Metric Dimension Parameterized by Feedback Vertex Set and Other Structural Parameters

Esther Galby
CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

Liana Khazaliya
Saint Petersburg State University, Saint Petersburg, Russia

Fionn Mc Inerney
CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

Roohani Sharma
Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany

Prafullkumar Tale
CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

Abstract

For a graph G, a subset $S \subseteq V(G)$ is called a *resolving set* if for any two vertices $u, v \in V(G)$, there exists a vertex $w \in S$ such that $d(w,u) \neq d(w,v)$. The Metric Dimension problem takes as input a graph G and a positive integer k, and asks whether there exists a resolving set of size at most k. This problem was introduced in the 1970s and is known to be NP-hard [GT 61 in Garey and Johnson's book]. In the realm of parameterized complexity, Hartung and Nichterlein [CCC 2013] proved that the problem is W[2]-hard when parameterized by the natural parameter k. They also observed that it is FPT when parameterized by the vertex cover number and asked about its complexity under *smaller* parameters, in particular the feedback vertex set number. We answer this question by proving that Metric Dimension is W[1]-hard when parameterized by the feedback vertex set number. This also improves the result of Bonnet and Purohit [IPEC 2019] which states that the problem is W[1]-hard parameterized by the treewidth. Regarding the parameterization by the vertex cover number, we prove that Metric Dimension does not admit a polynomial kernel under this parameterization unless NP $\subseteq$ coNP/*poly*. We observe that a similar result holds when the parameter is the distance to clique. On the positive side, we show that Metric Dimension is FPT when parameterized by either the distance to cluster or the distance to co-cluster, both of which are smaller parameters than the vertex cover number.

2012 ACM Subject Classification Theory of computation → Parameterized complexity and exact algorithms

Keywords and phrases Metric Dimension, Parameterized Complexity, Feedback Vertex Set

Digital Object Identifier 10.4230/LIPIcs.MFCS.2022.51

Related Version *Full Version*: https://arxiv.org/abs/2206.15424

Funding Research supported by the European Research Council (ERC) consolidator grant No. 725978 SYSTEMATICGRAPH.

Acknowledgements The authors would like to thank Florent Foucaud for pointing us to Gutin et al. [21]. The article contains a result that subsumes our result conditionally refuting the polynomial kernel for Metric Dimension parameterized by the vertex cover number.

47th International Symposium on Mathematical Foundations of Computer Science (MFCS 2022).
Editors: Stefan Szeider, Robert Ganian, and Alexandra Silva; Article No. 51; pp. 51:1–51:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Problems dealing with distinguishing the vertices of a graph have attracted a lot of attention over the years, with the metric dimension problem being a classic one that has been vastly studied since its introduction in the 1970s by Slater [31], and independently by Harary and Melter [22]. Formally, given a graph G and an integer $k \geq 1$, the METRIC DIMENSION problem asks whether there exists a subset $S \subseteq V(G)$ of vertices of G of size at most k such that, for any two vertices $u, v \in V(G)$, there exists a vertex $w \in S$ such that $d(w, u) \neq d(w, v)$. If such a subset $S \subseteq V(G)$ exists, it is called a *resolving set*. The size of a smallest resolving set of a graph G is the *metric dimension* of G, and is denoted by $MD(G)$.

There are many variants and problems associated to the metric dimension, with *identifying codes* [27], *adaptive identifying codes* [4], and *locating dominating sets* [32] asking for the vertices to be distinguished by their neighborhoods in the subset chosen. Other variants of note are the *k-metric dimension*, where each pair of vertices must be resolved by k vertices in $S \subseteq V(G)$ instead of just one [15], and the *truncated metric dimension*, where the distance metric is the minimum of the distance in the graph and some integer k [34]. Along similar lines, in the *centroidal dimension* problem, each vertex must be distinguished by its relative distances to the vertices in $S \subseteq V(G)$ [17]. The metric dimension has also been considered in digraphs, with Bensmail et al. [6] providing a summary of the related work in this area. Interestingly, there are many game-theoretic variants of the metric dimension, such as *sequential metric dimension* [5], the *localization game* [9, 24], and the *centroidal localization game* [8]. The metric dimension and its variants have been studied for both their theoretical interest and their numerous applications such as in network verification [2], fault-detection in networks [36], pattern recognition and image processing [30], graph isomorphism testing [1], chemistry [11, 26], and genomics [35]. For more on these variants and others, see [28] for the latest survey.

Much of the related work around the metric dimension problem focuses on its computational complexity. METRIC DIMENSION was first shown to be NP-complete in general graphs in [19]. Later, it was also shown to be NP-complete in split graphs, bipartite graphs, co-bipartite graphs, and line graphs of bipartite graphs in [14], in bounded-degree planar graphs [12], and interval and permutation graphs of diameter 2 [18]. On the positive side, there are linear-time algorithms for METRIC DIMENSION in trees [31], cographs [14], and cactus block graphs [25], and a polynomial-time algorithm for outerplanar graphs [12].

Since the problem is NP-hard even for very restricted cases, it is natural to ask for ways to confront this hardness. In this direction, the parameterized complexity paradigm allows for a more refined analysis of the problem's complexity. In this setting, we associate each instance I with a parameter ℓ, and are interested in an algorithm with running time $f(\ell) \cdot |I|^{\mathcal{O}(1)}$ for some computable function f. Parameterized problems that admit such an algorithm are called fixed parameter tractable (FPT) with respect to the parameter under consideration. On the other hand, under standard complexity assumptions, parameterized problems that are hard for the complexity class W[1] or W[2] do not admit such fixed-parameter algorithms. A parameter may originate from the formulation of the problem itself (called natural parameters) or it can be a property of the input graph (called structural parameters).

Hartung and Nichterlein [23] proved that METRIC DIMENSION is W[2]-hard when parameterized by the natural parameter, the solution size k, even when the input graph is bipartite and has maximum degree 3. This motivated the study of the parameterized complexity of the problem under structural parameterizations. It was observed in [23] that the problem admits a simple FPT algorithm when parameterized by the vertex cover number. It took a considerable amount of work and/or meta-results to prove that there are FPT algorithms parameterized by the max leaf number [13], the modular width or treelength plus the maximum degree [3], and the treedepth [20]. In [14], they gave an XP algorithm parameterized by the feedback edge set number. Only recently, it was shown that Metric Dimension is W[1]-hard parameterized by the treewidth [7], answering an open question mentioned in [3, 12, 13]. This result was improved upon since, with it being shown that Metric Dimension is even NP-hard in graphs of treewidth 24 [29]. For more on the metric dimension, see [33] for a recent survey.

Our contributions. In this paper, we continue the analysis of structural parameterizations of Metric Dimension. See the Hasse diagram in Figure 1 for a summary of known results and our contributions. As mentioned before, it is known that Metric Dimension is W[1]-hard parameterized by the treewidth [7]. There are two natural directions to improve this result. One direction was to show that Metric Dimension is para-NP-hard parameterized by the treewidth, which was proven in [29]. Another direction is to prove that Metric Dimension is W[1]-hard for a higher parameter than treewidth, *i.e.*, one for which the treewidth is upper bounded by a function of it. A parameter fitting this profile is the feedback vertex set number since the treewidth of a graph G is upper bounded by the feedback vertex set number of G plus one. Moreover, the complexity of Metric Dimension parameterized by the feedback vertex set number is left as an open problem in [23], the seminal paper on the parameterized complexity of Metric Dimension. We take this direction and answer this open question of [23] by proving that Metric Dimension is W[1]-hard parameterized by the feedback vertex set number (see Sec. 2). We then revisit the complexity of the problem when parameterized by the vertex cover number. Recall that the problem is known to admit an FPT algorithm, and hence, a *kernel*, under this parameterization. We prove that, however, Metric Dimension does not admit a polynomial kernel unless $\mathsf{NP} \subseteq \mathsf{coNP}/poly$ when parameterized by the vertex cover number (see Sec. 3)[1]. On the positive side, we then show that Metric Dimension is FPT for the structural parameters the distance to cluster and the distance to co-cluster both of which are smaller parameters than the vertex cover number (see Sec. 4). Note that the FPT algorithm for the distance to cluster parameter implies an FPT algorithm for the distance to clique parameter. With a slight modification of the reduction in Sec. 3, we establish the problem does not admit a polynomial kernel, under the same assumption, when the parameter is the distance to clique.

