

Multiset Metric Dimension of Binomial Random Graphs

Austin Eide*, Paweł Prałat†

July 15, 2025

Abstract

For a graph $G = (V, E)$ and a subset $R \subseteq V$, we say that R is *multiset resolving* for G if for every pair of vertices v, w , the *multisets* $\{d(v, r) : r \in R\}$ and $\{d(w, r) : r \in R\}$ are distinct, where $d(x, y)$ is the graph distance between vertices x and y . The *multiset metric dimension* of G is the size of a smallest set $R \subseteq V$ that is multiset resolving (or ∞ if no such set exists). This graph parameter was introduced by Simanjuntak, Siagian, and Vitrík in 2017 [19], and has since been studied for a variety of graph families. We prove bounds which hold with high probability for the multiset metric dimension of the binomial random graph $\mathcal{G}(n, p)$ in the regime $d = (n - 1)p = \Theta(n^x)$ for fixed $x \in (0, 1)$.

1 Introduction

1.1 Background

For a finite, connected graph $G = (V, E)$ and vertices $v, w \in V$, we let $d(v, w)$ denote the distance between v and w , that is, the length of a shortest path in G between v and w . For a vertex $v \in V$ and integer $k \geq 0$, we let

$$S_k(v) := \{w \in V : d(v, w) = k\} \quad \text{and} \quad N_k(v) := \{w \in V : d(v, w) \leq k\} = \bigcup_{j=0}^k S_j(v)$$

be the sphere and, respectively, the ball of radius k around v . Similarly, for a subset $V' \subseteq V$ and $k \geq 0$, let

$$S_k(V') := \bigcup_{v \in V'} S_k(v) \quad \text{and} \quad N_k(V') := \bigcup_{v \in V'} N_k(v).$$

*Department of Mathematics, Toronto Metropolitan University, Toronto, ON, Canada; e-mail: austin.eide@torontomu.ca

†Department of Mathematics, Toronto Metropolitan University, Toronto, ON, Canada; e-mail: pralat@torontomu.ca

Finally, for $V', R \subseteq V$, let

$$S_k^R(V') := S_k(V') \cap R \quad \text{and} \quad N_k^R(V') := N_k(V') \cap R.$$

The *metric signature* of v with respect to a set $R \subseteq V$ is the vector $\mathbf{s}_R(v) := (d(v, w))_{w \in R} \in \mathbb{N}_0^{|R|}$, where $\mathbb{N}_0$ is the set of non-negative integers. (We assume that the vertices of R are ordered and that the indices of $\mathbf{s}_R(v)$ are indexed with respect to this ordering.) We say that R is *metric resolving* if $\mathbf{s}_R(v) \neq \mathbf{s}_R(w)$ for all $v \neq w$. The *metric dimension* of G , denoted $\beta(G)$, is the size of a smallest metric resolving set for G . Note that V itself is always metric resolving, so $\beta(G) \leq |V|$. In fact, it is easy to see that for any connected graph G , $\beta(G) \leq |V(G)| - 1$ and the equality holds for $G = K_n$, the complete graph on n vertices. We remark that the definition of a metric resolving set can be extended to cover graphs which are not connected by declaring that vertices in distinct components have distance ∞ ; with this definition, the metric dimension of G is simply the sum of the metric dimensions of its components.

The study of the metric dimension of graphs dates to the 1970s, when it was first proposed by Slater [20] and Harary and Melter [12]. Since then it has received significant attention from mathematicians and computer scientists, and the metric dimension is now known for a wide variety of graph families. We refer to the recent survey [21] for a fairly comprehensive list of these families, as well as some interesting applications. We make particular note of previous works on the metric dimension of random graphs. In [3], Bollobás, Mitsche, and Prałat investigated the metric dimension of $\mathcal{G}(n, p)$ when the average degree $d = (n-1)p$ is at least $\log^5 n$; a recent paper of Díaz, Hartle, and Moore treats the cases $\log n < d < \log^5 n$ [5]. (Note that $\mathcal{G}(n, p)$ is disconnected w.h.p.¹ for $d \leq (1 - \epsilon) \log n$, where $\epsilon > 0$ is an arbitrarily small constant.) Ódor and Thiran also consider a sequential version of metric dimension on $\mathcal{G}(n, p)$ in which a player attempts to locate a hidden vertex v^* by making sequential queries of the distances $d(v, v^*)$ for $v \in V$ [18, 17]. In [6], Dudek, English, Frieze, MacRury, and Prałat study a similar process on $\mathcal{G}(n, p)$ in which the hidden vertex v^* is allowed to move along an edge in each round. See also [7] for weaker bounds in the regime in which $\mathcal{G}(n, p)$ has diameter two w.h.p. Lichev, Mitsche, and Prałat additionally study this process on random geometric graphs [15]. See also [16] for the metric dimension of random trees and very-sparse ($d < 1$) random graphs and [23] for related results on the stochastic block model.

