children vertices of w' in $T_{\ell+1}^\delta$. For each sub-branch T_{w_k} of $T_{\ell+1}^\delta$, with $k \in \{1, \ldots, \delta\}$, let ϕ_k be an isomorphism from T_ℓ^δ to T_{w_k}. It follows that $\hat{S}_k = \{\phi_k(u) | u \in S_k\}$ is an outer multiset basis of T_{w_k}, for every $k \in \{1, \ldots, \delta+1\}$. Moreover, given that $\forall_{k \in \{1,\ldots,\delta+1\}} \mathrm{m}_{T_{w_k}}(w_k|\hat{S}_k) = \mathrm{m}_{T_\ell^\delta}(w|S_k)$, we conclude that

$$\forall_{i \neq j \in \{1,\ldots,\delta+1\}} \mathrm{m}_{T_{w_i}}(w_i|\hat{S}_i)[\epsilon_{T_{w_i}}(w_i)] \neq \mathrm{m}_{T_{w_j}}(w_j|\hat{S}_j)[\epsilon_{T_{w_j}}(w_j)]$$

By Theorem 5.2, we obtain that, for every $i \in \{1, \ldots, \delta+1\}$, the set $R_i = \bigcup_{j \in \{1,\ldots,\delta+1\} \setminus \{i\}} \hat{S}_j$ is an outer multiset basis of $T_{\ell+1}^\delta$. Moreover, for every $i \in \{1, \ldots, \delta+1\}$, the following holds

$$\mathrm{m}_{T_{\ell+1}^\delta}(w'|R_i)[\epsilon_{T_{\ell+1}^\delta}(w')] = \sum_{j \in \{1,\ldots,\delta+1\} \setminus \{i\}} \mathrm{m}_{T_{w_j}}(w_j|\hat{S}_i)[\epsilon_{T_{w_j}}(w_j)]$$

From the equation above we obtain that for every $i, j \in \{1, \ldots, \delta+1\}$,

$$\begin{aligned} &\mathrm{m}_{T_{\ell+1}^\delta}(w|R_i)[\epsilon_{T_{\ell+1}^\delta}(w)] - \mathrm{m}_{T_{\ell+1}^\delta}(w|R_j)[\epsilon_{T_{\ell+1}^\delta}(w)] = \\ &\qquad \mathrm{m}_{T_{w_j}}(w_j|\hat{S}_j)[\epsilon_{T_{w_j}}(w_j)] - \mathrm{m}_{T_{w_i}}(w_i|\hat{S}_i)[\epsilon_{T_{w_i}}(w_i)] \end{aligned}$$

Recall that $\forall_{k \in \{1,\ldots,\delta+1\}} \mathrm{m}_{T_{w_k}}(w_k|\hat{S}_k) = \mathrm{m}_{T_\ell^\delta}(w|S_k)$, which means that $i \neq j \implies \mathrm{m}_{T_{w_j}}(w_j|\hat{S}_j)[\epsilon_{T_{w_j}}(w_j)] \neq \mathrm{m}_{T_{w_i}}(w_i|\hat{S}_i)[\epsilon_{T_{w_i}}(w_i)]$. Therefore, we conclude that $T_{\ell+1}^\delta$ and $R_1, \ldots, R_{\delta+1}$ satisfy the first condition of the induction hypothesis, *i.e.*

$$\forall_{i \neq j \in \{1,\ldots,\delta+1\}} \mathrm{m}_{T_{\ell+1}^\delta}(w|R_i)[\epsilon_{T_{\ell+1}^\delta}(w)] \neq \mathrm{m}_{T_{\ell+1}^\delta}(w|R_j)[\epsilon_{T_{\ell+1}^\delta}(w)]$$