In this extended abstract, we omit the standard terminology and some formal proofs (which are marked with $\star$) due to space constraints and present them in the full version on arXiv. Recall that any two vertices $u, v \in V(G)$ are *true twins* if $N[u] = N[v]$, and are *false twins* if $N(u) = N(v)$. A subset of vertices $S \subseteq V(G)$ *resolves* a pair of vertices $u, v \in V(G)$ if there exists a vertex $w \in S$ such that $d(w, u) \neq d(w, v)$. A vertex $u \in V(G)$ is *distinguished* by a subset of vertices $S \subseteq V(G)$ if, for any $v \in V(G) \setminus \{u\}$, there exists a vertex $w \in S$ such that $d(w, u) \neq d(w, v)$. We end this section with the following simple observation.

▶ **Observation 1.** *Let G be a graph. Then, for any (true or false) twins $u, v \in V(G)$ and any resolving set S of G, $S \cap \{u, v\} \neq \emptyset$.*

[1] After this paper was short-listed for the proceedings of MFCS 2022, Florent Foucaud informed us of the paper of Gutin et al. [21], which contains a slightly stronger result.

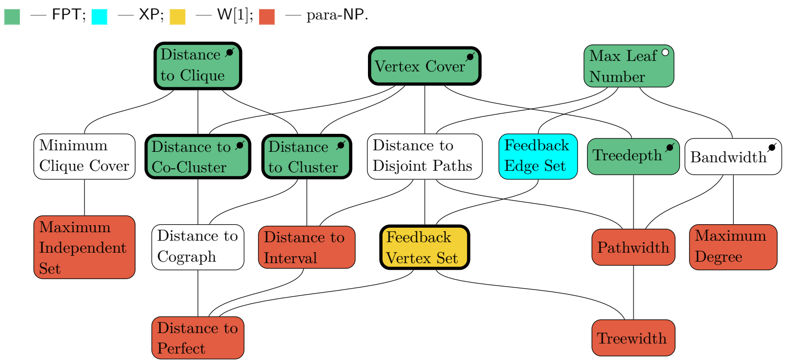

Figure 1 Hasse diagram of graph parameters and associated results for Metric Dimension. An edge indicates that the lower parameter is upper bounded by a function of the higher one. Colors correspond to the known hardness with respect to the highlighted parameter. The parameters for which the hardness remains an open question are not colored. The crossed bold circle in the upper-right corner means that Metric Dimension does not admit a polynomial kernel when parameterized by the marked parameter unless NP $\subseteq$ coNP/*poly*; the white one if a polynomial kernel exists. The bold borders highlight parameters that are covered in this paper. Also see Footnote 1.

2 The Feedback Vertex Set Number

In this section, we prove that Metric Dimension is W[1]-hard parameterized by the feedback vertex set number. To prove this, we reduce from the NAE-Integer-3-Sat problem defined as follows. An instance of this problem consists of a set X of variables, a set C of clauses, and an integer d. Each variable takes a value in $\{1,\dots,d\}$, and clauses are of the form $(x \le a_x, y \le a_y, z \le a_z)$, where $a_x, a_y, a_z \in \{1,\dots,d\}$. A clause is satisfied if not all three inequalities are true and not all are false. The goal is to find an assignment of the variables that satisfies all given clauses. This problem was shown to be W[1]-hard parameterized by the number of variables [10].

▶ **Theorem 2.** Metric Dimension *is* W[1]*-hard parameterized by the feedback vertex set number.*

Proof. We reduce from NAE-Integer-3-Sat: given an instance (X, C, d) of this problem, we construct an instance (G, k) of Metric Dimension as follows. For each variable $x \in X$, we introduce a cycle G_x of length $2d+2$ which has two distinguished *anchor vertices* u_1^x and u_2^x as depicted in Figure 2a; for convenience, we may also refer to u_1^x as v_0^x or w_0^x, and to u_2^x as v_{d+1}^x or w_{d+1}^x. For each clause $c = (x \le a_x, y \le a_y, z \le a_z)$, we introduce the gadget G_c depicted in Figure 2b consisting of two vertex-disjoint copies H_c and $H_{\bar{c}}$ of the same graph. More precisely, for $\ell \in \{c, \bar{c}\}$, H_ℓ consists of a $K_{1,3}$ on the vertex set $\{\ell, v^\ell, p_1^\ell, p_2^\ell\}$, where v^ℓ has degree three, and a path P_{b^ℓ} of length d connects ℓ to b^ℓ. The subgraph of G_c induced by $\{\ell, v^\ell, p_1^\ell, p_2^\ell \mid \ell \in \{c, \bar{c}\}\}$ is referred to as the *core of* G_c.

We further connect G_c to G_x, G_y, and G_z as follows. For every $t \in \{x, y, z\}$, we connect b^c to u_1^t by a path $P_1^{t,c}$ of length $4d - a_t$, and v^c to u_2^t by a path $P_2^{t,c}$ of length $4d + a_t - 1$. Furthermore, letting $w^{t,c}$ be the neighbor of v^c on $P_2^{t,c}$, we attach a copy $W^{t,c}$ of $K_{1,3}$ to $w^{t,c}$ by identifying $w^{t,c}$ with one of the leaves; we denote by $t_1^{t,c}$ and $t_2^{t,c}$ the two remaining leaves and refer to $W^{t,c}$ as a *pendant claw*. Similarly, for every $t \in \{x, y, z\}$, we connect $b^{\bar{c}}$ to u_2^t by a path $P_2^{t,\bar{c}}$ of length $3d + a_t$, and $v^{\bar{c}}$ to u_1^t by a path $P_1^{t,\bar{c}}$ of length $5d - a_t$. Furthermore, letting $w^{t,\bar{c}}$ be the neighbor of $v^{\bar{c}}$ on $P_1^{t,\bar{c}}$, we attach a copy $W^{t,\bar{c}}$ of $K_{1,3}$ to $w^{t,\bar{c}}$ by identifying $w^{t,\bar{c}}$ with one of the leaves; we denote by $t_1^{t,\bar{c}}$ and $t_2^{t,\bar{c}}$ the two remaining

(a) The variable gadget G_x.

(b) The clause gadget G_c is the disjoint union of H_c (left) and $H_{\bar{c}}$ (right).

Figure 2 The gadgets in the proof of Theorem 2.

leaves and refer to $W^{t,\bar{c}}$ as a *pendant claw*. Finally, we introduce a path $P = t_1 p t_2$ which we connect to the clause gadgets as follows. For every clause $c \in C$ and $\ell \in \{c, \bar{c}\}$, we connect p to v^ℓ by a path P_ℓ of length $2d$. Furthermore, letting w^ℓ be the neighbor of p on P_ℓ, we attach a copy W^ℓ of $K_{1,3}$ to w^ℓ by identifying w^ℓ with one of the leaves; we denote by t_1^ℓ and t_2^ℓ the two remaining leaves and refer to W^ℓ as a *pendant claw*. This concludes the construction of G (see Figure 3).

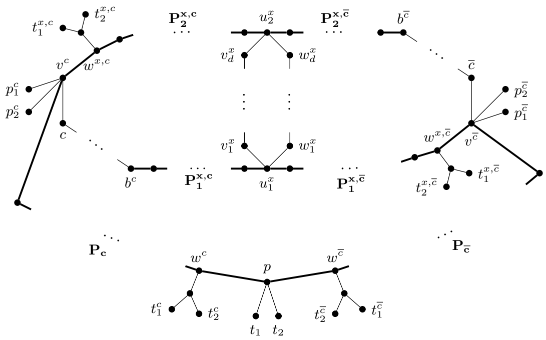

Figure 3 An illustration of the reduction in the proof of Theorem 2.

We set $k = |X| + 10|C| + 1$. Observe that the feedback vertex set number of G is at most $2|X| + 1$: indeed, removing $\{p\} \cup \{u_1^x, u_2^x \mid x \in X\}$ from G results in a graph without cycles. We next show that the instance (X, C, d) is satisfiable if and only if (G, k) is a YES-instance for METRIC DIMENSION. To this end, we first prove the following.

▷ Claim 3 (⋆). For any two distinct $s, t \in \{c, \bar{c} \mid c \in C\}$ and any two distinct variables $x, y \in X$, the following hold.

(i) The shortest path from H_s to H_t contains P_s and P_t as subpaths and has length $4d$.

(ii) $d(V(G_x), V(G_y)) \geq 6d$.

(iii) If x appears in the clause corresponding to s, then $d(V(G_x), V(H_s)) \geq 3d$.

(iv) If x does not appear in the clause corresponding to s, then any shortest path from G_x to H_s contains P_s as a subpath and has length at least $8d$.

