On the metric dimension of Cartesian powers of a graph

Zilin Jiang [a,1], Nikita Polyanskii [b,2]

[a] *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *CDISE, Skolkovo Institute of Science and Technology, Moscow, 121205, Russian Federation*

A R T I C L E I N F O

A B S T R A C T

A set of vertices S resolves a graph if every vertex is uniquely determined by its vector of distances to the vertices in S. The metric dimension of a graph is the minimum cardinality of a resolving set of the graph. Fix a connected graph G on $q \geq 2$ vertices, and let M be the distance matrix of G. We prove that if there exists $w \in \mathbb{Z}^q$ such that $\sum_i w_i = 0$ and the vector Mw, after sorting its coordinates, is an arithmetic progression with nonzero common difference, then the metric dimension of the Cartesian product of n copies of G is $(2 + o(1))n/\log_q n$. In the special case that G is a complete graph, our results close the gap between the lower bound attributed to Erdős and Rényi and the upper bounds developed subsequently by Lindström, Chvátal, Kabatianski, Lebedev and Thorpe.

E-mail addresses: zilinj@mit.edu (Z. Jiang), nikita.polyansky@gmail.com (N. Polyanskii).

1. Introduction

A set of vertices S *resolves* a graph if every vertex is uniquely determined by its vector of distances to the vertices in S. The *metric dimension* of a graph is the minimum cardinality of a resolving set of the graph. The *Cartesian product* of graphs $G_1, \ldots, G_n$ is the graph with vertex set $V(G_1) \times \cdots \times V(G_n)$ such that $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ are adjacent whenever there exists $j \in [n]$ such that $u_i = v_i$ for all $i \neq j$ and u_j is adjacent to v_j in G_j.

For a graph G and $n \in \mathbb{N}$, denote by $G^{\square n}$ the Cartesian product of n copies of G, and by $m(G, n)$ the metric dimension of $G^{\square n}$. This paper undertakes the study of the asymptotic behavior of $m(G, n)$ when the connected graph G is fixed and n tends to infinity, especially when G is a complete graph on q vertices, which we denote by K_q. In this context, the definition of a resolving set can be rephrased in the following way. Denote the distance between vertices u, v in G by $d(u, v)$. Given a subset S of V^n, define $d_S \colon V^n \to \mathbb{N}^S$ by $(d_S(v))_s = d(v_1, s_1) + \cdots + d(v_n, s_n)$ for every $v = (v_1, \ldots, v_n) \in V^n$ and $s = (s_1, \ldots, s_n) \in S$. The set of vertices S is a resolving set of $G^{\square n}$ if and only if d_S is an injection.

The concept of resolving set and that of metric dimension date back to the 1950s – they were defined by Bluementhal [2] in the context of metric space. These notions were introduced to graph theory by Harary and Melter [8] and Slater [21] in the 1970s.

Under the guise of a coin weighing problem, the metric dimension of a hypercube was first studied by Erdős and Rényi. The coin weighing problem, posed by Söderberg and Shapiro [22], assumes n coins of weight a or b, where a and b are known, and an accurate scale. Söderberg and Shapiro asked the question of how many weighings are needed to determine which of n coins are of weight a and which of weight b if the numbers of each are not known. The variant of the problem, where the family of weighings has to be given in advance, is connected to the metric dimension of the hypercube $K_2^{\square n}$. It was observed that the minimum number of weighings differs from $m(K_2, n)$ by at most 1 (see [20, Section 1]). A lower bound on the number of weighings by Erdős and Rényi [7] and an upper bound by Lindström [12] and independently by Cantor and Mills [5] imply that $m(K_2, n) = (2 + o(1))n / \log_2 n$.

The metric dimension of the Hamming graph $K_q^{\square n}$ is also connected to the Mastermind game. Mastermind is a deductive game for two players, the codemaker and the codebreaker.[3] In this game, the codemaker conceals a vector $x = (x_1, \ldots, x_n) \in [q]^n$, and the codebreaker, who knows both q and n, tries to identify x by asking a number of questions, which are answered by the codemaker. Each question is a vector $y = (y_1, \ldots, y_n) \in [q]^n$; each answer consists of a pair of numbers $a(x, y)$, the number of subscripts i such that $x_i = y_i$, and $b(x, y)$, the maximum number of $a(x, \tilde{y})$ with $\tilde{y}$ running through all the permutations of y. Knuth [11] has shown that four questions suffice to determine x in

---

[3] In honor of Erdős, Chvátal [6] referred to the codemaker and the codebreaker as SF and PGOM. See [19, p. 41 and p. 70] for what SF and PGOM stand for.

the commercial version of the game where $n = 4$ and $q = 6$. Suppose for the time being that we remove the second number $b(x, y)$ from the answers given by the codemaker and we require that the questions from the codebreaker are sent all at once. In this version of Mastermind, the minimum number of questions required to determine x is exactly $m(K_q, n)$ (see [4, Section 6]).