Define the *multiset signature* of v with respect to $R \subseteq V$ as the vector

$$\mathbf{m}_R(v) := (|S_k^R(v)|)_{k=0}^{\text{diam}(G)} \in \mathbb{N}_0^{\text{diam}(G)+1},$$

where $\text{diam}(G) = \max_{u, v \in V} d(u, v)$ is the diameter of graph G . Observe that $\mathbf{m}_R(v)$ is a representation of $\mathbf{s}_R(v)$ when $\mathbf{s}_R(v)$ is considered as a multiset rather than a vector. A set R is *multiset resolving* if $\mathbf{m}_R(v) \neq \mathbf{m}_R(w)$ for all $v \neq w$, and we define the multiset metric dimension $\beta_{\text{ms}}(G)$ to be the size of a smallest multiset resolving set for G if such a set exists, and define $\beta_{\text{ms}}(G) = \infty$ otherwise. This parameter was introduced by Simanjuntak,

¹An event holds with high probability (w.h.p.) if it holds with probability tending to one as $n \rightarrow \infty$.

Siagian, and Vitřík in 2017 [19], and has since been studied by multiple authors—see, for instance, [1], [4], and [11]. As with metric resolving sets, the definition of a multiset resolving set can be extended to apply to graphs which are not connected.

A connected graph G need not have finite multiset metric dimension; for instance, it is known that any non-path graph of diameter 2 has no multiset resolving set [19, Theorem 3.1]. To remediate this issue, Gil-Pons et. al. in [10] define the *outer multiset metric dimension* $\beta_{\text{ms}}^{\text{out}}(G)$ as the size of a smallest R such that $\mathbf{m}_R(v) \neq \mathbf{m}_R(w)$ for all $v \neq w$ with $v, w \in V \setminus R$. Trivially, any set R of size $|V| - 1$ is outer-multiset resolving, and hence always $\beta_{\text{ms}}^{\text{out}}(G) \leq |V| - 1$; in fact, it is shown in [14] that $\beta_{\text{ms}}^{\text{out}}(G) = |V| - 1$ if and only if G is regular and has diameter at most 2. Clearly we have $\beta_{\text{ms}}(G) \geq \beta_{\text{ms}}^{\text{out}}(G)$ for any graph G . On the other hand, we claim that $\beta_{\text{ms}}^{\text{out}}(G) \geq \beta(G)$. Indeed, if R is outer-multiset resolving, then $\mathbf{s}_R(v) \neq \mathbf{s}_R(w)$ for all $v \neq w$ with $v, w \in V \setminus R$; but it is easy to see that this is sufficient to ensure that $\mathbf{s}_R(v) \neq \mathbf{s}_R(w)$ for all $v \neq w$, so R is also metric resolving. Hence,

$$\beta_{\text{ms}}(G) \geq \beta_{\text{ms}}^{\text{out}}(G) \geq \beta(G).$$

Multiset resolving sets inherit some interesting applications from metric resolving sets. One example is source localization [21, Section 7.1]. Suppose that at time $t = 0$ a single (unknown) vertex v_0 of a connected graph G is infected, and that any uninfected vertex which is adjacent to an infected vertex at time $t - 1$ becomes infected at time t , for each $t = 1, 2, \dots$. We would like to identify the source vertex v_0 by observing the spread of the infection. One strategy for achieving this is to place a sensor at each vertex of some multiset resolving set R for G , with each sensor emitting a signal at the first moment its vertex becomes infected. For $t = 0, 1, 2, \dots$, let N_t be the number of signals emitted by vertices in R at time t . (Note that N_t is simply the number of vertices in R at distance t from v_0 .) Then since R is multiset-resolving for G , the identity of v_0 is completely determined by the sequence $\{N_t\}_{t=0}^{\text{diam}(G)}$.

Metric resolving sets have also been shown to be useful as a tool for certain instances of *graph embedding*, where the goal is map the vertices of G to real N -dimensional space while preserving structural properties of the graph. In applications, it is preferable for N to be small. For example, in [22] the authors use small resolving sets to find low-dimensional embeddings of graphical representations of genomic sequences, allowing for the application of a variety of clustering algorithms. The embedding is obtained by mapping each vertex to its metric signature with respect to some small metric resolving set; thus, given a metric resolving set R , the embedding maps the graph into $\mathbb{R}^{|R|}$. Importantly, for the application considered in [22], the metric resolving set embedding is nearly an isometry, i.e., the Euclidean distance between the images of any pair of vertices is quite close to the graph distance between that pair.

Multiset signatures with respect to a multiset resolving set also provide a natural graph embedding, and we propose that this embedding could prove useful for some applications. While $\beta_{\text{ms}}(G) \geq \beta(G)$ for any graph G , we observe that the dimension of the *embedding* produced by a multiset resolving set for G has dimension $\text{diam}(G) + 1$, as multiset signatures are vectors of this length. (Metric resolving set embeddings have dimension at least $\beta(G)$.)

Our results demonstrate that for random graphs on n vertices with average degree on the order of n^x for $x > 0$ sufficiently small, a multiset resolving set exists w.h.p.; since the diameter of this graph is $O(\frac{1}{x})$ w.h.p. (see [2]), we get an embedding into $\mathbb{R}^{O(\frac{1}{x})}$. On the other hand, results from [3] show that corresponding metric resolving set embedding has dimension which is logarithmic in n for the “best case” values of x and polynomial in n for the “worst case.” The extent to which either embedding preserves graph distances is an interesting question for future research.