Finally, observe that $|R_1| = |\hat{S}_2| \times \cdots \times |\hat{S}_{\delta+1}| = \delta \times \dim_{\mathrm{ms}}(T^\delta_\ell)$. The second condition of the induction hypothesis states that $\dim_{\mathrm{ms}}(T^\delta_\ell) = \delta^{\ell-n} \times \dim_{\mathrm{ms}}(T^\delta_n)$, which gives that $\dim_{\mathrm{ms}}(T^\delta_{\ell+1}) = \delta \times \dim_{\mathrm{ms}}(T^\delta_\ell) = \delta^{\ell+1-n} \times \dim_{\mathrm{ms}}(T^\delta_n)$. □

We end this section by addressing the problem of finding the smallest n such that T^δ_n contains $\delta + 1$ outer multiset bases $S_1, \ldots, S_{\delta+1}$ satisfying the premises of Theorem 5.3. We do so by developing a computer program[1] that calculates such number via exhaustive search. The pseudocode for this computer program can be found in Algorithm 5. It reduces the search space by bounding the size of an outer multiset basis with the help of Lemma 5.2 (see Step 13 of Algorithm 5). That said, we cannot guarantee termination of Algorithm 5, essentially for two reasons. First, the computational complexity of each iteration of the algorithm is exponential on the size of T^δ_n while, at the same time, the size of T^δ_n exponentially increases with n. Second, there is no theoretical guarantees that such an n can be found for every δ.

Despite the exponential computational complexity of Algorithm 5, it terminates for $\delta = 2$. In this case, the smallest n satisfying the premises of Theorem 5.3 is $n = 4$. The three outer multiset bases S_1, S_2 and S_3 of T^2_4 are illustrated in Table 1 below[2]. We refer the interested reader to Appendix 6 for a visual representation of the bases shown in Table 1. The main corollary of this result is the following.

Corollary 5.4. *The outer multiset dimension of a full* 2*-ary tree* T^2_ℓ *of depth* ℓ *is:*

$$\dim_{\mathrm{ms}}(T^2_\ell) = \begin{cases} 1, & \text{if } \ell = 1 \\ 3, & \text{if } \ell = 2 \\ 6, & \text{if } \ell = 3 \\ 13, & \text{if } \ell = 4 \\ 2^{\ell-4} \times 13, & \text{otherwise} \end{cases}$$

Proof. The first four cases are calculated by an exhaustive search using a computer program that can be found at `https://github.com/rolandotr/graph`. The last case follows from Theorem 5.3. □

[1]The computer program can be found at `https://github.com/rolandotr/graph`.

[2]Our program took about 3.25 hours in a DELL computer with processor i7-7600U and installed memory 16GB to find the result shown in Table 1.

Algorithm 1 Given a natural number δ, finds the smallest n such that the full δ-ary tree of depth n satisfies the premises of Theorem 5.3.

Let $n = 0$ and T_n^δ a full δ-tree of depth n rooted in w
$min = 1$ ▷ Lower bound on the cardinality of a basis in T_1^δ
$max = \delta - 1$ ▷ Upper bound on the cardinality of a basis in T_1^δ
repeat
  for $i = min$ **to** max **do** ▷ Each of these iterations can be ran in parallel
    Let B be an empty set
    for all $S \subseteq V(T_n^\delta)$ s.t. $|S| = i$ **do**
      if S is a resolving set **then**
        if $\forall_{S' \in B}\, \mathrm{m}_{T_n^\delta}(w|S')[\epsilon_{T_n^\delta}(w)] \neq \mathrm{m}_{T_n^\delta}(w|S)[\epsilon_{T_n^\delta}(w)]$ **then**
          $B = B \cup \{S\}$
    if $B \neq \emptyset$ **then**
      break ▷ The outer multiset dimension of T_n^δ has been found
  $min = \dim_{\mathrm{ms}}(T_n^\delta) \times \delta$ ▷ See Lemma 5.2
  $max = min + \delta - 1$ ▷ This is the trivial upper bound
  $n = n + 1$
until $|B| \geq \delta + 1$
return n

Table 1: Three outer multiset bases of T_4^2 satisfying the premises of Theorem 5.3. Vertices of T_4^2 have been labelled by using a breadth-first ascending order, starting by labelling the root node with 1 and finishing with the label $2^{n+1} - 1$.