Kabatianski, Lebedev and Thorpe [9] stated that a straightforward generalization of the lower bound on $m(K_2, n)$ by Erdős and Rényi [7] gives $m(K_q, n) \geq (2+o(1))n/\log_q n$. Kabatianski et al. also asserted that more precise calculations, based on the probabilistic method of Chvátal [6, Theorem 1], would show that $m(K_q, n) \leq (2 + o(1)) \log_q(1 + (q-1)q) \cdot n/\log_q n$. Very recently, these calculations were carried out by Kabatianski and Lebedev [10]. Moreover, they proved that $m(K_q, n) = (2+o(1))n/\log_q n$ for $q = 3, 4$, which was previously announced in [9, Theorem 1], and they conjectured that $m(K_q, n) = (2 + o(1))n/\log_q n$ for all $q \geq 2$. We emphasize that the asymptotic behavior is different when q varies and n is fixed. For example, Cáceres et al. [4, Theorem 6.1] showed that $m(K_q, 2) = \lfloor 2(2q - 1)/3 \rfloor$.

Motivated by the above applications, we establish an upper bound and a lower bound on $m(G, n)$ for every connected graph G in Section 2 and Section 3 respectively. For certain families of graphs, the lower bound and the upper bound are asymptotically equivalent. In particular, we show that $m(K_q, n) = (2 + o(1))n/\log_q n$ for all $q \geq 2$ in Section 4. We conclude with a generalization to integer matrices and some open problems in Section 5.

2. An upper bound on $m(G, n)$

We establish the following upper bound on $m(G, n)$.

Theorem 1. *Given a connected graph G on $q \geq 2$ vertices, let M be the distance matrix of G. For every $n \in \mathbb{N}$, the metric dimension $m(G, n)$ of $G^{\square n}$ is at most*

$$\left(2 + O\left(\frac{\log \log n}{\log n} \right) \right) \frac{n}{\log_r n},$$

where $r = r(G)$ is defined by

$$r(G) = \min \left\{ \frac{\max Mw - \min Mw}{\gcd_{i<j}((Mw)_i - (Mw)_j)} : w \in \mathbb{Z}^q, \ \sum_i w_i = 0, \right.$$

$$\left. (Mw)_i \neq (Mw)_j \text{ for all } i \neq j \right\} + 1. \quad (1)$$

Remark 1. Alternatively, $r(G)$ is the shortest length of an arithmetic progression with nonzero common difference that contains Mw, after sorting its coordinates, as a subsequence for some $w \in \mathbb{Z}^q$ such that $\sum_i w_i = 0$. Clearly, $r(G) \geq q$. It is less clear

that $r(G) < \infty$. We claim that there exists $w \in \mathbb{Q}^q$ such that $\sum_i w_i = 0$ and $(Mw)_i \neq (Mw)_j$ for all $i \neq j$. Denote the ith row of M by M_i. For $i \neq j$, because $(M_i - M_j)_i + (M_i - M_j)_j = 0$, the equation $(M_i - M_j)w = 0$, or $(Mw)_i = (Mw)_j$, defines a subspace of $\mathbb{Q}^n$ different from $\{w \in \mathbb{Q}^n : \sum_i w_i = 0\}$. In other words, $(Mw)_i = (Mw)_j$ defines a 1-codimensional subspace of $\{w \in \mathbb{Q}^n : \sum_i w_i = 0\}$, and so $\{w \in \mathbb{Q}^n : \sum_i w_i = 0\} \setminus \cup_{i \neq j} \{w \in \mathbb{Q}^n : (Mw)_i = (Mw)_j\}$ is nonempty. Finally, we scale w properly so that it becomes a vector in $\mathbb{Z}^q$.

Our construction of a resolving set of $G^{\square n}$ is inspired by the upper bound for the coin weighing problem by Lindström [13]. Among various constructions such as the recursive construction by Cantor and Mills [5] and the construction by Bshouty [3] based on Fourier transform, we find the one using the theory of Möbius functions by Lindström [15] best suits our needs.

We recall the basics of Möbius functions. Let $(P, \prec)$ be a locally finite partially ordered set. The Möbius function $\mu \colon P \times P \to \mathbb{Z}$ can be defined inductively by the following relation:

$$\mu(x, y) = \begin{cases} 1 & \text{if } x = y, \\ -\sum_{x \preceq z \prec y} \mu(x, z) & \text{for } x \prec y, \\ 0 & \text{otherwise.} \end{cases}$$

The classical Möbius function in number theory is essentially the Möbius function of the set of positive integers partially ordered by divisibility. For our purpose, we first consider binary representation of natural numbers, and we instead partially order $\mathbb{N} = \{0, 1, \dots\}$ in the following way: $x \preceq y$ if and only if $x = x \wedge y$, where $\wedge$ is the bitwise AND operation.[4] The Möbius function is thus

$$\mu(x, y) = (-1)^{n(x) - n(y)}, \text{if } x \preceq y,$$

where $n(x)$ is the number of ones in the binary representation of x. With the binary operator $\wedge$, the partially ordered set $(\mathbb{N}, \prec)$ is indeed a *meet-semilattice* – a partially ordered set in which any pair of elements has the greatest lower bound. We need the following identity for our meet-semilattice.