1.2 Definitions

The binomial random graph $\mathcal{G}(n, p)$ is formally defined as a distribution over the class of graphs with the set of vertices $[n] := \{1, \dots, n\}$ in which every pair $\{i, j\} \in \binom{[n]}{2}$ appears independently as an edge in G with probability p . Note that $p = p(n)$ may (and usually does) tend to zero as n tends to infinity. Most results in this area are asymptotic by nature. For more about this model see, for example, [2, 9, 13].

We let $d = (n - 1)p$, where n is the number of vertices and p is the edge probability of $\mathcal{G}(n, p)$. Given d , we reserve the index i^* to denote the largest integer $i = i(n)$ such that $d^i = o(n)$ as $n \rightarrow \infty$. We let $c = c(n) = \frac{d^{i^*+1}}{n}$. (These parameters essentially determine the diameter of $\mathcal{G}(n, p)$.) To make our results simpler to state, we assume throughout that $d = n^{x+O(\log^{-1} n)} = \Theta(n^x)$ for some fixed $x \in (0, 1)$, though we remark that the upper bound we obtain on $\beta_{\text{ms}}(\mathcal{G}(n, p))$ can be extended to a more general setting. (See Section 5.) With this simplifying assumption on d , we note that $i^* = \lfloor 1/x \rfloor$, unless $x = 1/k$ for some $k \geq 2$, in which case $i^* = k - 1$. In particular, i^* is a function of x alone (and not n).

1.3 Concentration tools

Let $X \in \text{Bin}(n, p)$ be a random variable distributed according to a Binomial distribution with parameters n and p . We will use the following consequences of *Chernoff's bound* (see e.g. [13, Theorem 2.1]): for any $t \geq 0$ we have

$$\Pr(X \geq \mathbb{E}X + t) \leq \exp\left(-\frac{t^2}{2(\mathbb{E}X + t/3)}\right) \quad (1)$$

$$\Pr(X \leq \mathbb{E}X - t) \leq \exp\left(-\frac{t^2}{2\mathbb{E}X}\right). \quad (2)$$

Note that if $t \leq \frac{3\mathbb{E}X}{2}$, the right-hand side of (1) is at most $\exp\left(-\frac{t^2}{3\mathbb{E}X}\right)$; since this also upper bounds (2), we have the following corollary: for any $0 \leq t \leq \frac{3\mathbb{E}X}{2}$,

$$\Pr(|X - \mathbb{E}X| \geq t) \leq 2 \exp\left(-\frac{t^2}{3\mathbb{E}X}\right). \quad (3)$$

1.4 The function $f_x(y)$

For $x \in (0, 1]$, define the function f_x on $[0, 1]$ by

$$f_x(y) := \sum_{i=0}^{\lfloor \frac{1}{x} \rfloor} \max\{ix + y - 1, 0\}.$$

The function f_x arises naturally in the study of the lower and upper bounds for $\beta_{\text{ms}}(\mathcal{G}(n, p))$, and we motivate its appearance in the context of each bound at the beginnings of their respective sections (Sections 3 and 4, respectively.) Here, we remark on some basic properties of f_x .

Lemma 1.1. *For $x \in (0, 1]$, the function f_x is continuous, identically 0 on $[0, 1 - \lfloor \frac{1}{x} \rfloor x]$, and strictly increasing on $[1 - \lfloor \frac{1}{x} \rfloor x, 1]$.*

The proof of Lemma 1.1 is elementary and follows immediately from the definition of the function $f_x(y)$, and so we omit it.

Lemma 1.2. *Let $x \in (0, 1]$ and $y \in [0, 1]$ be such that $f_x(y) > 0$ (or, equivalently, that $1 - \lfloor \frac{1}{x} \rfloor x < y < 1$). Then for any $\epsilon \in [0, 1 - y]$, we have*

$$f_x(y + \epsilon) - f_x(y) \geq \epsilon.$$

Proof. Since we assume that $f_x(y) > 0$, it must be the case that the largest summand in $f_x(y)$ —namely, $\max\{\lfloor \frac{1}{x} \rfloor x + y - 1, 0\}$ —is positive. Increasing y by ϵ increases this summand by ϵ ; none of the other summands decrease after the change in y , so the result follows. $\square$

Lemma 1.3. *If $x \in (0, 1/8]$, then there is a unique solution $y_4 = y_4(x) \in (0, 1)$ to $f_x(y) = 4$. Similarly, if $x \in (0, 1/2]$, then there is a unique solution $y_1 = y_1(x) \in (0, 1)$ to $f_x(y) = 1$.*

Proof. We observe that, since the function f_x is continuous, identically 0 on $[0, 1 - \lfloor \frac{1}{x} \rfloor x]$, and strictly increasing on $[1 - \lfloor \frac{1}{x} \rfloor x, 1]$ by Lemma 1.1, as long as $f_x(1) > 4$, there is necessarily a unique solution to $f_x(y) = 4$ in $(0, 1)$. We have

$$f_x(1) = \sum_{i=0}^{\lfloor 1/x \rfloor} ix = \frac{(\lfloor \frac{1}{x} \rfloor + 1) \lfloor \frac{1}{x} \rfloor \cdot x}{2}.$$

If $x \leq \frac{1}{8}$, then

$$\frac{(\lfloor \frac{1}{x} \rfloor + 1) \lfloor \frac{1}{x} \rfloor \cdot x}{2} > \frac{\lfloor \frac{1}{x} \rfloor}{2} \geq \frac{8}{2} = 4$$

where we use $(\lfloor 1/x \rfloor + 1)x > 1$ for $x > 0$.