▷ Claim 4 (⋆). For every clause $c = (x \leq a_x, y \leq a_y, z \leq a_z)$ and every $t \in \{x, y, z\}$, the following hold.

(i) For every $i \in \{0, \ldots, d+1\}$, if $i \leq a_t$, then the shortest path from v_i^t to c contains $P_1^{t,c}$ as a subpath and has length $5d + i - a_t$. Otherwise, the shortest path from v_i^t to c contains $P_2^{t,c}$ as a subpath and has length $5d + 1 + a_t - i$.

(ii) For every $i \in \{0, \ldots, d+1\}$, if $i \leq a_t - 1$, then the shortest path from v_i^t to v^c contains $P_1^{t,c}$ as a subpath and has length $5d + 1 + i - a_t$. Otherwise, the shortest path from v_i^t to v^c contains $P_2^{t,c}$ as a subpath and has length $5d + a_t - i$.

(iii) For every $i \in \{0, \ldots, d+1\}$, if $i \leq a_t - 2$, then the shortest path from v_i^t to $t_1^{t,c}$ contains $P_1^{t,c}$ as a subpath and has length $5d + 4 + i - a_t$. Otherwise, the shortest path from v_i^t to $t_1^{t,c}$ contains $P_2^{t,c}[u_2^t, w^{t,c}]$ as a subpath and has length $5d + 1 + a_t - i$.

▷ Claim 5 (★). For every clause $c = (x \leq a_x, y \leq a_y, z \leq a_z)$ and every $t \in \{x, y, z\}$, the following hold.

(i) For every $i \in \{0, \ldots, d+1\}$, if $i \leq a_t$, then the shortest path from v_i^t to $\overline{c}$ contains $P_1^{t,\overline{c}}$ as a subpath and has length $5d + 1 + i - a_t$. Otherwise, the shortest path from v_i^t to $\overline{c}$ contains $P_2^{t,\overline{c}}$ as a subpath and has length $5d + 1 + a_t - i$.

(ii) For every $i \in \{0, \ldots, d+1\}$, if $i \leq a_t + 1$, then the shortest path from v_i^t to $v^{\overline{c}}$ contains $P_1^{t,\overline{c}}$ as a subpath and has length $5d + i - a_t$. Otherwise, the shortest path from v_i^t to $v^{\overline{c}}$ contains $P_2^{t,\overline{c}}$ as a subpath and has length $5d + 2 + a_t - i$.

(iii) For every $i \in \{0, \ldots, d+1\}$, if $i \leq a_t + 2$, then the shortest path from v_i^t to $t_1^{t,\overline{c}}$ contains $P_1^{t,\overline{c}}[u_1^t, w^{t,\overline{c}}]$ as a subpath and has length $5d + 1 + i - a_t$. Otherwise, the shortest path from v_i^t to $t_1^{t,\overline{c}}$ contains $P_2^{t,\overline{c}}$ as a subpath and has length $5d + 5 + a_t - i$.

Assume first that (X, C, d) is satisfiable and let $\phi : X \to \{1, \ldots, d\}$ be an assignment of the variables satisfying every clause in C. We construct a resolving set S of G as follows. First, we add t_1 to S. For every variable $x \in X$, we add $v_{\phi(x)}^x$ to S. Finally, for every clause $c \in C$, we add $p_1^c, p_1^{\overline{c}}, t_1^c, t_1^{\overline{c}}$ to S and further add, for every variable t appearing in c, $t_1^{t,c}, t_1^{t,\overline{c}}$ to S. Note that $|S| = k$ and that every vertex of S is distinguished by itself. Let us show that S is indeed a resolving set of G. To this end, consider two distinct vertices $u, v \in V(G)$. We distinguish the following cases to show that there exists $w \in S$ such that $d(w, u) \neq d(w, v)$.

Case 1. *At least one of u and v belongs to a pendant claw.* W.l.o.g., assume first that $u \in V(W^\ell)$, where $\ell \in \{c, \overline{c} \mid c \in C\}$. If $v \in V(G) \setminus V(W^\ell)$, then $d(t_1^\ell, v) > 2 \geq d(t_1^\ell, u)$. Suppose therefore that $v \in V(W^\ell)$ as well. If $\{u, v\} \neq \{w^\ell, t_2^\ell\}$, then $d(t_1^\ell, u) \neq d(t_1^\ell, v)$. If $\{u, v\} = \{w^\ell, t_2^\ell\}$, then $d(t_1, w^\ell) = 2 < 4 = d(t_1, t_2^\ell)$. Second, assume that $u \in V(W^{t,\ell})$, where $\ell \in \{c, \overline{c}\}$ for some clause $c \in C$ and t is a variable appearing in clause c. If $v \in V(G) \setminus V(W^{t,\ell})$, then $d(t_1^{t,\ell}, v) > 2 \geq d(t_1^{t,\ell}, u)$. Suppose therefore that $v \in V(W^{t,\ell})$ as well. If $\{u, v\} \neq \{w^{t,\ell}, t_2^{t,\ell}\}$, then $d(t_1^{t,\ell}, u) \neq d(t_1^{t,\ell}, v)$. If $\{u, v\} = \{w^{t,\ell}, t_2^{t,\ell}\}$, then $d(p_1^\ell, w^{t,\ell}) = 2 < 4 = d(p_1^\ell, t_2^{t,\ell})$.

Case 2. *At least one of u and v belongs to the core of a clause gadget.* Assume, w.l.o.g., that $u \in \{\ell, v^\ell, p_1^\ell, p_2^\ell\}$, where $\ell \in \{c, \overline{c}\}$ for some clause $c = (x \leq a_x, y \leq a_y, z \leq a_z)$. If v is not a neighbor of v^ℓ, then $d(p_1^\ell, v) > 2 \geq d(p_1^\ell, u)$. If v is the neighbor of v^ℓ on the path P_ℓ, then $d(t_1, v) = d < d(t_1, u)$. Also, $v = w^{t,\ell}$ was covered by the previous case. So, consider $v \in \{\ell, v^\ell, p_1^\ell, p_2^\ell\}$. If $\{u, v\} \neq \{\ell, p_2^\ell\}$, then clearly $d(p_1^\ell, u) \neq d(p_1^\ell, v)$. Assume therefore that $\{u, v\} = \{\ell, p_2^\ell\}$. Since ϕ satisfies c, there exist $t, f \in \{x, y, z\}$ such that $\phi(t) \leq a_t$ and $\phi(f) > a_f$. Then, either $\phi(t) < a_t$, in which case, by Claim 4(i) and (ii),

$$d(v_{\phi(t)}^t, c) = 5d + \phi(t) - a_t < 5d + \phi(t) - a_t + 2 = d(v_{\phi(t)}^t, v^c) + 1 = d(v_{\phi(t)}^t, p_2^c),$$

or $\phi(t) = a_t$, in which case, by Claim 4(i) and (ii),

$$d(v_{\phi(t)}^t, c) = 5d < 5d + 1 = d(v_{\phi(t)}^t, v^c) + 1 = d(v_{\phi(t)}^t, p_2^c).$$

Similarly, either $\phi(f) = a_f + 1$, in which case, by Claim 5(i) and (ii),

$$d(v_{\phi(f)}^f, \overline{c}) = 5d < 5d + 2 = d(v_{\phi(f)}^f, v^{\overline{c}}) + 1 = d(v_{\phi(f)}^f, p_2^{\overline{c}}),$$

or $\phi(f) > a_f + 1$, in which case, by Claim 5(i) and (ii),

$$d(v_{\phi(f)}^f, \overline{c}) = 5d + 1 + a_f - \phi(f) < 5d + 3 + a_f - \phi(f) = d(v_{\phi(f)}^f, v^{\overline{c}}) + 1 = d(v_{\phi(f)}^f, p_2^{\overline{c}}).$$

In all cases, we conclude that there exists $w \in S$ such that $d(w, \ell) \neq d(w, p_2^\ell)$.