Lemma 2 *(Lemma of Lindström [14]). Let $(P, \prec, \wedge)$ be a locally finite meet-semilattice with Möbius function $\mu(x, y)$. Let $a, b \in P$ and $b \npreceq a$. Let $f(x)$ be defined for all $x \preceq a \wedge b$ with values in a commutative ring with unity. Then we have*

$$\sum_{x \preceq b} f(x \wedge a)\mu(x, b) = 0.$$

---

[4] A bitwise AND takes two binary representations and perform the logical AND operation on each pair of the corresponding bits. For example, $6 \wedge 11 = 0110_2 \wedge 1011_2 = 0010_2 = 2$.

The last ingredient is the following estimation on the partial sum of $n(\cdot)$.

Theorem 3 *(Theorem 1 of Bellman and Shapiro [1]).*

$$\sum_{i=0}^{x} n(i) = \tfrac{1}{2}x \log_2 x + O(x \log \log x) \ \text{as } x \to \infty.$$

We now construct a resolving set for Theorem 1 using the Möbius function of $(\mathbb{N}, \prec, \wedge)$.

Proof of Theorem 1. Let $w \in \mathbb{Z}^q$ be such that $\sum_i w_i = 0$ and the coordinates of Mw are distinct integers such that

$$r = r(G) > \frac{\max Mw - \min Mw}{\gcd_{i<j}((Mw)_i - (Mw)_j)}. \tag{2}$$

Set $|w|_1 := \sum_i |w_i|$. For each $j \in \mathbb{N}$, let $b(j)$ be the largest integer such that

$$r^{b(j)} \cdot |w|_1 \le 2^{n(j)}, \tag{3}$$

that is, $b(j) := \lfloor n(j) \log_r 2 - \log_r |w|_1 \rfloor$.

Let J be the set of the first n elements of $\{(j,k) : j \in \mathbb{N}, 0 \le k \le b(j)\}$ under the lexicographical order. We label the n copies of G in $G^{\square n}$ by J, namely each vertex of $G^{\square n}$ is an element of V^J, where $V = \{v_1, \ldots, v_q\}$ is the vertex set of G. Set $m := \max\{j : (j,k) \in J\}$.

Our resolving set will be described by a matrix S whose rows and columns are indexed by $\{0, 1, \ldots, m\}$ and J respectively with entries from V. Note that each row of S is an element of V^J, thus can be seen as a vertex of $G^{\square n}$. For $i \in \{0, 1, \ldots, m\}$ and $(j,k) \in J$, we denote the entry of S on row i and column (j,k) by $S(i,j,k) \in V$.

We claim that a matrix S can be chosen to satisfy the following properties.

$$\sum_{i \preceq j} S(i,j,k)\mu(i,j) = r^k(w_1 v_1 + \cdots + w_q v_q), \quad \text{for all } (j,k) \in J; \tag{4a}$$

$$\sum_{i \preceq j} S(i,j',k)\mu(i,j) = 0, \quad \text{for all } (j',k) \in J \text{ and } j' < j \le m. \tag{4b}$$

We remark that (4a) and (4b) happen in the commutative ring $\mathbb{Z}[v_1, \ldots, v_q]$ with unity.

For example, when $G = K_3$, $q = 3$, we take $w = (-1 \ \ 0 \ \ 1)^T$ and $r = 3$. In Table 1, we supply the values of $S(i,j,k)$ for $(j,k) = (7,0), (7,1)$. The reader can verify (4a) in this case.

In general, pick arbitrary $(j,k) \in J$. On the left hand side of (4a), the summation consists of $2^{n(j)}$ terms, moreover $2^{n(j)-1}$ of them have $\mu(i,j) = +1$ (respectively -1). Since $r^k(w_1 + \cdots + w_q) = 0$ and $r^k(|w_1| + \cdots + |w_q|) = r^k |w|_1 \le r^{b(j)} |w|_1 \le 2^{n(j)}$ by (3), it is easy to assign one of $\{v_1, \ldots, v_q\}$ to $S(i,j,k)$ for all $i \preceq j$, possibly in many ways,

Table 1
Values of $S(i, j, k)$ for $(j, k) = (7, 0), (7, 1)$ for $G = K_3$.

i	0	1	2	3	4	5	6	7	8	9	
$(7, 0)$	v_3	v_3	v_3	v_2	v_2	v_1	v_1	v_1	v_3	v_3	$\ldots$
$(7, 1)$	v_1	v_3	v_3	v_2	v_2	v_1	v_1	v_3	v_1	v_3	
$\mu(i, j)$	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$	0	0	

to satisfy (4a). For $i \not\preceq j$, we take $S(i, j, k) = S(i \wedge j, j, k)$. For every $(j', k) \in J$ and $j' < j \le m$, as $j \not\preceq j'$, the left hand side of (4b) equals $\sum_{i \preceq j} S(i \wedge j', j', k)\mu(i, j) = 0$ by applying Lemma 2 to the function $f_{j',k}(i) = S(i, j', k)$.