The existence of a unique solution y_1 to $f_x(y) = 1$ for $x \in (0, 1/2]$ can be established using the same argument, since $f_x(1) > 1$ for any $x \leq 1/2$. $\square$

1.5 Main result

Our main result is the following.

Theorem 1.4. *If $x \in (0, \frac{1}{8}]$ and $d = (n-1)p = n^{x+O(\log^{-1} n)}$, then w.h.p.*

$$\beta_{\text{ms}}(\mathcal{G}(n, p)) \leq n^{y_4+O(\log^{-1} n)},$$

where $y_4 = y_4(x) \in (0, 1)$ is the unique solution to $f_x(y) = 4$. On the other hand, if $x \in (0, \frac{1}{2}]$ and $d = n^{x+O(\log^{-1} n)}$, then w.h.p.

$$\beta_{\text{ms}}(\mathcal{G}(n, p)) \geq n^{y_1+O(\log^{-1} n)},$$

where $y_1 = y_1(x)$ is the unique solution to $f_x(y) = 1$. Finally, if $x > \frac{1}{2}$, then w.h.p.

$$\beta_{\text{ms}}(\mathcal{G}(n, p)) = \infty.$$

We remark that Theorem 1.4 is equivalent to saying that if $d = \Theta(n^x)$ for some fixed $x \in (0, 1/8]$, then $\beta_{\text{ms}}(\mathcal{G}(n, p)) = O(n^{y_4(x)})$ w.h.p., and if $d = \Theta(n^x)$ for fixed $x \in (0, 1/2]$, then $\beta_{\text{ms}}(\mathcal{G}(n, p)) = \Omega(n^{y_1(x)})$ w.h.p. (Recall from Lemma 1.3 that the equation $f_x(y) = 4$ has a unique solution if $x \in (0, 1/8]$, and $f_x(y) = 1$ has a unique solution if $x \in (0, 1/2]$.) The gap between the lower and upper bounds is still quite large; indeed, for the range of x for which both bounds apply, the ratio of the upper bound to the lower goes to infinity polynomially in n . It is also still uncertain whether $\beta_{\text{ms}}(\mathcal{G}(n, p))$ is even finite w.h.p. for $x \in (\frac{1}{8}, \frac{1}{2}]$. The result for $x > \frac{1}{2}$ follows from the fact that $\beta_{\text{ms}}(G) = \infty$ for any non-path graph G of diameter at most 2 [19, Theorem 3.1], and the fact that $\mathcal{G}(n, p)$ has diameter 2 w.h.p. in the regime $d = n^x$ for $\frac{1}{2} < x < 1$ [2].

In Figure 1 we plot the functions $y_1(x)$ and $y_4(x)$. Each curve exhibits the type of “zig-zag” behaviour also observed for the metric dimension of $\mathcal{G}(n, p)$ in [2], with jumps occurring at reciprocals of integers $\frac{1}{k}$ for $k \geq 1$. Note that these are precisely the points at which the diameter of $\mathcal{G}(n, p)$ decreases from $k+1$ to k . While obtaining a general formula for $y_1(1/k)$ and $y_4(1/k)$ appears difficult, with an ad hoc approach they are manageable to compute for specific values of k . For instance, we found that $y_1(1/2) = 3/4$, $y_1(1/3) = 2/3$, $y_1(1/4) = 7/12$, and so on.

2 Expansion properties

Here we outline some useful expansion properties of $\mathcal{G}(n, p)$ which hold w.h.p. in the regime we are interested in. We are particularly concerned with the sizes of the shells $S_i(v)$ and $S_i(\{u, v\})$ for vertices u, v , and distances $i \geq 0$. Let $d = (n-1)p$ satisfy $\omega(\log n) = d = o(n)$. Recall that we let $i = i^*(n)$ be the largest integer i such that $d^i = o(n)$ and $c = c(n) = \frac{d^{i^*+1}}{n}$. Define $\gamma = \max \left\{ \sqrt{\frac{\log n}{d}}, \frac{d^{i^*}}{n} \right\} = o(1)$. For $V' \subseteq V$ and $i \in \{0, 1, 2, \dots, i^*\}$, define the events

