How to guess an n-digit number

Zilin Jiang* Nikita Polyanskii†

Abstract

In a deductive game for two players, SF and PGOM, SF conceals an n-digit number $x = x_1, \dots, x_n$ in base q, and PGOM, who knows n and q, tries to identify x by asking a number of questions, which are answered by SF. Each question is an n-digit number $y = y_1, \dots, y_n$ in base q; each answer is the number of subscripts i such that $x_i = y_i$. Moreover, we require PGOM send all the questions at once. We show that the minimum number of questions required to determine x is $(2+o_q(1))n/\log_q n$. Our result closes the gap between the lower bound attributed to Erdős and Rényi and the upper bounds developed subsequently by Lindström, Chvátal, Kabatianski, Lebedev and Thorpe. A more general problem is to determine the asymptotic formula of the metric dimension of Cartesian powers of a graph. We state the class of graphs for which the formula can be determined, and the smallest graphs for which we did not manage to settle.

1 Introduction

Mastermind is a deductive game for two players, the codemaker and the codebreaker. In this game, the codemaker conceals a vector $x = (x_1, \dots, x_n) \in [q]^n$, 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 $y = (y_1, \dots, y_n) \in [q]^n$; 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.

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 1983, Chvátal [7] initiated the study of this version of Mastermind, and in honor of Erdős, he referred to the codemaker and the codebreaker as SF and PGOM[1].

Cáceres et al. [5, §6] noticed that the minimum number of questions required to determine x is exactly the metric dimension of a Hamming graph. Recall that a set of vertices S *resolves* a graph if every vertex is uniquely determined by its vector of distances to the vertices in S, and the *metric dimension* of a graph is the minimum cardinality of a resolving set of the graph. The mystery vector x can be seen as a vertex in the Hamming graph $K_q^{\square n}$, that is, the Cartesian product[2] of n copies of the complete graph K_q on q vertices. The questions can similarly be viewed as a set S of vertices in $K_q^{\square n}$. Since the distance between x and y in $K_q^{\square n}$ is precisely $n - b(x, y)$, the answers thus are equivalent to the vector of distances from x to S. To determine x uniquely for every x, S must resolve $K_q^{\square n}$ and so the minimum number of the questions is the metric dimension of $K_q^{\square n}$.

This paper undertakes the study of the asymptotic behavior of $m(K_q, n)$, the metric dimension of $K_q^{\square n}$, when q is fixed and n tends to infinity. The concept of resolving sets and that of metric dimension date back to the 1950s — they were defined by Bluementhal [3] in the context of metric spaces. These notions were introduced to graph theory by Harary and Melter [9] and Slater [22] in the 1970s. For the relationship between the metric dimension of a graph and the base size of a permutation group, see [1].

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 [23], 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

*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: `zilinj@mit.edu`.
The work was done when Z. Jiang was a postdoctoral fellow at Technion – Israel Institute of Technology, and was supported in part by Israel Science Foundation (ISF) grant nos. 1162/15, 936/16.

†CDISE, Skolkovo Institute of Science and Technology, and Department of Mathematics, Technion – Israel Institute of Technology. Email: `nikita.polyansky@gmail.com`.
Supported in part by ISF grant nos. 1162/15, 326/17, and by the Russian Foundation for Basic Research through grant nos. 16-01-00440 A, 18-07-01427 A, 18-31-00310 MOL_A.

[1] See [20, p. 41 and p. 70] for what SF and PGOM stand for.

[2] The *Cartesian product* of graphs $G_1, \dots, G_n$ is the graph with vertex set $V(G_1) \times \cdots \times V(G_n)$ such that $(u_1, \dots, u_n)$ and $(v_1, \dots, v_n)$ 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.

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 [21, §1]). A lower bound on the number of weighings by Erdős and Rényi [8] and an upper bound by Lindström [13] and independently by Cantor and Mills [6] imply that $m(K_2, n) = (2 + o(1))n/\log_2 n$.

Kabatianski, Lebedev and Thorpe [11] stated that a straightforward generalization of the lower bound on $m(K_2, n)$ by Erdős and Rényi [8] 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 [7, Theorem 1], show that $m(K_q, n) \leq (2 + o(1)) \log_q(1 + (q - 1)q) \cdot n/\log_q n$. Very recently, Kabatianski and Lebedev [12] proved that $m(K_q, n) = (2 + o(1))n/\log_q n$ for $q = 3, 4$, which was previously announced in [11, Theorem 1], and they conjectured that $m(K_q, n) = (2 + o(1))n/\log_q n$ for all q.