To show that S resolves $G^{\Box n}$, it suffices to demonstrate that every $X : J \to V$ is uniquely determined by the vector

$$
D := \left(\sum_{(j,k) \in J} d(X(j, k), S(i, j, k)) \right)_{i=0}^{m}.
$$

Suppose this vector $D = (D_0, \ldots, D_m)$ is provided. We shall gradually uncover $\{X(j, k) : 0 \le k \le b(j)\}$ for $j = m, m - 1, \ldots, 0$. Assume that $\{X(j, k) : 0 \le k \le b(j)\}$ is known for every $j > j_0$. We extend the distance function $d : V \times V \to \mathbb{N}$ of G to the bilinear form

$$
d \left(\sum_{i=1}^{q} \alpha_i v_i, \sum_{i=1}^{q} \beta_i v_i \right) = \sum_{i=1}^{q} \sum_{j=1}^{q} \alpha_i \beta_j d(v_i, v_j),
$$

where $\alpha_1, \ldots, \alpha_q$ and $\beta_1, \ldots, \beta_q$ are in $\mathbb{Q}$. Observe that

$$
\begin{aligned}
\sum_{i \preceq j_0} D_i \mu(i, j_0) &= \sum_{i \preceq j_0} \left(\sum_{(j,k) \in J} d(X(j, k), S(i, j, k)) \right) \mu(i, j_0) \\
&= \sum_{(j,k) \in J} d \left(X(j, k), \sum_{i \preceq j_0} S(i, j, k)\mu(i, j_0) \right) \\
&\overset{(4b)}{=} \sum_{j=j_0}^{m} \sum_{k=0}^{b(j)} d \left(X(j, k), \sum_{i \preceq j_0} S(i, j, k)\mu(i, j_0) \right).
\end{aligned}
$$

Since both $(D_0, \ldots, D_m)$ and $\{X(j, k) : j_0 < j \le m, 0 \le k \le b(j)\}$ are known, we are able to determine

$$\sum_{k=0}^{b(j_0)} d\left(X(j_0,k), \sum_{i \preceq j_0} S(i,j_0,k)\mu(i,j_0)\right) \overset{(4a)}{=} \sum_{k=0}^{b(j_0)} d\left(X(j_0,k), r^k \sum_{i=1}^{q} w_i v_i\right)$$

$$= \sum_{k=0}^{b(j_0)} r^k \sum_{i=1}^{q} w_i d(X(j_0,k), v_i) = \sum_{k=0}^{b(j_0)} r^k \sum_{i=1}^{q} M_{X(j_0,k),v_i} w_i = \sum_{k=0}^{b(j_0)} r^k \cdot (Mw)_{X(j_0,k)}. \quad (5)$$

Let $g = \gcd_{i<j}((Mw)_i - (Mw)_j)$. We can thus deduce from (5) the value of

$$\sum_{k=0}^{b(j_0)} r^k \cdot \frac{1}{g}\left((Mw)_{X(j_0,k)} - \min Mw\right). \quad (6)$$

Notice that, according to our choice of w and (2), $\left(\frac{1}{g}((Mw)_i - \min Mw)\right)_{i=1}^{q}$ are distinct integers in $[0, r)$. The value of (6) uniquely decides $\{X(j_0,k) : 0 \le k \le b(j_0)\}$.

Finally, we estimate $m+1$, the cardinality of the resolving set. Our choice of m implies that m is the smallest integer such that $\sum_{j=0}^{m} \max\{b(j)+1,0\} \ge n$. For every $x \in \mathbb{N}$, by Theorem 3,

$$\sum_{j=0}^{x} \max\{b(j)+1,0\} > \sum_{j=0}^{x}(n(j)\log_r 2 - \log_r |w|_1) = \tfrac{1}{2}x \log_r x - O(x \log\log x). \quad (7)$$

One can check that $x = 2n/\log_r n + O(n \log\log n/\log^2 n)$ ensures the right hand side of (7) is $\ge n$. $\quad \square$

3. A lower bound on $m(G,n)$

A straightforward generalization of the lower bound on the coin weighing problem by Erdős and Rényi gives a lower bound on the metric dimension of $G^{\Box n}$ (see Moser [17] and Pippenger [18] for different proofs using the second moment method and the information-theoretic method).

Theorem 4. *Given a connected graph G on $q \ge 2$ vertices, for every $n \in \mathbb{N}$, the metric dimension $m(G,n)$ of $G^{\Box n}$ is at least*

$$\left(2 - O\left(\frac{\log\log n}{\log n}\right)\right)\frac{n}{\log_q n}.$$

Proof. Let $S \subset V^n$ be a resolving set of $G^{\Box n}$ of size $m = m(G,n)$, where V is the vertex set of G. We may assume without loss that $m = O(n)$. For every $s = (s_1, \ldots, s_n) \in S$, let $X_1, \ldots, X_n$ be independent random variables defined by $X_i = d(Y_i, s_i)$, where the independent random variables $Y_1, \ldots, Y_n$ are chosen uniformly at random from V, and define

$$A_s := \left\{ (v_1, \ldots, v_n) \in V^n : \left| \sum_{i=1}^{n} d(v_i, s_i) - \mathrm{E}\left[\sum_{i=1}^{n} X_i \right] \right| < \sqrt{n \ln n} \cdot D \right\},$$