$$\mathcal{V}_i(V') = \{ |S_i(V')| = (1 + O(\gamma))|V'|d^i \}.$$

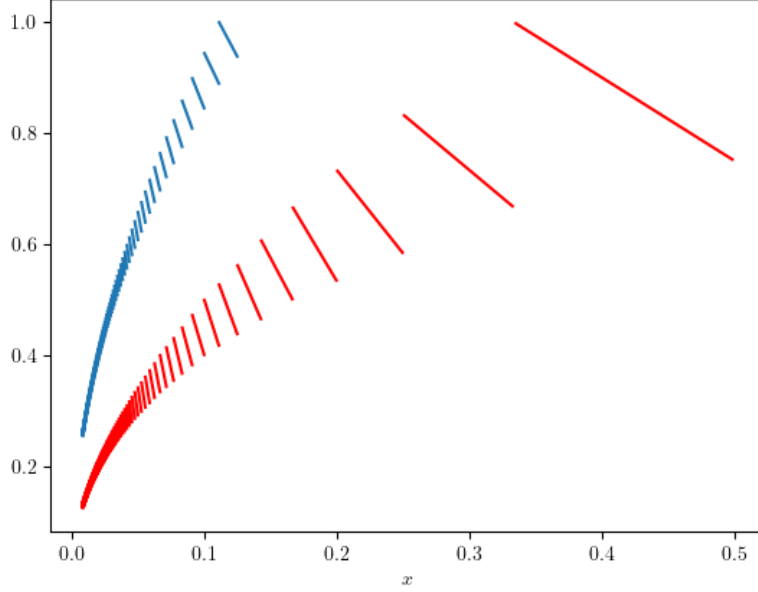


Figure 1: Plots of the points (x, y_4) satisfying $f_x(y_4) = 4$ (blue) for $0 < x \leq \frac{1}{8}$ and (x, y_1) satisfying $f_x(y_1) = 1$ (red) for $0 < x \leq \frac{1}{2}$.

and

$$\mathcal{V}_{i^*+1}(V') = \left\{ |S_i(V')| = \left(1 - e^{-|V'|^c} - \frac{|V'|d^{i^*}}{n} + O\left(\gamma + \frac{\log n}{\sqrt{n}}\right) \right) n \right\}.$$

Let

$$\mathcal{V} := \bigcap_{V': |V'| \in \{1, 2\}} \bigcap_{i=0}^{i^*+1} \mathcal{V}_i(V').$$

All of our results rely on the following lemma. We note that while our main results on the multiset metric dimension only apply to random graphs with average degree on the order of n^x for fixed $x > 0$, the lemma establishes typical expansion properties of random graphs down to degree $d = \omega(\log n)$.

Lemma 2.1. *Suppose $d = (n-1)p$ satisfies $d = \omega(\log n)$ and $d = o(n)$. Then the event $\mathcal{V}$ holds w.h.p. in $\mathcal{G}(n, p)$.*

A version of Lemma 2.1 was originally shown to hold for $(\log n)^5(\log \log n)^2 \leq d = o(n)$ in [3]; it was subsequently shown in [18] to hold with the hypothesis $\omega(\log n) = d = o(n)$. We refer the reader to those sources for the proof. In particular, for levels $i = 0, 1, \dots, i^*$ see [18, Lemma 5.1], and for level $i = i^* + 1$ see [18, Lemma 5.3] (a mild extension is required for the latter case). We summarize some useful consequences of Lemma 2.1 in the following corollary, which is easily deduced from the lemma.

Corollary 2.2. *Let $d = (n-1)p$ satisfy $d = \omega(\log n)$ and $d = o(n)$. Then the following hold w.h.p. for $\mathcal{G}(n, p)$:*

- (i) *For any vertex v and any $i \in \{0, 1, \dots, i^*\}$, $|N_i(v)| = (1 + o(1))d^i$.*
- (ii) *For any pair of vertices v and w and any $i \in \{0, 1, \dots, i^*\}$, $|S_i(w) \setminus N_i(v)| = (1 + o(1))d^i$.*
- (iii) *If $c = \Theta(1)$, for any vertex v , $|N_{i^*+1}(v)| = (1 + o(1))(1 - e^{-c})n$.*
- (iv) *If $c = \Theta(1)$, for any pair of vertices v and w , $|S_{i^*+1}(v) \setminus N_{i^*}(w)| = (1 + o(1))(1 - e^{-c})n$ and $|S_{i^*+1}(w) \setminus N_{i^*+1}(v)| = (1 + o(1))e^{-c}(1 - e^{-c})n$.*

3 Upper bound

Throughout this section, we will condition on $\mathcal{G}(n, p)$ having the typical expansion property $\mathcal{V}$, so that in particular all statements in Corollary 2.2 hold. To prove the upper bound, we use a standard application of the probabilistic method. Our strategy is to show that a random subset of vertices R chosen by including each vertex independently with probability r/n has the following property when $r = r(n)$ is sufficiently large with respect to $d = (n-1)p$: for any pair of vertices v, w ,

$$\Pr(R \text{ fails to distinguish } v \text{ and } w \mid \mathcal{V}) \leq \frac{1}{2n^2}. \quad (4)$$

Then by a union bound, the probability that R fails to distinguish some pair is at most $\binom{n}{2} \cdot \frac{1}{2n^2} \leq \frac{1}{4}$. By Markov's inequality, the probability that R is at least $2\mathbb{E}[|R|] = 2r$ is at most $\frac{1}{2}$. Thus the probability that R is a multiset resolving set and $|R| \leq 2r$ is at least $\frac{1}{4}$, allowing us to conclude that $\mathcal{G}(n, p)$ has a multiset resolving set of size at most $2r$ conditional on $\mathcal{V}$. (In practice, r will be $o(n)$, so the bound is nontrivial.) Since $\mathcal{V}$ holds w.h.p., we get $\beta_{\text{ms}}(\mathcal{G}(n, p)) \leq 2r$ w.h.p.