Case 3. *At least one of u and v belongs to a variable gadget.* Assume, w.l.o.g., that $u \in V(G_x)$ for some variable $x \in X$. By the previous cases, we may assume that v does not belong to the core of a clause gadget or a pendant claw. If $v \in V(G_y)$ for some variable $y \neq x$, then by Claim 3(ii),

$$d(v^x_{\phi(x)}, u) \leq d + 1 < 6d \leq d(V(G_x), V(G_y)) \leq d(v^x_{\phi(x)}, v).$$

Now, suppose that $v \in V(G_x)$ as well. If $\{u, v\} = \{v^x_i, w^x_i\}$ for some $i \in [d]$, then

$$d(v^x_{\phi(x)}, v^x_i) = |\phi(x) - i| < d(v^x_{\phi(x)}, w^x_i) = \min\{\phi(x) + i, 2d + 2 - \phi(x) - i\}.$$

Suppose next that $u = v^x_i$ and $v = v^x_j$ for two distinct $i, j \in \{0, \dots, d+1\}$, say $i < j$, w.l.o.g. Consider a clause $c = (x \leq a_x, y \leq a_y, z \leq a_z)$ containing x. If $j < a_x$, then by Claim 4(ii),

$$d(p^c_1, v^x_i) = d(v^c, v^x_i) + 1 = 5d + 2 + i - a_x < 5d + 2 + j - a_x = d(v^c, v^x_j) + 1 = d(p^c_1, v^x_j).$$

Now, suppose that $i < a_x \leq j$. Then, by Claim 4(ii),

$$d(p^c_1, v^x_i) - d(p^c_1, v^x_j) = 5d + 2 + i - a_x - (5d + a_x - j + 1) = i + j + 1 - 2a_x.$$

Thus, if $i + j + 1 - 2a_x \neq 0$, then $d(p^c_1, v^x_i) \neq d(p^c_1, v^x_j)$. Now, if $i + j + 1 - 2a_x = 0$, then either $j = a_x$ and $i = a_x - 1$, in which case, by Claim 4(iii),

$$d(t^{x,c}_1, v^x_i) = 5d + 2 > 5d + 1 = d(t^{x,c}_1, v^x_j),$$

or $j > a_x$ and $i < a_x - 1$, in which case, by Claim 4(iii),

$$d(t^{x,c}_1, v^x_j) = 5d + 1 + a_x - j = 5d + 2 + i - a_x < 5d + 4 + i - a_x = d(t^{x,c}_1, v^x_i).$$

Finally, if $a_x \leq i < j$, then by Claim 4(ii),

$$d(p^c_1, v^x_j) = 5d + 1 + a_x - j < 5d + 1 + a_x - i = d(p^c_1, v^x_i).$$

Since for any $t \in V(G) \setminus V(G_x)$ and $k \in [d]$, $d(t, v^x_k) = d(t, w^x_k)$, we conclude similarly if either $u = v^x_i$ and $v = w^x_j$, or $u = w^x_i$ and $v = w^x_j$ for two distinct $i, j \in \{0, \dots, d+1\}$. Assume, henceforth, that $v \notin \bigcup_{x \in X} V(G_x)$. If v does not belong to a path connecting G_x to some clause gadget, then

$$d(v^x_{\phi(x)}, v) \geq \min_{c \in C} d(V(G_x), V(G_c)) \geq 3d > d + 1 \geq d(v^x_{\phi(x)}, u)$$

by Claim 3(iii) and (iv). Suppose therefore that $v \in V(P^{x,\ell}_i)$, where $i \in \{1, 2\}$ and $\ell \in \{c, \overline{c}\}$ for some clause $c = (x \leq a_x, y \leq a_y, z \leq a_z)$ containing x. W.l.o.g., let us assume that $u = v^x_j$ where $j \in \{0, \dots, d+1\}$.

Assume first that $\ell = c$ and $i = 1$. Let $P^{x,c}_1 = z_0 \dots z_{4d-a_x}$, where $z_0 = b^c$ and $z_{4d-a_x} = u^x_1$. Let $v = z_k$, where $k \in [4d - a_x - 1]$. If $j \leq a_x - 1$, then by Claim 4(ii), the shortest path from p^c_1 to v^x_j contains $P^{x,c}_1$ as a subpath, which implies in particular that $d(p^c_1, v) < d(p^c_1, u)$. Suppose therefore that $j \geq a_x$. Then, by Claim 4(ii),

$$d(p^c_1, u) - d(p^c_1, v) = 5d + 1 + a_x - j - (d + k + 2).$$

Thus, if $5d + 1 + a_x - j - (d + k + 2) \neq 0$, then $d(p^c_1, u) \neq d(p^c_1, v)$. Now, if $5d + 1 + a_x - j - (d + k + 2) = 0$, then by Claim 4(iii),

$$d(t^{x,c}_1, u) = 5d + 1 + a_x - j = d + k + 2 < d + k + 4 = d(t^{x,c}_1, v).$$

Second, assume that $\ell = c$ and $i = 2$. Let $P_2^{x,c} = z_0 \dots z_{4d+a_x-1}$, where $z_0 = v^c$ and $z_{4d+a_x-1} = u_2^x$. Let $v = z_k$, where $k \in [4d + a_x - 2]$ (note that since v does not belong to the core of a clause gadget or a pendant claw by assumption, in fact $k \geq 2$). If $j \geq a_x$, then by Claim 4(ii), the shortest path from p_1^c to u contains $P_2^{x,c}$ as a subpath, which implies in particular that $d(p_1^c, v) < d(p_1^c, u)$. Otherwise, $j \leq a_x - 1$, in which case

$$d(p_1^c, u) - d(p_1^c, v) = 5d + 2 + j - a_x - (k + 1).$$

Thus, if $5d+2+j-a_x-(k+1) \neq 0$, then $d(p_1^c, u) \neq d(p_1^c, v)$. Now, if $5d+2+j-a_x-(k+1) = 0$, then $j < a_x - 1$ since $k < 5d$, and so, by Claim 4(iii),

$$d(t_1^{x,c}, u) = 5d + 4 + j - a_x = k + 3 > k + 1 = d(t_1^{x,c}, v).$$

Third, assume that $\ell = \overline{c}$ and $i = 1$. Let $P_1^{x,\overline{c}} = z_0 \dots z_{5d-a_x}$, where $z_0 = v^{\overline{c}}$ and $z_{5d-a_x} = u_1^x$. Let $v = z_k$, where $k \in [5d - a_x - 1]$ (note that since v does not belong to the core of a clause gadget or a pendant claw by assumption, in fact $k \geq 2$). If $j \leq a_x + 1$, then by Claim 5(ii), the shortest path from $p_1^{\overline{c}}$ to v_j^x contains $P_1^{x,\overline{c}}$ as a subpath which implies in particular that $d(p_1^{\overline{c}}, v) < d(p_1^{\overline{c}}, u)$. Suppose therefore that $j \geq a_x + 2$. Then, by Claim 5(ii),

$$d(p_1^{\overline{c}}, v_j^x) - d(p_1^{\overline{c}}, z_k) = 5d + 3 + a_x - j - (k + 1).$$

Thus, if $5d + 3 + a_x - j - (k + 1) \neq 0$, then $d(p_1^{\overline{c}}, v_j^x) \neq d(p_1^{\overline{c}}, z_k)$. Now, if $5d + 3 + a_x - j - (k + 1) = 0$, then $j > a_x + 2$ since $k < 5d$, and so, by Claim 5(iii),

$$d(t_1^{x,\overline{c}}, v_j^x) = 5d + 5 + a_x - j = k + 3 > k + 1 = d(t_1^{x,\overline{c}}, z_k).$$

Assume finally that $\ell = \overline{c}$ and $i = 2$. Let $P_2^{x,\overline{c}} = z_0 \dots z_{3d+a_x}$, where $z_0 = b^{\overline{c}}$ and $z_{3d+a_x} = u_2^x$. Let $v = z_k$, where $k \in [3d + a_x - 1]$. If $j \geq a_x + 2$, then by Claim 5(ii), the shortest path from $p_1^{\overline{c}}$ to u contains $P_2^{x,\overline{c}}$ as a subpath, which implies in particular that $d(p_1^{\overline{c}}, v) < d(p_1^{\overline{c}}, u)$. Suppose therefore that $j \leq a_x + 1$. Then, by Claim 5(ii),

$$d(p_1^{\overline{c}}, v_j^x) - d(p_1^{\overline{c}}, z_k) = 5d + 1 + j - a_x - (d + k + 2).$$

Thus, if $5d + 1 + j - a_x - (d + k + 2) \neq 0$, then $d(p_1^{\overline{c}}, v_j^x) \neq d(p_1^{\overline{c}}, z_k)$. Now, if $5d + 1 + j - a_x - (d + k + 2) = 0$, then $j < a_x + 1$ since $k < 4d$, and so, by Claim 5(iii),

$$d(t_1^{x,\overline{c}}, v_j^x) = 5d + 1 + j - a_x = d + k + 2 < d + k + 4 = d(t_1^{x,\overline{c}}, z_k).$$

In all the subcases, we conclude that there exists $w \in S$ such that $d(w, u) \neq d(w, v)$.