where D is the diameter of the graph. Since each X_i is bounded by $[0, D]$, Hoeffding's inequality provides an upper bound on the cardinality of the complement of A_s:

$$\frac{|V^n \setminus A_s|}{|V^n|} = \Pr\left(\left| \sum_{i=1}^{n} d(Y_i, s_i) - \mathrm{E}\left[\sum_{i=1}^{n} X_i \right] \right| \geq \sqrt{n \ln n} \cdot D \right)$$

$$= \Pr\left(\left| \sum_{i=1}^{n} X_i - \mathrm{E}\left[\sum_{i=1}^{n} X_i \right] \right| \geq \sqrt{n \ln n} \cdot D \right)$$

$$\leq 2 \exp\left(-\frac{2\left(\sqrt{n \ln n} \cdot D \right)^2}{n \cdot D^2} \right) = \frac{2}{n^2}.$$

From the equivalent definition of a resolving set mentioned in Section 1, the function $d_S \colon V^n \to \mathbb{N}^S$, defined by $(d_S(v_1, \ldots, v_n))_s := d(v_1, s_1) + \cdots + d(v_n, s_n)$ for every $(v_1, \ldots, v_n) \in V^n$ and $s = (s_1, \ldots, s_n) \in S$, is injective. Since the image of $\cap_{s \in S} A_s$ under d_S is contained in a cube of side length $< 2\sqrt{n \ln n} \cdot D$ in $\mathbb{N}^S$, we obtain

$$\left(2\sqrt{n \ln n} \cdot D \right)^m \geq \left| \bigcap_{s \in S} A_s \right| \geq |V^n| - \sum_{s \in S} |V^n \setminus A_s|$$

$$\geq q^n \left(1 - \frac{2m}{n^2} \right) = q^n \left(1 - O\left(\frac{1}{n} \right) \right).$$

Taking logarithm gives

$$m \geq \frac{n \ln q - O\left(\frac{1}{n} \right)}{\frac{1}{2} \ln n + O\left(\log \log n \right)} = \frac{2n}{\log_q n} \cdot \frac{1 - O\left(\frac{1}{n^2} \right)}{1 + O\left(\frac{\log \log n}{\log n} \right)}$$

$$= \left(2 - O\left(\frac{\log \log n}{\log n} \right) \right) \frac{n}{\log_q n}. \qquad \square$$

4. Asymptotically tight cases

The bounds in Theorem 1 and Theorem 4 are asymptotically equivalent if and only if $r(G)$ defined by (1) equals q. We characterize the equality case.

Lemma 5. *Given a connected graph G on $q \geq 2$ vertices, let M be the distance matrix of G. The following statements are equivalent.*

1. *The technical parameter $r(G)$ defined by (1) equals q.*

2. *There exists $w \in \mathbb{Z}^q$ such that $\sum_i w_i = 0$ and the vector Mw, after sorting its coordinates, is an arithmetic progression with nonzero common difference.*
3. *There exists a permutation π on $[q]$ such that*

$$
\begin{pmatrix} \pi(1) \\ \vdots \\ \pi(q) \\ 0 \end{pmatrix} \text{ is in the column space of } \begin{pmatrix} M & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix},
$$

where the column space is understood as a subspace of $\mathbb{Q}^{q+1}$, and $\mathbf{1}$ is the q-dimensional all-ones column vector.

Proof. Let $w \in \mathbb{Z}^q$ be a vector such that $\sum_{i=1}^q w_i = 0$ and the coordinates of Mw are distinct integers, and let $g := \gcd_{i<j}((Mw)_i - (Mw)_j)$. Clearly, $\max Mw - \min Mw \geq (q-1)g$, and equality holds if and only if the vector Mw, after sorting its coordinates, is an arithmetic progression with common difference $g > 0$. This shows the implication from Statement 1 to Statement 2. The converse is evident.

Lastly, we demonstrate the equivalence between Statement 2 and Statement 3. Suppose that there exists $w \in \mathbb{Z}^n$ such that $\mathbf{1}^T w_i = 0$ and the vector Mw, after sorting its coordinates, is an arithmetic progression with nonzero common difference. Thus there exists $a, b \in \mathbb{Z}$ with $b \neq 0$ and a permutation π such that

$$
Mw = \begin{pmatrix} a + b\pi(1) \\ \vdots \\ a + b\pi(q) \end{pmatrix} = a\mathbf{1} + b \begin{pmatrix} \pi(1) \\ \vdots \\ \pi(q) \end{pmatrix}.
$$

We obtain that

$$
\begin{pmatrix} M & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} w \\ -a \end{pmatrix} = \begin{pmatrix} Mw - a\mathbf{1} \\ 0 \end{pmatrix} = b \begin{pmatrix} \pi(1) \\ \vdots \\ \pi(q) \\ 0 \end{pmatrix},
$$

which implies Statement 3. Reversing the argument, one can show that Statement 3 indicates the existence of $w \in \mathbb{Q}^q$ satisfying the conditions in Statement 2. However, one can always scale w properly so that it becomes a vector in $\mathbb{Z}^q$. $\square$

Corollary 6. *Given a connected graph G on $q \geq 2$ vertices, let M be the distance matrix of G. If G is a complete graph, a path, a cycle or a complete bipartite graph, or the matrix*

$$
M' := \begin{pmatrix} M & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}
$$

Table 2
Construction of $w \in \mathbb{Z}^q$ for complete graphs, paths and cycles.