We resolve this conjecture in §2 via a construction of a resolving set of $K_q^{\square n}$ of size $(2+o(1))n/\log_q n$. A more general problem is to estimate the metric dimension $m(G, n)$ of $G^{\square n}$, the Cartesian product of n copies of a connected graph G. We establish an upper bound on $m(G, n)$ for a class of connected graphs in §3 and a lower bound for every connected graph in §4. In particular, we show in addition that $m(G, n) = (2 + o(1))n/\log_q n$ when G is a path, a cycle, or a complete bipartite graph on q vertices in §5. We emphasize that the situation is totally different when q varies and n is fixed. For example, Cáceres et al. [5, Theorem 6.1] showed that $m(K_q, 2) = \lfloor 2(2q - 1)/3\rfloor$. We conclude with an open problem in §6.

2 The asymptotic formula of $m(K_q, n)$

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

THEOREM 2.1. *Given $q \geq 2$, for every $n \in \mathbb{N}$, the metric dimension $m(K_q, n)$ of $K_q^{\square n}$ is at most*

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

Our construction of a resolving set of $K_q^{\square n}$ is inspired by the upper bound for the coin weighing problem by Lindström [14]. Among various constructions such as the recursive construction by Cantor and Mills [6] and the construction by Bshouty [4] based on Fourier transform, we find the one using the theory of Möbius functions by Lindström [16] 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 natural numbers $\mathbb{N} = \{0, 1, \dots\}$ partially ordered by divisibility. For our purpose, we first consider binary representation of natural numbers, and we instead partially order $\mathbb{N}$ in the following way: $x \preceq y$ if and only if $x = x \wedge y$, where $\wedge$ is the bitwise AND operation. 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 a greatest lower bound. We need the following identity of Lindström [15, Lemma] for our meet-semilattice.

LEMMA 2.1. *Let $(P, \prec, \wedge)$ be a locally finite meet-semilattice with Möbius function $\mu(x, y)$. Let $a, b \in P$ and $b \not\preceq a$. Let $f(x)$ be defined for all $x \preceq a \wedge b$ with values in a commutative ring with unity. Then we have*

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

The last ingredient is the following estimation on the partial sum of $n(\cdot)$ due to Bellman and Shapiro [2, Theorem 1].

THEOREM 2.2. *As $x \to \infty$,*

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

We now construct a resolving set for Theorem 2.1 using the Möbius function of $(\mathbb{N}, \prec, \wedge)$.

Proof. Let M be the distance matrix of K_q, and let $w \in \mathbb{Z}^q$ be such that $\sum_i w_i = 0$ and the vector Mw, after sorting its coordinates, is an arithmetic progression with common difference $g \neq 0$. For example, we could take w with $w_i = 2i - (q + 1)$ for $i \in [q]$ and $g = 2$. The specific choice of w does not matter. Set $|w|_1 := \sum_i |w_i|$. For each $j \in \mathbb{N}$, let $b(j)$ be the largest integer such that

$$q^{b(j)} \cdot |w|_1 \leq 2^{n(j)}, \tag{2.1}$$

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

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 K_q in $K_q^{\square n}$ by J, namely each vertex of $K_q^{\square n}$ is an element of V^J, where $V = \{v_1, \ldots, v_q\}$ is the vertex set of K_q. 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 $K_q^{\square n}$. 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.

$$\sum_{i \preceq j} S(i,j,k)\mu(i,j) = q^k(w_1 v_1 + \cdots + w_q v_q), \quad \text{for all } (j,k) \in J; \tag{2.2a}$$

$$\sum_{i \preceq j} S(i,j',k)\mu(i,j) = 0, \quad \text{for all } (j',k) \in J \text{ and } j' < j \le m. \tag{2.2b}$$

We remark that (2.2a) and (2.2b) happen in the commutative ring $\mathbb{Z}[v_1, \ldots, v_q]$ with unity.

For example, when $q = 3$, $(w_1, w_2, w_3) = (-2, 0, 2)$. In the table below, we supply the values of $S(i,j,k)$ for $(j,k) = (7,0), (7,1)$. The reader can verify (2.2a) in this case.