Case 4. *None of the above.* First, note that p is distinguished by S since it is the unique vertex of G at distance 1 from t_1. Second, t_2 is distinguished by S since it is the unique vertex of G at distance 2 from t_1 and distance 4 from t_1^c and $t_1^{\overline{c}}$ for all $c \in C$. Thus, in this last case, we can assume that both u and v belong either to paths connecting gadgets or to some path P_{b^ℓ}, where $\ell \in \{c, \overline{c} \mid c \in C\}$. Assume first that $u \in V(P_\ell)$ for some $\ell \in \{c, \overline{c} \mid c \in C\}$. If $v \in V(P_\ell)$ as well, then surely $d(t_1, u) \neq d(t_1, v)$. If $v \in V(P_q)$ for some $q \in \{c, \overline{c} \mid c \in C\}$ different from ℓ, then $d(p_1^\ell, u) < d(p_1^\ell, v)$ since the unique shortest path from p_1^ℓ to v contains P_ℓ as a subpath. Finally, if there exists $q \in \{c, \overline{c} \mid c \in C\}$ such that v belongs to P_{b_q} or to some path connecting H_q to a variable gadget, then $d(t_1, v) > d(t_1, v^q) \geq d(t_1, u)$.

Second, assume that $u \in V(P_i^{x,\ell})$, where $i \in [2]$ and $\ell \in \{c, \overline{c}\}$ for some clause c containing variable x. Note that by the previous paragraph, we may assume that $v \notin \bigcup_{q \in C} V(P_q) \cup V(P_{\overline{q}})$. Suppose first that $v \in V(P_j^{y,q})$, where $j \in [2]$ and $q \in \{c', \overline{c}'\}$ for some clause c' containing variable y. Note that $d(t_1^\ell, u) = d(t_1, u)$. So, if $q \neq \ell$, then either $d(t_1, u) \neq d(t_1, v)$, or

$$d(t_1^\ell, v) - d(t_1^\ell, u) = d(t_1^\ell, p) + d(t_1, v) - 1 - d(t_1^\ell, u) = d(t_1, v) + 2 - d(t_1, u) = 2.$$

Thus, assume that $q = \ell$. Suppose first that $x = y$. If $i = j$, then surely $d(p_1^\ell, u) \neq d(p_1^\ell, v)$. Otherwise, assume, w.l.o.g., that u belongs to the path containing $w^{x,\ell}$. Note that $d(t_1^{x,\ell}, u) = d(p_1^\ell, u)$. Then, either $d(p_1^\ell, u) \neq d(p_1^\ell, v)$, or

$$d(t_1^{x,\ell}, v) - d(t_1^{x,\ell}, u) = d(t_1^{x,\ell}, v^\ell) + d(p_1^\ell, v) - 1 - d(t_1^{x,\ell}, u) = d(p_1^\ell, v) + 2 - d(p_1^\ell, u) = 2.$$

Second, suppose that $x \neq y$. If u belongs to the path containing $w^{x,\ell}$, then we argue as previously. By symmetry, we may also assume that v does not belong to the path containing $w^{y,\ell}$. This implies, in particular, that $i = j$ and b^ℓ is the endpoint in H_ℓ of both $P_i^{x,\ell}$ and $P_j^{y,\ell}$. Thus, $d(v_{\phi(x)}^x, u) \leq d(v_{\phi(x)}^x, u_i^x) + d(u_i^x, b^\ell) \leq d + 4d$. First, note that if a shortest path P from $v_{\phi(x)}^x$ to v contains $P_i^{x,\ell}$ as a subpath, then since $u \in V(P_i^{x,\ell})$, it follows that $d(v_{\phi(x)}^x, u) < d(v_{\phi(x)}^x, v)$. Hence, we may assume that P contains a vertex in G_y or both v^ℓ and b^ℓ. By Claim 3(ii), if P contains a vertex in G_y, then

$$d(v_{\phi(x)}^x, v) > d(V(G_x), V(G_y)) \geq 6d > 5d \geq d(v_{\phi(x)}^x, u).$$

Otherwise, P contains v^ℓ and b^ℓ, and so, letting $t \in [2] \setminus \{i\}$, we get that

$$\mathsf{lgt}(P) \geq d(v_{\phi(x)}^x, u_t^x) + d(u_t^x, v^\ell) + d(v^\ell, b^\ell) + d(b^\ell, v) \geq 1 + 4d + d + 1 > 5d \geq d(v_{\phi(x)}^x, u).$$

Suppose finally that $u \in V(P_{b^\ell})$ for some $\ell \in \{c, \overline{c} \mid c \in C\}$. By the two previous paragraphs, we may assume that $v \in V(P_{b^q})$ for some $q \in \{c, \overline{c} \mid c \in C\}$. If $q = \ell$, then surely $d(p_1^\ell, u) \neq d(p_1^\ell, v)$. Otherwise, by Claim 3(i),

$$d(p_1^\ell, v) \geq d(V(H_\ell), V(H_q)) = 4d > d + 1 \geq d(p_1^\ell, u), \text{ which concludes case 4.}$$

By the above case analysis, we infer that, for any $u, v \in V(G)$, there exists $w \in S$ such that $d(w, u) \neq d(w, v)$, that is, S is a resolving set of G. Since $|S| = k$, it follows that (G, k) is a YES-instance for METRIC DIMENSION.

Conversely, assume that (G, k) is a YES-instance for METRIC DIMENSION and let S be a resolving set of size at most k. By Observation 1, for any clause $c \in C$ and any variable $x \in X$ appearing in c,

$$|S \cap \{p_1^c, p_2^c\}| \geq 1, \ |S \cap \{p_1^{\overline{c}}, p_2^{\overline{c}}\}| \geq 1, \ |S \cap \{t_1^c, t_2^c\}| \geq 1, \text{ and } |S \cap \{t_1^{\overline{c}}, t_2^{\overline{c}}\}| \geq 1. \tag{1}$$

$$|S \cap \{t_1^{x,c}, t_2^{x,c}\}| \geq 1 \text{ and } |S \cap \{t_1^{x,\overline{c}}, t_2^{x,\overline{c}}\}| \geq 1. \tag{2}$$

$$|S \cap \{t_1, t_2\}| \geq 1. \tag{3}$$

Consider now a variable x. Since any path from a vertex in $V(G) \setminus V(G_x)$ to a vertex in $\{v_i^x, w_i^x \mid i \in [d]\}$ contains u_1^x or u_2^x, and, for any $i \in [d]$ and $u \in \{u_1^x, u_2^x\}$, $d(u, v_i^x) = d(u, w_i^x)$, no vertex in $V(G) \setminus \{v_i^x, w_i^x \mid i \in [d]\}$ can resolve v_i^x and w_i^x for any $i \in [d]$. It follows that

$$|S \cap \{v_i^x, w_i^x \mid i \in [d]\}| \geq 1. \tag{4}$$

Now, note that S has size at most $k = |X| + 10|C| + 1$, and so, equality must in fact hold in every inequality of Equations (1)–(4). W.l.o.g., let us assume that $t_1 \in S$ and that, for every clause $c \in C$ and variable $x \in X$ appearing in c, we have that $p_1^c, p_1^{\overline{c}}, t_1^c, t_1^{\overline{c}}, t_1^{x,c}, t_1^{x,\overline{c}} \in S$.

For every variable $x \in X$, assume, w.l.o.g., that $S \cap \{v_i^x, w_i^x \mid i \in [d]\} = S \cap \{v_i^x | \; i \in [d]\}$, and let $i_x \in [d]$ be the index of the vertex in $S \cap \{v_i^x \mid i \in [d]\}$. We contend that the assignment which sets each variable x to i_x satisfies every clause in C. Indeed, consider a clause $c = (x \leq a_x, y \leq a_y, z \leq a_z)$. We first aim to show that, for every $w \in S \setminus \{V(G_x) \cup V(G_y) \cup V(G_z)\}$ and $\ell \in \{c, \overline{c}\}$, $d(w, \ell) = d(w, p_2^\ell)$. Note that it suffices to show that any shortest path from $w \in S \setminus \{V(G_x) \cup V(G_y) \cup V(G_z)\}$ to $\ell \in \{c, \overline{c}\}$ contains v^ℓ, as then $d(w, \ell) = d(w, v^\ell) + 1 = d(w, p_2^\ell)$. Now, if $w \in V(G_t)$ for some $t \in \{c', \overline{c'} \mid c' \in C\}$ different from ℓ, then this readily follows from Claim 3(i); and if $w \in V(G_t)$ for some $t \in X \setminus \{x, y, z\}$, then this readily follows from Claim 3(iv). If $w = t_1^{r,q}$ for some $r \in X$ and $q \in \{c', \overline{c'} \mid c' \in C\}$, then $d(t_1^{r,q}, \ell) = d(t_1^{r,q}, v^q) + d(v_q, \ell)$, and so, by Claim 3(i), any path from w to ℓ contains v^ℓ. Finally, if $w \in \{t_1^{c'}, t_1^{\overline{c'}} \mid c' \in C\}$, then clearly any shortest path from w to ℓ contains v^ℓ.