Complete graph	Path	Even cycle	Odd cycle
$w_i = 2i - (q+1)$	$w_i = \begin{cases} -1 & \text{if } i = 1 \\ 1 & \text{if } i = q \\ 0 & \text{otherwise} \end{cases}$	$w_i = \begin{cases} +1 & \text{if } i = 1 \\ -\frac{q+2}{2} & \text{if } i = \frac{q}{2} \\ \frac{q}{2} & \text{if } i = \frac{q+2}{2} \\ 0 & \text{otherwise} \end{cases}$	$w_i = \begin{cases} \frac{q-3}{2} & \text{if } i = \frac{q+1}{2} \\ \frac{q-1}{2} & \text{if } i = q \\ -1 & \text{otherwise} \end{cases}$

is invertible, then the metric dimension $m(G, n)$ of $G^{\Box n}$ is

$$\left(2 + O\left(\frac{\log \log n}{\log n}\right)\right) \frac{n}{\log_q n}.$$

Proof. When M' is invertible, Statement 3 in Lemma 5 applies here. When G is a complete graph, a path or a cycle, by Statement 2 in Lemma 5, it suffices to construct a vector $w \in \mathbb{Z}^q$ such that $\sum_i w_i = 0$ and the vector Mw, after sorting its coordinates, is an arithmetic progression with nonzero common difference. We list the construction of w in Table 2 and leave the verification to the readers.

Lastly, because $K_{2,2}$ is a cycle of length 4, for a complete bipartite graph $G = K_{q_1,q_2}$, it suffices to check that M' is invertible for $q_1 \neq 2$. Denote by J_q the q-dimensional all-ones matrix, and by I_q the q-dimensional identity matrix. Recall that J_{q_1} has eigenvalues 0 and q_1. As $q_1 \neq 2$, $J_{q_1} - 2I_{q_1}$ is invertible and $(J_{q_1} - 2I_{q_1})\mathbf{1} = (q_1 - 2)\mathbf{1}$, hence $\mathbf{1}^T(J_{q_1} - 2I_{q_1})^{-1}\mathbf{1} = \frac{q_1^2}{q_1 - 2}$. Using row operations and Schur complements,[5] we have the following matrix equivalence:

$$\begin{pmatrix} M & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = \begin{pmatrix} 2J_{q_1} - 2I_{q_1} & J & \mathbf{1} \\ J & 2J_{q_2} - 2I_{q_2} & \mathbf{1} \\ \mathbf{1}^T & \mathbf{1}^T & 0 \end{pmatrix} \sim \begin{pmatrix} J_{q_1} - 2I_{q_1} & O & \mathbf{1} \\ O & J_{q_2} - 2I_{q_2} & \mathbf{1} \\ \mathbf{1}^T & \mathbf{1}^T & 0 \end{pmatrix}$$

$$\sim \begin{pmatrix} J_{q_1} - 2I_{q_1} & O & \mathbf{0} \\ O & J_{q_2} - 2I_{q_2} & \mathbf{1} \\ \mathbf{0}^T & \mathbf{1}^T & -\frac{q_1^2}{q_1 - 2} \end{pmatrix}$$

$$\sim \begin{pmatrix} J_{q_1} - 2I_{q_1} & O & \mathbf{0} \\ O & \left(1 + \frac{q_1 - 2}{q_1^2}\right) J_{q_2} - 2I_{q_2} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & -\frac{q_1^2}{q_1 - 2} \end{pmatrix}.$$

Notice that $\left(1 + \frac{q_1 - 2}{q_1^2}\right) J_{q_2} - 2I_{q_2}$ has eigenvalues $\left(1 + \frac{q_1 - 2}{q_1^2}\right) q_2 - 2$ and -2, which are nonzero. Therefore M' is invertible for a complete bipartite graph. $\square$

---

[5] Suppose $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is a block matrix and A is invertible. The Schur complement of the block A is $M/A := D - CA^{-1}B$, which gives rise to the matrix equivalence $M \sim \begin{pmatrix} A & O \\ O & M/A \end{pmatrix}$.
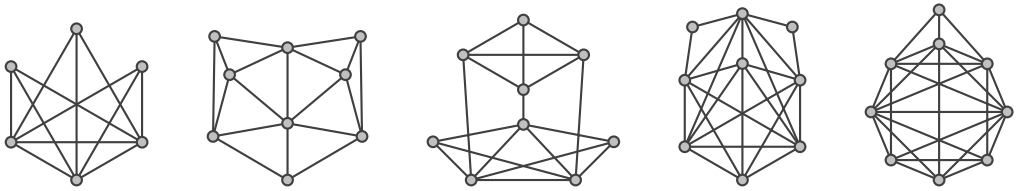
Fig. 1. All the connected graphs on $2 \leq q \leq 9$ vertices with $r(G) > q$.