In the proof, we show that if $d = n^{x+O(\log^{-1}n)}$ for $x \in (0, 1)$, the bound (4) holds when the parameter r is of the order n^y , where y satisfies $f_x(y) \geq 4$. A heuristic explanation is as follows. Recall that i^* is the largest i so that $d^i = o(n)$. Note first that if R fails to distinguish v and w , then $|S_i^R(v) \setminus S_i(w)| = |S_i^R(w) \setminus S_i(v)|$ for all i , since each vertex in $S_i^R(\{v, w\})$ contributes equally to $|S_i^R(v)|$ and $|S_i^R(w)|$. For each $i \in \{0, 1, \dots, i^*\}$, $|S_i^R(v) \setminus S_i(w)|$ and $|S_i^R(w) \setminus S_i(v)|$ are independent binomial random variables with means $(1 + o(1))\mu_i$, where we define $\mu_i = \frac{d^i r}{n} = \Theta(n^{ix+y-1})$ when $r = n^y$. We can then bound the probability that $|S_i^R(v) \setminus S_i(w)| = |S_i^R(w) \setminus S_i(v)|$ by $\min\{1, O(\mu_i^{-1/2})\} = O\left(n^{-\frac{1}{2} \max\{ix+y-1, 0\}}\right)$. (This bound corresponds to the maximum of the probability mass function of the $\text{Bin}(d^i, r/n)$ random variable, see Lemma 3.2.) If all of the variables $|S_i^R(v) \setminus S_i(w)|$ and $|S_i^R(w) \setminus S_i(v)|$ were independent for $i \in \{0, 1, \dots, i^*\}$, we could bound the probability that R fails to distinguish v and w by

$$O\left(\prod_{i=0}^{i^*} n^{-\frac{1}{2} \max\{ix+y-1, 0\}}\right) = n^{-\frac{1}{2} \sum_{i=0}^{i^*} \max\{ix+y-1, 0\} + O(\log^{-1}n)}. \quad (5)$$

In the event, the independence assumption does not hold, since $(S_i(v) \setminus S_i(w)) \cap (S_j(w) \setminus S_j(v))$ can be nonempty for $i \neq j$. Even so, by taking a bit more care in the proof we can show that a bound like (5) still holds. When $\frac{1}{k+1} < x < \frac{1}{k}$ for some positive integer k , we have $i^* = k = \lfloor 1/x \rfloor$, and thus the righthand side of (5) is precisely $n^{-\frac{1}{2}f_x(y) + O(\log^{-1} n)}$, suggesting that y must be large enough so that $f_x(y) \geq 4$ in order for (4) to hold. An extension of this argument shows that the same situation holds when $x = \frac{1}{k}$. Lemma 1.3 gives that we may choose such a y provided that $0 < x \leq 1/8$.

The main result on the upper bound can now be properly stated.

Theorem 3.1. *Let $x \in (0, \frac{1}{8}]$ be fixed and let $d = (n-1)p = n^{x+O(\log^{-1} n)}$. W.h.p., $\mathcal{G}(n, p)$ has a multiset resolving set of size at most $n^{y_4 + O(\log^{-1} n)}$, where $y_4 = y_4(x) \in (0, 1)$ is the unique solution to $f_x(y) = 4$.*

In the proof we also use the following technical result, which bounds the maximum of the probability mass function of a binomial random variable.

Lemma 3.2. *Let Z be a binomial random variable with mean μ . Then*

$$\max_z \Pr(Z = z) \leq \min \left\{ 1, O \left(\frac{1}{\sqrt{\mu}} \right) \right\}.$$

We omit the proof of Lemma 3.2, as it follows without too much difficulty from the observation that the maximum of the probability mass function is attained at a z satisfying $z = \mu + O(1)$ and application of Stirling's formula (similar to the proof of the De Moivre-Laplace Theorem, see e.g., [8, VII.3, Theorem 1]). Now we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. We condition on $\mathcal{G}(n, p)$ satisfying $\mathcal{V}$. Let $\epsilon = \frac{C}{\log n}$ with $C = C(x) > 0$ a large constant to be determined later, let $y = y(n) = y_4 + \epsilon$, and let $r = r(n) = n^y$. Let R be a random subset of V obtained by including each vertex v in R independently with probability r/n . Our goal is to show that for each pair $v, w \in V$,

$$\Pr(R \text{ fails to distinguish } v \text{ and } w \mid \mathcal{V}) = n^{-\frac{1}{2}f_x(y) + O(\log^{-1} n)}. \quad (6)$$