Since S is a resolving set, it follows that, for every clause $c \in C$, there exist $t, f \in \{x, y, z\}$ such that $d(v_{i_t}^t, c) \neq d(v_{i_t}^t, p_2^c)$ and $d(v_{i_f}^f, \overline{c}) \neq d(v_{i_f}^f, p_2^{\overline{c}})$. Now, by Claim 4(i) and (ii), if $i_t > a_t$, then $d(v_{i_t}^t, c) = 5d + 1 + a_t - i_t = d(v_{i_t}^t, v^c) + 1 = d(v_{i_t}^t, p_2^c)$, a contradiction to our assumption. Therefore, $i_t \leq a_t$. Similarly, if $i_f \leq a_f$, then by Claim 5(i) and (ii), $d(v_{i_f}^f, \overline{c}) = 5d + 1 + i_f - a_f = d(v_{i_f}^f, v^{\overline{c}}) + 1 = d(v_{i_f}^f, p_2^{\overline{c}})$, a contradiction to our assumption. Therefore, $i_f > a_f$, and so, the assignment constructed indeed satisfies every clause in C. ◀

3 The Vertex Cover Number and the Distance to clique

In this section, we prove that Metric Dimension parameterized by either the vertex cover number or the distance to clique does not admit a polynomial kernel unless $\mathsf{NP} \subseteq \mathsf{coNP}/poly$. Both reductions are similar ones from the SAT problem, in which we are given a conjunctive normal form (CNF) formula ϕ on n variables and m clauses, and we are asked whether there exists an assignment of either true or false to each of the variables, such that ϕ is true (satisfied). SAT is known to not admit a polynomial kernel unless $\mathsf{NP} \subseteq \mathsf{coNP}/poly$ [16]. We first prove that Metric Dimension parameterized by the vertex cover number does not admit a polynomial kernel unless $\mathsf{NP} \subseteq \mathsf{coNP}/poly$, with the same result for distance to clique to follow after from a small modification to this reduction.

▶ **Theorem 6.** Metric Dimension *parameterized by the vertex cover number does not admit a polynomial kernel unless* $\mathsf{NP} \subseteq \mathsf{coNP}/poly$.

Proof. ($\star$) By a reduction from SAT, we prove that Metric Dimension parameterized by the vertex cover number does not admit a polynomial kernel unless $\mathsf{NP} \subseteq \mathsf{coNP}/poly$. Let ϕ be an instance of SAT, *i.e.*, a SAT formula on n variables $x_1, \ldots, x_n$ and m clauses $C_1, \ldots, C_m$. Since any SAT formula on n variables trivially has at most $3^n - 1$ unique clauses, we may assume that $m \leq 3^n - 1$.

From ϕ, we construct an instance (G, k) of Metric Dimension as follows. For each $i \in [n]$, construct a cycle $(t_i a_i^1 b_i^1 f_i b_i^2 a_i^2 t_i)$ on 6 vertices, and let $I_i := \{a_i^1, a_i^2, b_i^1, b_i^2\}$. Construct a path $g^1 g g^2$ on 3 vertices and, for each $i \in [n]$, make both f_i and t_i adjacent to g. For each $j \in [m]$, add a pair of vertices c_j^1 and c_j^2, and let $C_j := \{c_j^1, c_j^2\}$. For each $j \in [m]$, make c_j^2 adjacent to both f_i and t_i for each $i \in [n]$. For each $j \in [m]$ and each $i \in [n]$, if $x_i = \texttt{True}$ does not satisfy the clause $\mathcal{C}_j$ in ϕ, then make c_j^1 adjacent to t_i, and if $x_i = \texttt{False}$ does not satisfy $\mathcal{C}_j$, then make c_j^1 adjacent to f_i. Let $\alpha = \lceil n \cdot \log_2 3 \rceil$, and, for each $\ell \in [\alpha]$, construct a path $z_\ell^1 z_\ell z_\ell^2$ on 3 vertices. For each $j \in [m]$, consider the binary representation $\text{bin}(j)$ of j, and connect both c_j^1 and c_j^2 with z_ℓ if $\text{bin}(j)[\ell] = 1$, where $[\ell]$ is the ℓ^{th} bit of j in its binary representation from right to left. Finally, construct a clique on the vertices $z_1, \ldots, z_\alpha, g$. This completes the construction of G (see Figure 4).

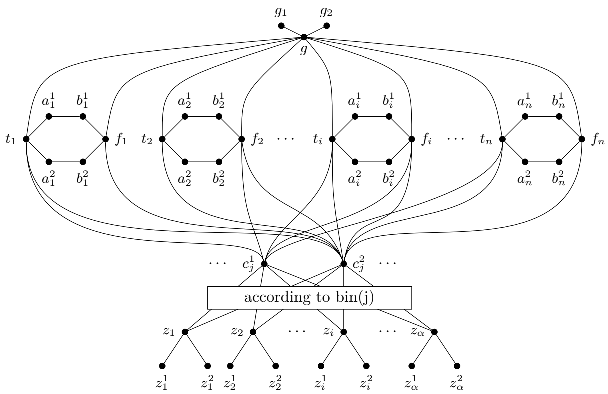

Figure 4 Illustration of the graph G constructed in the proof of Theorem 6. The vertices $z_1, \ldots, z_\alpha, g$ are in a clique that is not drawn. In this particular case, ϕ has a clause $(\overline{x_1} \vee x_2 \vee \overline{x_n})$.

To simplify notation for the proof, let $I := I_1 \cup \cdots \cup I_n$. Set $k = n + \alpha + 1$. Note that the vertex cover number of G is at most $4n + \alpha + 1$ since $\{g\} \cup \{a_i^1, a_i^2, t_i, f_i, z_\ell \mid i \in [n], \ell \in [\alpha]\}$ is a vertex cover of G. We now show that the instance ϕ is satisfiable if and only if (G, k) is a Yes-instance for Metric Dimension. We just sketch the proof from here. To prove that if ϕ is satisfiable, then (G, k) is a Yes-instance for Metric Dimension, we build a resolving set R of G of size k as follows. First, put the vertices of $\{g_1, z_1^1, \ldots, z_\alpha^1\}$ in R. Then, for all $i \in [n]$, if, according to the satisfying truth value assignment of ϕ, $x_i = \texttt{True}$ ($x_i = \texttt{False}$, resp.), then add a_i^1 (b_i^1, resp.) to R. Clearly, $|R| = k$, and it is not difficult to check that R is a resolving set of G.

Now, we prove that if (G, k) is a Yes-instance for Metric Dimension, then ϕ is satisfiable. For any resolving set R of G, one can show that since $|R| \leq n + \alpha + 1$, then $|R \cap \{g^1, g^2\}| = 1$, $|R \cap \{z_\ell^1, z_\ell^2\}| = 1$ for all $\ell \in [\alpha]$, and $|R \cap I_i| = 1$ for all $i \in [n]$. W.l.o.g., assume that $\{g^1, z_1^1, \ldots, z_\alpha^1\} \subset R$. Consider $j \in [m]$. It can be shown that no vertex in $\{g^1, z_1^1, \ldots, z_\alpha^1\}$ can resolve the two vertices of C_j, and thus, there must exist $w \in R \cap I$ such that $d(w, c_j^1) \neq d(w, c_j^2)$. Since for every $i \in [n]$ such that x_i does not appear in the clause C_j, $d(u, c_j^1) = d(u, c_j^2)$ for every $u \in I_i$, there must exist $i \in [n]$ such that x_i appears in the clause C_j and $w \in R \cap I_i$. In particular, c_j^1 must be non-adjacent to one of t_i and f_i. Now, if c_j^1 is non-adjacent to t_i, then c_j^1 is adjacent to f_i, and so, $w \in \{a_i^1, a_i^2\}$, as otherwise $d(w, c_j^1) = d(w, c_j^2)$. Symmetrically, if c_j^1 is non-adjacent to f_i, then c_j^1 is adjacent to t_i, and so, $w \in \{b_i^1, b_i^2\}$, as otherwise $d(w, c_j^1) = d(w, c_j^2)$. So, the truth assignment obtained by setting a variable x_i to True if $R \cap I_i \subseteq \{a_i^1, a_i^2\}$, and to False otherwise, satisfies ϕ. ◀