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	

In general, pick arbitrary $(j,k) \in J$. On the left hand side of (2.2a), the summation consists of $2^{n(j)}$ terms, moreover $2^{n(j)-1}$ of them has $\mu(i,j) = +1$ (respectively -1). Since $q^k(w_1 + \cdots + w_q) = 0$ and $q^k(|w_1| + \cdots + |w_q|) = q^k \, |w|_1 \le q^{b(j)} \, |w|_1 \le 2^{n(j)}$ by (2.1), 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, to satisfy (2.2a). 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 (2.2b) equals $\sum_{i \preceq j} S(i \wedge j', j', k)\mu(i,j) = 0$ by applying Lemma 2.1 to the function $f_{j',k}(i) = S(i,j',k)$.

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

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

where $d \colon V \times V \to \mathbb{N}$ is the distance function of K_q. 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 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 $\sum_{i \preceq j_0} D_i \mu(i, j_0)$ equals

$$\begin{aligned}
&\sum_{i \preceq j_0} \left(\sum_{(j,k) \in J} d(X(j,k), S(i,j,k)) \right) \mu(i, j_0) \\
&= \sum_{(j,k) \in J} d\left(X(j,k), \sum_{i \preceq j_0} S(i,j,k)\mu(i,j_0) \right) \\
&\overset{(2.2b)}{=} \sum_{j=j_0}^{m} \sum_{k=0}^{b(j)} d\left(X(j,k), \sum_{i \preceq j_0} S(i,j,k)\mu(i,j_0) \right).
\end{aligned}$$

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

$$\begin{aligned}
&\sum_{k=0}^{b(j_0)} d\left(X(j_0,k), \sum_{i \preceq j_0} S(i,j_0,k)\mu(i,j_0) \right) \\
&\overset{(2.2a)}{=} \sum_{k=0}^{b(j_0)} d\left(X(j_0,k), q^k \sum_{i=1}^{q} w_i v_i \right) \\
&= \sum_{k=0}^{b(j_0)} q^k \sum_{i=1}^{q} w_i d(X(j_0,k), v_i) \\
&= \sum_{k=0}^{b(j_0)} q^k \sum_{i=1}^{q} M_{X(j_0,k), v_i} w_i \\
&= \sum_{k=0}^{b(j_0)} q^k \cdot (Mw)_{X(j_0,k)}.
\end{aligned}$$

We can thus deduce the value of

$$\sum_{k=0}^{b(j_0)} q^k \cdot \tfrac{1}{g} \left((Mw)_{X(j_0,k)} - \min Mw \right), \tag{2.3}$$

where $\min Mw$ is the minimum coordinate of the vector Mw. Notice that, according to our choice of w,

$$\left(\tfrac{1}{g}((Mw)_i - \min Mw) \right)_{i=1}^{q}$$

are distinct integers in $[0, q)$. The value of (2.3) 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 $\sum_{j=0}^{m}(\max\{b(j), 0\} + 1) \ge n$. For every $x \in \mathbb{N}$, by Theorem 2.2,

$$\begin{aligned}(2.4) \quad \sum_{j=0}^{x}(\max\{b(j), 0\} + 1) &> \sum_{j=0}^{x}(n(j)\log_q 2 - \log_q |w|_1) \\ &= \tfrac{1}{2}x\log_q x - O(x\log\log x).\end{aligned}$$

One can check that

$$x = \frac{2n}{\log_q n} + O\left(\frac{n\log\log n}{\log^2 n}\right)$$

ensures the right hand side of (2.4) is $\ge n$.

3 An upper bound on $m(G, n)$

In the proof of Theorem 2.1, we have made use of one property of the distance matrix M of K_q, that is, the existence of $w \in \mathbb{Z}^q$ such that $\sum_i w_i = 0$ and the vector Mw, after sorting its coordinates, is an arithmetic progression with nonzero common difference. It turns out that the same proof goes through without alternation for connected graphs with this property.

THEOREM 3.1. *Given a connected graph G on $q \ge 2$ vertices, let M be the distance matrix of G. If there exists $w \in \mathbb{Z}^q$ such that $\sum_i w_i = 0$ and the vector, after sorting its coordinates, is an arithmetic progression with nonzero common difference, then for every $n \in \mathbb{N}$, the metric dimension $m(G, n)$ of $G^{\square n}$ is at most*

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

4 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 $G^{\square n}$ (see Moser [18] and Pippenger [19] for different proofs using the second moment method and the information-theoretic method).