By Lemma 1.2 we have $f_x(y) \geq f_x(y_4) + \epsilon = 4 + \frac{C}{\log n}$. Thus we can choose C sufficiently large so that the right-hand side of (6) is at most $\frac{1}{2n^2}$, and hence by a union bound we have

$$\Pr(R \text{ fails to distinguish some pair of vertices} \mid \mathcal{V}) \leq \binom{n}{2} \cdot \frac{1}{2n^2} \leq \frac{1}{4}.$$

By Markov's inequality the probability that $|R|$ exceeds $2\mathbb{E}[R] = 2r$ is at most $\frac{1}{2}$. Using this and the above, the probability that R distinguishes all vertices and has size at most $2r = n^{y_4 + O(\log^{-1} n)}$ is at least $\frac{1}{4}$. We may then conclude that any graph satisfying $\mathcal{V}$ has a multiset resolving set of size at most $n^{y_4 + O(\log^{-1} n)}$, and thus $\mathcal{G}(n, p)$ has a multiset resolving set of size at most $n^{y_4 + O(\log^{-1} n)}$ w.h.p.

Now we focus on showing (6). Fix v and w . We will reveal the set R in stages with respect to these two vertices. Initially none of R is revealed. In stage 0 we reveal $R \cap N_0(\{v, w\}) = \{v, w\}$. In stage 1 we reveal $R \cap (N_1(\{v, w\}) \setminus N_0(\{v, w\}))$, and in general in stage i we reveal $R \cap (N_i(\{v, w\}) \setminus N_{i-1}(\{v, w\}))$. We reveal R in this way through stage $\lfloor 1/x \rfloor$, though we need to take some care to handle the final stage when $x = \frac{1}{k}$ for some integer $k \geq 8$, as in this case we have $i^* = \lfloor 1/x \rfloor - 1$. We start with the analysis of stages 0 through i^* , which applies to all $x \in (0, 1/8]$.

In stage 0, we simply reveal whether v and w are in R . We condition on neither vertex being in r , as this occurs with probability $1 - O(r/n) = 1 - o(1)$. For $i \geq 1$, condition on the outcomes of stages $0, 1, \dots, i-1$. To process stage i , we first reveal which vertices in $S_i(v) \setminus N_{i-1}(w)$ are in R and also condition on these. At this point we have the following information:

1. $|S_i^R(v)|$ is deterministic (i.e., it is a function of the information revealed so far);
2. $|S_i^R(w)|$ can be expressed as $z_i + Z_i$, where $z_i = |S_i^R(w) \cap N_i(v)|$ is deterministic and $Z_i = |S_i^R(w) \setminus N_i(v)|$ is a random variable;
3. Z_i is a $\text{Bin}(s_i, r/n)$ random variable with $s_i = |S_i(w) \setminus N_i(v)|$.

By the event $\mathcal{V}$, we have $s_i = (1 + o(1))d^i = (1 + o(1))n^{ix + O(\log^{-1} n)} = \Theta(n^{ix})$. Observe that $|S_i^R(v)| = |S_i^R(w)|$ if and only if $Z_i = |S_i^R(v)| - z_i$. Conditioned on the information revealed so far, the probability that $Z_i = |S_i^R(v)| - z_i$ is at most

$$\begin{aligned}
\max_z \Pr(\text{Bin}(s_i, r/n) = z) &\leq \min \left\{ 1, O \left(\sqrt{\frac{n}{rs_i}} \right) \right\} \\
&= O \left(\min \left\{ 1, \sqrt{\frac{n}{rn^{ix}}} \right\} \right) \\
&= O \left(\min \left\{ 1, n^{\frac{1-ix-y}{2}} \right\} \right) \\
&= O \left(n^{-\frac{1}{2} \max\{ix+y-1, 0\}} \right),
\end{aligned}$$

where we use Lemma 3.2 for the inequality in the first line. Suppose $x \in (\frac{1}{k+1}, \frac{1}{k})$ for some integer $k \geq 8$. In this case, we have $i^* = \lfloor 1/x \rfloor = k$, and from the above we inductively have the bound

$$\begin{aligned}
\Pr(\mathbf{m}_R(v) = \mathbf{m}_R(w) \mid \mathcal{V}) &\leq \Pr \left(\bigcap_{i=0}^{i^*} \{ |S_i^R(v)| = |S_i^R(w)| \} \mid \mathcal{V} \right) \\
&\leq \prod_{i=0}^{i^*} O \left(n^{-\frac{1}{2} \max\{ix+y-1, 0\}} \right) \\
&= O \left(n^{-\frac{1}{2} \sum_{i=0}^{\lfloor 1/x \rfloor} \max\{ix+y-1, 0\}} \right) \\
&= n^{-\frac{1}{2} f_x(y) + O(\log^{-1} n)},
\end{aligned}$$

which establishes (6) for this case.