By making the vertices of $\{C_j \mid j \in [m]\}$ into a clique in the construction of G in the proof of Theorem 6, observe that the distance to clique of the resulting graph is at most $6n + 3\alpha + 3$, and that none of the distances described in the proof change. Then, from the proof of Theorem 6 for this modified G, we obtain the following:

▶ **Theorem 7.** Metric Dimension *parameterized by the distance to clique does not admit a polynomial kernel unless* NP $\subseteq$ coNP/*poly.*

4 The Distance to Cluster and the Distance to co-cluster

In this section, we prove that Metric Dimension is FPT parameterized by either the distance to cluster or the distance to co-cluster. In fact, we show that the problem admits an exponential kernel parameterized by the distance to cluster (or co-cluster). Since the main ideas for these two parameters are the same, we focus on the distance to cluster parameter. Applying Reduction Rule 2 for false twins (instead of true twins) and defining equivalence classes over the independent sets (instead of cliques) for Reduction Rule 3, we get the similar result for the distance to co-cluster. Recall that, for a graph G, the *distance to cluster* of G is the minimum number of vertices of G that need to be deleted so that the resulting graph is a *cluster graph*, *i.e.*, a disjoint union of cliques.

▶ **Theorem 8.** Metric Dimension *is* FPT *parameterized by the distance to cluster.*

Proof. Let (G, k) be an instance of Metric Dimension and let $X \subseteq V(G)$ be such that $G - X$ is a disjoint union of cliques. To obtain a kernel for the problem, we present a set of reduction rules. The safeness of the following reduction rule is trivial.

▶ **Reduction Rule 1.** *If $V(G) \neq \emptyset$ and $k \leq 0$, then return a trivial* No*-instance.*

▶ **Reduction Rule 2.** *If there exist $u, v, w \in V(G)$ such that u, v, w are true (or false) twins, then remove u from G and decrease k by one.*

▷ Claim 9 (⋆). Reduction Rule 2 is safe.

We assume, henceforth, that Reduction Rule 2 has been exhaustively applied to (G, k). This implies, in particular, that for every clique C of $G - X$, there are at most two vertices in C with the same neighborhood in X. Since the number of distinct neighborhoods in X is at most $2^{|X|}$, each clique in $G - X$ has order at most $2^{|X|+1}$. We now aim to bound the number of cliques in $G - X$. To this end, we define a notion of *equivalence classes* over the set of cliques of $G - X$. It will easily be seen that the number of equivalence classes is at most $2^{2^{|X|+1}}$. The number of cliques in each equivalence class will then be bounded by using Reduction Rule 3.

For every clique C of $G - X$, the *signature* $\texttt{sign}(C)$ of C is the *multiset* containing the neighborhoods in X of each vertex of C, that is, $\texttt{sign}(C) = \{N(u) \cap X : u \in C\}$. For any two cliques C_1, C_2 of $G - X$, we say that C_1 and C_2 are *identical*, which we denote by $C_1 \sim C_2$, if and only if $\texttt{sign}(C_1) = \texttt{sign}(C_2)$. It is not difficult to see that $\sim$ is in fact an equivalence relation with at most $2^{2^{|X|+1}}$ equivalence classes: indeed, since the number of distinct neighborhoods in X is at most $2^{|X|}$, and at most two vertices of each clique have the same neighborhood in X, the number of distinct signatures is at most $2^{2^{|X|+1}}$. Consider now an equivalence class $\mathcal{C}$ of $\sim$. Note that since the signature of a clique is a multiset, the number of vertices in each $C \in \mathcal{C}$ is equal to $|\texttt{sign}(C)|$. For any $C_1, C_2 \in \mathcal{C}$, we say that two vertices $u \in C_1$ and $v \in C_2$ are *clones* if $N(u) \cap X = N(v) \cap X$ (in particular, if $C_1 = C_2$ and $u \neq v$, then u, v are true twins). For any $C_1, C_2 \in \mathcal{C}$ and any $u \in C_1$, we denote by $c(u, C_2)$

the set of clones of u in C_2 (note that $|c(u, C_2)| \leq 2$). Now observe that, for any two cliques $C_1, C_2 \in \mathcal{C}$, the number of pairs of true twins in C_1 and C_2 is the same: we let $t(\mathcal{C})$ be the number of pairs of true twins in each clique of $\mathcal{C}$. We highlight that there are exactly $2t(\mathcal{C})$ vertices in each clique of $\mathcal{C}$ that have true twins. The following claim for clones is the analog of Observation 1 for twins.

▷ Claim 10 ($\star$). Let C_1 and C_2 be two cliques of an equivalence class $\mathcal{C}$ of $\sim$. Let $u \in C_1$ and $v \in C_2$ be clones. Then, for any $w \in V(G) \setminus (V(C_1) \cup V(C_2))$, $d(u, w) = d(v, w)$, and so, for any resolving set S of G, $S \cap (V(C_1) \cup V(C_2)) \neq \emptyset$.

It follows from the above claim that, for any equivalence class $\mathcal{C}$ of $\sim$ and any resolving set S, S contains at least $|\mathcal{C}| - 1$ vertices in $V(\mathcal{C}) := \bigcup_{C \in \mathcal{C}} V(C)$. We now present an upper bound on the size of $S \cap V(\mathcal{C})$ when $|\mathcal{C}| \geq |X| + 2$.

▷ Claim 11 ($\star$). For every equivalence class $\mathcal{C}$ of $\sim$, if $|\mathcal{C}| \geq |X| + 2$, then, for any minimum resolving set S of G, $|S \cap V(\mathcal{C})| \leq |X| + |\mathcal{C}| \cdot \max\{1, t(\mathcal{C})\}$.

Let $\mathcal{C}$ be an equivalence class of $\sim$, and let S be a resolving set of G. For every $i \geq 0$, we denote by $\mathcal{C}^S_{=i}$ ($\mathcal{C}^S_{\geq i}$, resp.) the set of cliques $C \in \mathcal{C}$ such that $|S \cap V(C)| = i$ ($|S \cap V(C)| \geq i$).

▷ Claim 12 ($\star$). Let $\mathcal{C}$ be an equivalence class of $\sim$ such that $|\mathcal{C}| \geq |X| + 2$. Then, for any minimum resolving set S of G, the following hold:

(i) if $t(\mathcal{C}) = 0$, then $|\mathcal{C}^S_{=0}| \leq 1$ and $|\mathcal{C}^S_{\geq 2}| \leq |X| + 1$;
(ii) if $t(\mathcal{C}) \neq 0$, then $|\mathcal{C}^S_{\geq t(\mathcal{C})+1}| \leq |X| + 1$.

The above claim states that if some equivalence class $\mathcal{C}$ of $\sim$ contains at least $|X| + 3$ cliques, then, for any minimum resolving set S of G, if $t(\mathcal{C}) = 0$, then $\mathcal{C}^S_{=1} \neq \emptyset$, and otherwise, $\mathcal{C}^S_{=t(\mathcal{C})} \neq \emptyset$. The following reduction rule is based on this claim.

▶ **Reduction Rule 3.** *If there exists an equivalence class $\mathcal{C}$ of $\sim$ such that $|\mathcal{C}| \geq 2^{|X|+2} + |X| + 2$, then remove a clique $C \in \mathcal{C}$ from G and reduce k by* $\max\{1, t(\mathcal{C})\}$.

▷ Claim 13 ($\star$). Reduction Rule 3 is safe.

Now observe that once Reduction Rule 3 has been exhaustively applied to (G, k), each equivalence class of $\sim$ contains at most $2^{|X|+2} + |X| + 1$ cliques. Since there are at most $2^{2^{|X|+1}}$ equivalence classes and each clique of $G - X$ has size at most $2^{|X|+1}$, we conclude that G contains at most $2^{2^{|X|+1}} \cdot (2^{|X|+2} + |X| + 1) \cdot 2^{|X|+1} + |X|$ vertices. ◀

5 Conclusion

As the Metric Dimension problem is W[2]-hard when parameterized by the solution size [23], the next natural step is to understand its parameterized complexity under structural parameterizations. We continued this line of research, following in the steps of [3, 13, 20], and more recently [7, 29]. Our most technical result is a proof that the Metric Dimension problem is W[1]-hard when parameterized by the feedback vertex set number of the graph. We thereby improved the result by Bonnet and Purohit [7] that states the problem is W[1]-hard when parameterized by the treewidth, and answered an open question in [23]. It is easy to see that the problem admits an FPT algorithm when parameterized by the larger parameter, the vertex cover number of the graph. On the positive side, we proved that the problem admits FPT algorithms when parameterized by the distance to cluster and the distance to co-cluster, which are smaller parameters than the vertex cover number.