THEOREM 4.1. *Given a connected graph G on $q \ge 2$ vertices, for every $n \in \mathbb{N}$, the metric dimension $m(G, n)$ of $G^{\square n}$ 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 $G^{\square n}$ 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 $s = (s_1, \dots, s_n) \in S$, let $X_1, \dots, X_n$ be independent random variables defined by $X_i = d(Y_i, s_i)$, where the independent random variables $Y_1, \dots, Y_n$ are chosen uniformly at random from V, and define A_s as the collection of $(v_1, \dots, v_n) \in V^n$ such that

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

where D is the diameter of the graph. Since each X_i is bounded by $[0, D]$, Hoeffding's inequality [10, Theorem 2] provides an upper bound on the cardinality of the complement of A_s: $|V^n \setminus A_s|/|V^n|$ can be interpreted as the probability that $(Y_1, \dots, Y_n)$ is not in A_s, which equals

$$\begin{aligned}&\Pr\left(\left|\sum_{i=1}^{n} d(Y_i, s_i) - \mathrm{E}\left[\sum_{i=1}^{n} X_i\right]\right| \ge \sqrt{n\ln n}\cdot D\right) \\ &= \Pr\left(\left|\sum_{i=1}^{n} X_i - \mathrm{E}\left[\sum_{i=1}^{n} X_i\right]\right| \ge \sqrt{n\ln n}\cdot D\right) \\ &\le 2\exp\left(-\frac{2(\sqrt{n\ln n}\cdot D)^2}{n\cdot D^2}\right) = \frac{2}{n^2}.\end{aligned}$$

From the definition of a resolving set mentioned, we know that the function $d_S\colon V^n \to \mathbb{N}^S$, defined by $(d_S(v_1, \dots, v_n))_s := d(v_1, s_1) + \dots + d(v_n, s_n)$ for every $(v_1, \dots, v_n) \in V^n$ and $s = (s_1, \dots, s_n) \in S$, 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

$$\begin{aligned}\left(2\sqrt{n\ln n}\cdot D\right)^m &\ge \left|\bigcap_{s\in S} A_s\right| \ge |V^n| - \sum_{s\in S}|V^n \setminus A_s| \\ &\ge q^n\left(1 - \frac{2m}{n^2}\right) = q^n\left(1 - O\left(\frac{1}{n}\right)\right).\end{aligned}$$

Taking logarithms gives

$$\begin{aligned}m &\ge \frac{n\ln q - O\left(\frac{1}{n}\right)}{\frac{1}{2}\ln n + O(\log\log n)} = \frac{2n}{\log_q n}\cdot\frac{1 - O\left(\frac{1}{n^2}\right)}{1 + O\left(\frac{\log\log n}{\log n}\right)} \\ &= \left(2 - O\left(\frac{\log\log n}{\log n}\right)\right)\frac{n}{\log_q n}.\end{aligned}$$

5 Asymptotically tight cases

We characterize the graphs which satisfy the technical condition in Theorem 3.1.

LEMMA 5.1. *Given a connected graph G on $q \ge 2$ vertices, let M be the distance matrix of G. The following statements are equivalent.*

1. *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.*

2. *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. 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 2. Reversing the argument, one can show that Statement 2 indicates the existence of $w \in \mathbb{Q}^q$ satisfying the conditions in Statement 1. However, one can always scale w properly so that it becomes a vector in $\mathbb{Z}^q$.

COROLLARY 5.1. *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*

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

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

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

Proof. When M' is invertible Statement 2 in Lemma 5.1 applies here. When G is a complete graph, a path or a cycle, 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 below and leave the verification to the readers.

$$\begin{array}{rl}
\text{complete graph} & w_i = 2i - (q+1) \\
\text{path} & w_i = \begin{cases} -1 & \text{if } i = 1 \\ 1 & \text{if } i = q \\ 0 & \text{otherwise} \end{cases} \\
\text{even cycle} & w_i = \begin{cases} +1 & \text{if } i = 1 \\ -\frac{q+2}{2} & \text{if } i = \frac{q}{2} \\ \frac{q}{2} & \text{if } i = \frac{q+2}{2} \\ 0 & \text{otherwise} \end{cases} \\
\text{odd cycle} & w_i = \begin{cases} \frac{q-3}{2} & \text{if } i = \frac{q+1}{2} \\ \frac{q-1}{2} & \text{if } i = q \\ -1 & \text{otherwise} \end{cases}
\end{array}$$

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, we have the following matrix equivalence:

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

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

REMARK 5.1. *Sebő and Tannier [21, §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.*

6 An open problem

Statement 2 in Lemma 5.1 allows us to search for connected graphs which violate the technical condition in Theorem 3.1. For each connected graph G on q vertices with distance matrix M, 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 [17] 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.