Remark 2. Sebő and Tannier [20, Section 1] claimed that $m(P_q, n) \leq (2+o(1))n/\log_q n$, where P_q is the path on q vertices, and they thought "this upper bound is probably the asymptotically correct value". Our result confirms their conjecture.

5. Open problems

Statement 3 in Lemma 5 allows us to search for connected graphs G on q vertices with $r(G) > q$. For each connected graph G on q vertices, we check if the system of equations

$$\begin{pmatrix} M & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_q \\ x_{q+1} \end{pmatrix} = \begin{pmatrix} \pi(1) \\ \vdots \\ \pi(q) \\ 0 \end{pmatrix}$$

has a solution for some permutation π on $[q]$. Using McKay's dataset [16] of connected graphs on up to 10 vertices, we find 1 graph on 6 vertices, 4 graphs on 9 vertices and 1709 graphs on 10 vertices for which $r(G) > q$ (see Fig. 1).

The graph on 6 vertices is $K_6 \setminus K_3$. We give a simple argument for $r(K_6 \setminus K_3) = 7$ in Appendix A. We believe that our construction of a resolving set can be significantly improved for such graphs.

Conjecture A. *Given a connected graph G on $q \geq 2$ vertices, the metric dimension $m(G, n)$ of $G^{\square n}$ is $(2+o(1))n/\log_q n$. In particular, $m(K_6 \setminus K_3, n) = (2+o(1))n/\log_6 n$.*

In the proofs of Theorem 1 and Theorem 4, we have made use of one property of graph distance, that is, it is integer valued. In addition, we used some other properties of graph distance in Remark 1 just to show that Theorem 1 is not vacuously true for any connected graph. In this sense, our results are more related to integer matrices than graphs.

Definition 1. Given a $p \times q$ integer matrix M and $n \in \mathbb{N}$, $m(M, n)$ is the minimum cardinality of a subset S of $[q]^n$ such that $M_S \colon [p]^n \to \mathbb{N}^S$, defined by $(M_S(i_1, \ldots, i_n))_s = M_{i_1, s_1} + \cdots + M_{i_n, s_n}$ for every $(i_1, \ldots i_n) \in [p]^n$ and $s = (s_1, \ldots, s_n) \in S$, is an injection.

If the difference between two rows of M, say the first two, is parallel to $\mathbf{1}^T$, then for every $n \geq 2$ and $S \subset [q]^n$, $M_S(1, 2, 1, 1, \dots) = M_S(2, 1, 1, 1, \dots)$, and so $m(M, n) = \infty$. Otherwise, Theorem 1 and Theorem 4 generalize to integer matrices naturally.

Theorem 7. *Given a $p \times q$ integer matrix M with $p \geq 2$, if none of the differences between two rows of M is parallel to $\mathbf{1}^T$, then for every $n \in \mathbb{N}$,*

$$\left(2 - O\left(\frac{\log\log n}{\log n}\right)\right) \frac{n}{\log_p n} \leq m(M, n) \leq \left(2 + O\left(\frac{\log\log n}{\log n}\right)\right) \frac{n}{\log_r n},$$

where $r = r(M)$ is defined by

$$r(M) = \min\left\{ \frac{\max Mw - \min Mw}{\gcd_{i<j}((Mw)_i - (Mw)_j)} : w \in \mathbb{Z}^q,\ \sum_i w_i = 0,\right.$$

$$\left. (Mw)_i \neq (Mw)_j \text{ for all } i \neq j \right\} + 1.$$

Remark 3. The same argument in Remark 1 shows that $p \leq r(M) < \infty$ in Theorem 7.

It is conceivable that the generalization of Conjecture A to integer matrices holds.

Conjecture B. *Given a $p \times q$ integer matrix M with $p \geq 2$, if none of the differences between two rows of M is parallel to $\mathbf{1}^T$, then $m(M, n) = (2 + o(1))n/\log_p n$.*

Acknowledgments

Appendix A. Proof of $r(K_6 \setminus K_3) = 7$

Proposition 8. *Let M be the distance matrix of $K_6 \setminus K_3$. For every $w \in \mathbb{Z}^6$ such that $\sum_i w_i = 0$, the vector Mw, after sorting its coordinates, is never an arithmetic progression with nonzero common difference. Moreover, there exists $w \in \mathbb{Z}^6$ such that $\sum_i w_i = 0$ and the coordinates of Mw consist of $0, 2, 3, 4, 5, 6$.*