T_4^2	$S_1 = \{22, 24, 14, 25, 26, 16, 28, 18, 2, 8, 30, 20, 21\}$
	$S_2 = \{22, 12, 24, 14, 26, 16, 28, 18, 6, 8, 30, 20, 21\}$
	$S_3 = \{22, 24, 14, 25, 26, 16, 17, 28, 18, 8, 30, 20, 21\}$

It is worth remarking that Algorithm 5 can be paralellised and hence benefit from a computer cluster. Running the algorithm in a high performance computing facility is thus part of future work, which may lead to termination of Algorithm 5 for values of δ higher than 2.

6. Conclusions

In this paper we have addressed the problem of uniquely characterising vertices in a graph by means of their multiset metric representations. We have generalised the traditional notion of resolvability in such a way that the new formulation allows for different structural characterisations of vertices, including as particular cases the ones previously proposed in the literature. We have pointed out a fundamental limitation affecting previously proposed resolvability parameters based on the multiset representation, and have introduced a new notion of resolvability, the outer multiset dimension, which effectively addresses this limitation. Additionally, we have conducted a study of the new parameter, where we have analysed its general behaviour, determined its exact value for several graph families, and proven the NP-hardness of its computation, while providing an algorithm that efficiently handles some particular cases.

Acknowledgements: The work reported in this paper was partially funded by Luxembourg's Fonds National de la Recherche (FNR), via grants C15/IS/10428112 (DIST) and C17/IS/11685812 (PrivDA).

Bibliography

[1] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 181–190, New York, NY, USA, 2007. ACM.

[2] Gary Chartrand, Linda Eroh, Mark A Johnson, and Ortrud R Oellermann. Resolvability in graphs and the metric dimension of a graph. *Discrete Applied Mathematics*, 105(1-3):99–113, 2000.

[3] Frank Harary and Robert A. Melter. On the metric dimension of a graph. *Ars Combinatoria*, 2:191–195, 1976.

[4] Carmen Hernando, Merce Mora, Ignacio M Pelayo, Carlos Seara, José Cáceres, and Mari L Puertas. On the metric dimension of some families of graphs. *Electronic Notes in Discrete Mathematics*, 22:129–133, 2005.

[5] Varanoot Khemmani and Supachoke Isariyapalakul. The multiresolving sets of graphs with prescribed multisimilar equivalence classes. *Int. J. Math. Mathematical Sciences*, 2018:8978193:1–8978193:6, 2018.

[6] Samir Khuller, Balaji Raghavachari, and Azriel Rosenfeld. Landmarks in graphs. *Discrete Applied Mathematics*, 70(3):217–229, 1996.

[7] E. A. Leicht, Petter Holme, and M. E. J. Newman. Vertex similarity in networks. *Phys. Rev. E*, 73:026120, Feb 2006.

[8] Kun Liu and Evimaria Terzi. Towards identity anonymization on graphs. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 93–106, New York, NY, USA, 2008. ACM.

[9] Sjouke Mauw, Yunior Ramírez-Cruz, and Rolando Trujillo-Rasua. Anonymising social graphs in the presence of active attackers. *Transactions on Data Privacy*, 11(2):169–198, 2018.

[10] Sjouke Mauw, Yunior Ramírez-Cruz, and Rolando Trujillo-Rasua. Conditional adjacency anonymity in social graphs under active attacks. *Knowledge and Information Systems*, to appear, 2018.

[11] Sjouke Mauw, Rolando Trujillo-Rasua, and Bochuan Xuan. Counteracting active attacks in social network graphs. In *Data and Applications Security and Privacy XXX - 30th Annual IFIP WG 11.3 Conference, DBSec 2016, Trento, Italy, July 18-20, 2016. Proceedings*, pages 233–248, 2016.