Although this work advances the understanding of structural parameterizations of Metric Dimension, it falls short of completing the picture (see Figure 1). We find it hard to extend the positive results to the parameters like the minimum clique cover, the distance to disjoint paths, feedback edge set, and the bandwidth. It would be interesting to find FPT algorithms or prove that such algorithms are highly unlikely to exist for these parameters. The FPT algorithm parameterized by the treedepth in [20] relies on the meta-result. Is it possible to get an FPT algorithm whose running time is a single or double exponent in the treedepth? It would also be interesting to investigate the problem parameterized by the distance to cograph. Recall that the problem is polynomial-time solvable in cographs [14].

Bonnet and Purohit [7] conjectured that the problem is W[1]-hard even when parameterized by the treewidth plus the solution size. Towards resolving this conjecture, an interesting question would be to investigate whether the problem admits an FPT algorithm when parameterized by the feedback vertex set number plus the solution size. Note that even an XP algorithm parameterized by the feedback vertex set number is not apparent.

References

1. L. Babai. On the complexity of canonical labelling of strongly regular graphs. *SIAM J. Comput.*, 9(1):212–216, 1980.
2. Z. Beerliova, F. Eberhard, T. Erlebach, A. Hall, M. Hoffman, M. Mihalák, and L. S. Ram. Network discovery and verification. *IEEE J. Sel. Area Comm.*, 24(12):2168–2181, 2006.
3. R. Belmonte, F. V. Fomin, P. A. Golovach, and M. S. Ramanujan. Metric dimension of bounded tree-length graphs. *SIAM J. Discrete Math.*, 31(2):1217–1243, 2017.
4. Y. Ben-Haim, S. Gravier, A. Lobstein, and J. Moncel. Adaptive identification in graphs. *J. Comb. Theory, Ser. A*, 115(7):1114–1126, 2008.
5. J. Bensmail, D. Mazauric, F. Mc Inerney, N. Nisse, and S. Pérennes. Sequential metric dimension. *Algorithmica*, 82(10):2867–2901, 2020.
6. J. Bensmail, F. Mc Inerney, and N. Nisse. Metric dimension: from graphs to oriented graphs. *Discrete Applied Mathematics*, in press. `doi:10.1016/j.dam.2020.09.013`.
7. E. Bonnet and N. Purohit. Metric dimension parameterized by treewidth. *Algorithmica*, 83:2606–2633, 2021.
8. B. Bosek, P. Gordinowicz, J. Grytczuk, N. Nisse, J. Sokól, and M. Sleszynska-Nowak. Centroidal localization game. *Electronic Journal of Combinatorics*, 25(4):P4.62, 2018.
9. B. Bosek, P. Gordinowicz, J. Grytczuk, N. Nisse, J. Sokól, and M. Sleszynska-Nowak. Localization game on geometric and planar graphs. *Discrete Applied Mathematics*, 251:30–39, 2018.
10. K. Bringmann, D. Hermelin, M. Mnich, and E. J. van Leeuwen. Parameterized Complexity Dichotomy for Steiner Multicut. In *32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015)*, volume 30 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 157–170, 2015.
11. G. Chartrand, L. Eroh, M. Johnson, and O. Oellermann. Resolvability in graphs and the metric dimension of a graph. *Discrete Applied Mathematics*, 105(1-3):99–113, 2000.
12. J. Díaz, O. Pottonen, M. J. Serna, and E. J. van Leeuwen. Complexity of metric dimension on planar graphs. *J. Comput. Syst. Sci.*, 83(1):132–158, 2017.
13. D. Eppstein. Metric dimension parameterized by max leaf number. *Journal of Graph Algorithms and Applications*, 19(1):313–323, 2015.
14. L. Epstein, A. Levin, and G. J. Woeginger. The (weighted) metric dimension of graphs: Hard and easy cases. *Algorithmica*, 72(4):1130–1171, 2015.
15. A. Estrada-Moreno, J. A. Rodriguez-Velázquez, and I. G. Yero. The k-metric dimension of a graph. *Applied Mathematics and Information Sciences*, 9(6):2829–2840, 2015.

16 L. Fortnow and R. Santhanam. Infeasibility of instance compression and succinct PCPs for NP. *J. Comput. Syst. Sci.*, 77(1):91–106, 2011.

17 F. Foucaud, R. Klasing, and P. J. Slater. Centroidal bases in graphs. *Networks*, 64(2):96–108, 2014.

18 F. Foucaud, G. B. Mertzios, R. Naserasr, A. Parreau, and P. Valicov. Identification, location-domination and metric dimension on interval and permutation graphs. II. algorithms and complexity. *Algorithmica*, 78(3):914–944, 2017.

19 M. R. Garey and D. S. Johnson. *Computers and Intractability – A guide to NP-completeness*. W.H. Freeman and Company, 1979.

20 T. Gima, T. Hanaka, M. Kiyomi, Y. Kobayashi, and Y. Otachi. Exploring the gap between treedepth and vertex cover through vertex integrity. *Theoretical Computer Science*, 2022.

21 G. Z. Gutin, M. S. Ramanujan, F. Reidl, and M. Wahlström. Alternative parameterizations of metric dimension. *Theoretical Computer Science*, 806:133–143, 2020.

22 F. Harary and R. A. Melter. On the metric dimension of a graph. *Ars Combinatoria*, 2:191–195, 1976.

23 S. Hartung and A. Nichterlein. On the parameterized and approximation hardness of metric dimension. In *Proceedings of the 28th Conference on Computational Complexity, CCC*, pages 266–276. IEEE Computer Society, 2013.

24 J. Haslegrave, R. A. B. Johnson, and S. Koch. Locating a robber with multiple probes. *Discrete Math.*, 341(1):184–193, 2018.

25 S. Hoffmann, A. Elterman, and E. Wanke. A linear time algorithm for metric dimension of cactus block graphs. *Algorithmica*, 72(4):1130–1171, 2015.

26 M. A. Johnson. Structure-activity maps for visualizing the graph variables arising in drug design. *J. Biopharm. Statist.*, 3:203–236, 1993.

27 M. G. Karpovsky, K. Chakrabarty, and L. B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Trans. Information Theory*, 44(2):599–611, 1998.

28 D. Kuziak and I. G. Yero. Metric dimension related parameters in graphs: A survey on combinatorial, computational and applied results. *arXiv*, 2021. `arXiv:2107.04877`.

29 S. Li and M. Pilipczuk. Hardness of metric dimension in graphs of constant treewidth. In *16th International Symposium on Parameterized and Exact Computation (IPEC 2021)*, volume 214 of *LIPIcs*, pages 24:1–24:13, 2021.

30 R. A. Melter and I. Tomescu. Metric bases in digital geometry. *Comput. Vision Graphics Image Process.*, 25:113–121, 1984.

31 P. J. Slater. Leaves of trees. In *Proceedings of the Sixth Southeastern Conference on Combinatorics, Graph Theory, and Computing*, pages 549–559. Congressus Numerantium, No. XIV. Utilitas Mathematica, 1975.

32 P. J. Slater. Domination and location in acyclic graphs. *Networks*, 17(1):55–64, 1987.

33 R. C. Tillquist, R. M. Frongillo, and M. E. Lladser. Getting the lay of the land in discrete space: A survey of metric dimension and its applications. *arXiv*, 2021. `arXiv:2104.07201`.

34 R. C. Tillquist, R. M. Frongillo, and M. E. Lladser. Truncated metric dimension for finite graphs. *arXiv*, 2021. `arXiv:2106.14314`.

35 R. C. Tillquist and M. E. Lladser. Low-dimensional representation of genomic sequences. *Journal of Mathematical Biology*, 79:1–29, 2019.

36 R. Ungrangsi, A. Trachtenberg, and D. Starobinski. An implementation of indoor location detection systems based on identifying codes. In *Proc. INTELLCOM 2004*, volume 3283 of *LNCS*, pages 175–189, 2004.