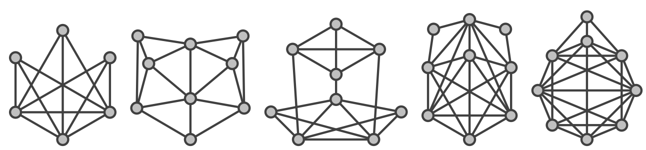

Figure 1: All the connected graphs on $2 \leq q \leq 9$ vertices which violate the technical condition in Theorem 3.1.

We believe that our construction of a resolving set can be extended for those graphs.

Conjecture 6.1. *Given a connected graph G on $q \geq 2$ vertices, the metric dimension $m(G, n)$ of $G^{\square n}$ is $(2 + o(1))n/\log_q n$.*

Acknowledgement

We are grateful to the anonymous reviewers for their close reading of this paper and many helpful comments and corrections.

References

[1] R. F. Bailey and P. J. Cameron. Base size, metric dimension and other invariants of groups and graphs. *Bull. Lond. Math. Soc.*, 43(2):209–242, 2011.

[2] R. Bellman and H. N. Shapiro. On a problem in additive number theory. *Ann. of Math. (2)*, 49:333–340, 1948.

[3] L. M. Blumenthal. *Theory and applications of distance geometry*. Oxford, at the Clarendon Press, 1953.

[4] N. 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.

[5] J. Cáceres, C. Hernando, M. Mora, I. M. Pelayo, M. a. L. Puertas, C. Seara, and D. R. Wood. On the metric dimension of Cartesian products of graphs. *SIAM J. Discrete Math.*, 21(2):423–441, 2007. `arXiv:math/0507527[math.CO]`.

[6] D. G. Cantor and W. H. Mills. Determination of a subset from certain combinatorial properties. *Canad. J. Math.*, 18:42–48, 1966.

[7] V. Chvátal. Mastermind. *Combinatorica*, 3(3-4):325–329, 1983.

[8] P. Erdős and A. Rényi. On two problems of information theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:229–243, 1963.

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

[10] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[11] G. Kabatianski, V. Lebedev, and J. Thorpe. The Mastermind game and the rigidity of the Hamming space. In *2000 IEEE International Symposium on Information Theory*, page 375, Italy, June 2000. IEEE.

[12] G. A. Kabatiansky and V. S. Lebedev. On metric dimension of nonbinary Hamming spaces. *Problems of Information Transmission*, 54(1):48–55, Jan 2018.

[13] B. Lindström. On a combinatory detection problem. I. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 9:195–207, 1964.

[14] B. Lindström. On a combinatorial problem in number theory. *Canad. Math. Bull.*, 8:477–490, 1965.

[15] B. Lindström. Determinants on semilattices. *Proc. Amer. Math. Soc.*, 20:207–208, 1969.

[16] B. Lindström. On Möbius functions and a problem in combinatorial number theory. *Canad. Math. Bull.*, 14:513–516, 1971.

[17] B. 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`.

[18] L. Moser. The second moment method in combinatorial analysis. In *Combinatorial Structures and their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969)*, pages 283–284. Gordon and Breach, New York, 1970.

[19] N. Pippenger. An information-theoretic method in combinatorial theory. *J. Combinatorial Theory Ser. A*, 23(1):99–104, 1977.

[20] B. Schechter. *My brain is open: The mathematical journeys of Paul Erdős*. Simon & Schuster, New York, 1998.

[21] A. Sebő and E. Tannier. On metric generators of graphs. *Math. Oper. Res.*, 29(2):383–393, 2004.

[22] P. J. Slater. Leaves of trees. pages 549–559. Congressus Numerantium, No. XIV, 1975.

[23] S. Söderberg and H. S. Shapiro. A combinatory detection problem. *Amer. Math. Monthly*, 70(10):1066–1070, 1963.