[12] Wei Peng, Feng Li, Xukai Zou, and Jie Wu. Seed and grow: An attack against anonymized social networks. In *9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON 2012, Seoul, Korea (South), June 18-21, 2012*, pages 587–595, 2012.

[13] R. Simanjuntak, T. Vetrik, and P. Bintang Mulia. The multiset dimension of graphs. *ArXiv e-prints*, November 2017.

[14] Peter J. Slater. Leaves of trees. *Congressus Numerantium*, 14:549–559, 1975.

[15] Rolando Trujillo-Rasua and Ismael G. Yero. k-metric antidimension: A privacy measure for social graphs. *Information Sciences*, 328:403–417, 2016.

[16] Bin Zhou and Jian Pei. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, ICDE '08, pages 506–515, Washington, DC, USA, 2008. IEEE Computer Society.

[17] Lei Zou, Lei Chen, and M. Tamer Özsu. K-automorphism: A general framework for privacy preserving network publication. *Proc. VLDB Endow.*, 2(1):946–957, August 2009.

Appendix

Here, the reader can find graphical representations for different outer multiset bases in a full 2-ary tree of depth 4. In the figures, a basis is formed by the red-coloured vertices.

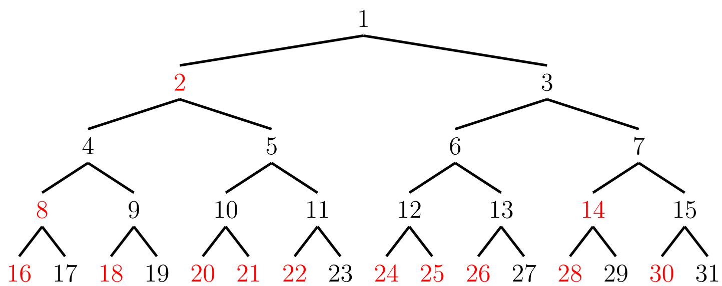

Figure 5: The multiset representation of the root vertex with respect to the set of red-coloured vertices is $\{1, 2^2, 4^{10}\}$.

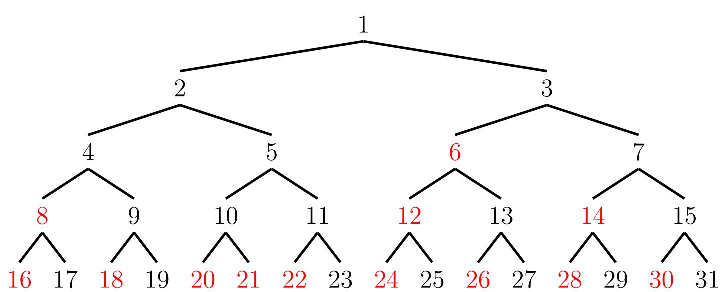

Figure 6: The multiset representation of the root vertex with respect to the set of red-coloured vertices is $\{2, 3^3, 4^9\}$.

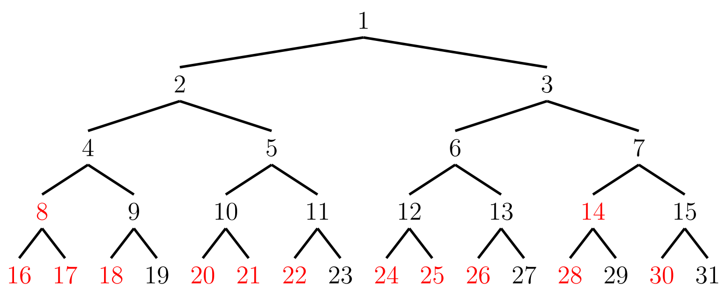

Figure 7: The multiset representation of the root vertex with respect to the set of red-coloured vertices is $\{3^2, 4^{11}\}$.