Proof. Label the vertices of $K_6 \setminus K_3$ of degree 3 by $1, 2, 3$ and those of degree 5 by $4, 5, 6$, and let M be the distance matrix of $K_6 \setminus K_3$:

$$M := \begin{pmatrix} 0 & 2 & 2 & 1 & 1 & 1 \\ 2 & 0 & 2 & 1 & 1 & 1 \\ 2 & 2 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Assume for the sake of contradiction that there exists $w \in \mathbb{Z}^6$ and a permutation π on $\{1, 2, 3, 4, 5, 6\}$ such that

$$(Mw)_i = c + \pi(i)d \text{ for all } i \in \{1, 2, 3, 4, 5, 6\}, \tag{8}$$

where $c \in \mathbb{Z}$ and $d \in \mathbb{Z} \setminus \{0\}$. Observe that

$$
\begin{aligned}
3c + (\pi(1) + \pi(2) + \pi(3))d &= (Mw)_1 + (Mw)_2 + (Mw)_3 \\
&= 4(w_1 + w_2 + w_3) + 3(w_4 + w_5 + w_6), \\
3c + (\pi(4) + \pi(5) + \pi(6))d &= (Mw)_4 + (Mw)_5 + (Mw)_6 \\
&= 3(w_1 + w_2 + w_3) + 2(w_4 + w_5 + w_6).
\end{aligned}
$$

Since $\sum_i w_i = 0$, we get that $\pi(1) + \pi(2) + \pi(3) = \pi(4) + \pi(5) + \pi(6)$, contradicting to $\sum_i \pi(i) = 15$. Finally, $w = \begin{pmatrix} 5 & 3 & 2 & -2 & -3 & -5 \end{pmatrix}^T$ satisfies $\sum_i w_i = 0$ and $Mw = \begin{pmatrix} 0 & 4 & 6 & 2 & 3 & 5 \end{pmatrix}^T$. $\square$

References

[1] Richard Bellman, Harold N. Shapiro, On a problem in additive number theory, Ann. of Math. (2) 49 (1948) 333–340.
[2] Leonard M. Blumenthal, Theory and Applications of Distance Geometry, Clarendon Press, Oxford, 1953.
[3] Nader H. Bshouty, Optimal algorithms for the coin weighing problem with a spring scale, in: Proceedings of the 22nd Annual Conference on Learning Theory, Montreal, Quebec, Canada, June 2009, Association for Computational Learning.
[4] José Cáceres, Carmen Hernando, Mercè Mora, Ignacio M. Pelayo, María L. Puertas, Carlos Seara, David R. Wood, On the metric dimension of Cartesian products of graphs, SIAM J. Discrete Math. 21 (2) (2007) 423–441, arXiv:math/0507527 [math.CO].
[5] David G. Cantor, W.H. Mills, Determination of a subset from certain combinatorial properties, Canad. J. Math. 18 (1966) 42–48.
[6] V. Chvátal, Mastermind, Combinatorica 3 (3–4) (1983) 325–329.
[7] Paul Erdős, Alfréd Rényi, On two problems of information theory, Magy. Tud. Akad. Mat. Kut. Intéz. Közl. 8 (1963) 229–243.
[8] Frank Harary, Robert A. Melter, On the metric dimension of a graph, Ars Combin. 2 (1976) 191–195.
[9] G. Kabatianski, V. Lebedev, J. Thorpe, The Mastermind game and the rigidity of the Hamming space, in: 2000 IEEE International Symposium on Information Theory, IEEE, Italy, June 2000, p. 375.
[10] G.A. Kabatiansky, V.S. Lebedev, On metric dimension of nonbinary Hamming spaces, Probl. Inf. Transm. 54 (1) (January 2018) 48–55.
[11] Donald E. Knuth, The computer as master mind, J. Recreat. Math. 9 (1) (1976/1977) 1–6.
[12] Bernt Lindström, On a combinatory detection problem. I, Magy. Tud. Akad. Mat. Kut. Intéz. Közl. 9 (1964) 195–207.
[13] Bernt Lindström, On a combinatorial problem in number theory, Canad. Math. Bull. 8 (1965) 477–490.
[14] Bernt Lindström, Determinants on semilattices, Proc. Amer. Math. Soc. 20 (1969) 207–208.
[15] Bernt Lindström, On Möbius functions and a problem in combinatorial number theory, Canad. Math. Bull. 14 (1971) 513–516.
[16] Brendan McKay, Combinatorial data: graphs, 2018, http://users.cecs.anu.edu.au/~bdm/data/graphs.html, archived at https://web.archive.org/web/20180410163655/http://users.cecs.anu.edu.au/~bdm/data/graphs.html.

[17] L. Moser, The second moment method in combinatorial analysis, in: Combinatorial Structures and Their Applications, Proc. Calgary Internat. Conf., Calgary, Alta., 1969, Gordon and Breach, New York, 1970, pp. 283–284.

[18] Nicholas Pippenger, An information-theoretic method in combinatorial theory, J. Combin. Theory Ser. A 23 (1) (1977) 99–104.

[19] Bruce Schechter, My Brain is Open: the Mathematical Journeys of Paul Erdős, Simon & Schuster, New York, 1998.

[20] András Sebő, Eric Tannier, On metric generators of graphs, Math. Oper. Res. 29 (2) (2004) 383–393.

[21] Peter J. Slater, Leaves of trees, in: Proceedings of the Sixth Southeastern Conference on Combinatorics, Graph Theory, and Computing, Boca Raton, Florida, February 1975, in: Congressus Numerantium, vol. 14, 1975, pp. 549–559.

[22] Staffan Söderberg, H.S. Shapiro, A combinatory detection problem, Amer. Math. Monthly 70 (10) (1963) 1066–1070.