Suppose now that $x = \frac{1}{k}$ for some $k \geq 8$. For this case $i^* = \lfloor 1/x \rfloor - 1 = k - 1$ and $c = \frac{d^{i^*+1}}{n} = n^{O(\log^{-1} n)} = \Theta(1)$. We run stage $i^* + 1$ in the same way as stages 0 through i^* . Let $s_{i^*+1} = |S_{i^*+1}(w) \setminus N_{i^*+1}(v)|$, and observe by Corollary 2.2 that $s_{i^*+1} = (1 + o(1))e^{-c}(1 - e^{-c})n = \Theta(n)$. Using Lemma 3.2, we can then upper bound probability that $|S_{i^*+1}^R(v)| = |S_{i^*+1}^R(w)|$ (conditioned on the history) by

$$\max_z \Pr(\text{Bin}(s_{i^*+1}, r/n) = z) \leq \min \left\{ 1, O \left(\sqrt{\frac{n}{rs_{i^*+1}}} \right) \right\} = O(r^{-1/2}) = n^{-y/2 + O(\log^{-1} n)}.$$

We then have the bound

$$\begin{aligned} \Pr(\mathbf{m}_R(v) = \mathbf{m}_R(w) \mid \mathcal{V}') &\leq n^{-\frac{1}{2} \sum_{i=0}^{i^*} \max\{ix+y-1, 0\} + O(\log^{-1} n)} \cdot n^{-\frac{1}{2}y + O(\log^{-1} n)} \\ &= n^{-\frac{1}{2} \sum_{i=0}^{\lfloor 1/x \rfloor - 1} \max\{ix+y-1, 0\} - \frac{1}{2}y + O(\log^{-1} n)}. \end{aligned}$$

Note that

$$\begin{aligned} \sum_{i=0}^{\lfloor 1/x \rfloor - 1} \max\{ix + y - 1, 0\} &= f_x(y) - \max\{\lfloor 1/x \rfloor x + y - 1, 0\} \\ &= f_x(y) - \max\{k \cdot \frac{1}{k} + y - 1, 0\} \\ &= f_x(y) - y. \end{aligned}$$

From this it follows that

$$\Pr(\mathbf{m}_R(v) = \mathbf{m}_R(w) \mid \mathcal{V}') \leq n^{-\frac{1}{2}f_x(y) + O(\log^{-1} n)},$$

which establishes (6) for this case. The proof of the theorem is finished. $\square$

4 Lower bound

To illustrate the proof of the lower bound, let $x \in (\frac{1}{k+1}, \frac{1}{k}]$ for some $k \geq 2$ and $d = (n-1)p = \Theta(n^x)$. We will make use of the following result on the diameter of $\mathcal{G}(n, p)$ from [2]:

Lemma 4.1 ([2, Corollary 10.12]). *Let $d = (n-1)p = \omega(\log n)$, and suppose*

$$\frac{d^{i-1}}{n} - 2 \log n \rightarrow -\infty \quad \text{and} \quad \frac{d^i}{n} - 2 \log n \rightarrow \infty.$$

Then, the diameter of $\mathcal{G}(n, p)$ equals i w.h.p.

For our choice of x , by Lemma 4.1 the diameter of $\mathcal{G}(n, p)$ is $k+1$ w.h.p. This means that for any set R and any vertex v , the multiset signature $\mathbf{m}_R(v)$ of v with respect to R is w.h.p. fully determined by the first k coordinates, i.e., by $|S_i^R(v)|$ for $i \in \{0, 1, \dots, k\}$. Using a simple counting argument, we will show that w.h.p. in $\mathcal{G}(n, p)$, for any set R ,

at least $(1 + o(1))\frac{n}{2}$ vertices have the property that $|S_i^R(v)| \leq \max\left\{1, O\left(\frac{d^i|R|}{n}\right)\right\}$ for all $i \in \{0, 1, \dots, k\}$. For these *typical* vertices, when $|R| \leq n^y$, the number of possible signatures with respect to R is then crudely at most

$$\prod_{i=0}^k \max\left\{1, O\left(\frac{d^i|R|}{n}\right)\right\} = O\left(\prod_{i=0}^k n^{\max\{ix+y-1, 0\}}\right) = n^{f_x(y) + O(\log^{-1} n)}.$$

If the righthand side above is less than, say, $0.49n$, then we may conclude that there must be some pair of typical vertices v, w with $\mathbf{m}_R(v) = \mathbf{m}_R(w)$. Recall that by Lemma 1.3 there exists a solution $y_1 = y_1(x) \in (0, 1]$ to the equation $f_x(y) = 1$ when $0 < x \leq 1/2$. Let $y = y_1 - \frac{C}{\log n}$ for a constant $C > 0$. By Lemma 1.2 we have $f_x(y) \leq f_x(y_1) - \frac{C}{\log n} = 1 - \frac{C}{\log n}$. Thus if C is chosen sufficiently large, then $n^{f_x(y) + O(\log^{-1} n)}$ can be made less than $0.49n$, which yields the desired result.

Theorem 4.2. *Let $x \in (0, \frac{1}{2}]$ be fixed and let $d = (n-1)p = n^{x+O(\log^{-1} n)}$. W.h.p., $\mathcal{G}(n, p)$ has no multiset resolving set of size less than $n^{y_1+O(\log^{-1} n)}$, where $y_1 = y_1(x) \in (0, 1)$ is the unique solution to $f_x(y) = 1$.*

Note that the diameter of $\mathcal{G}(n, p)$ is 2 for $x > \frac{1}{2}$, in which case there is no multiset resolving set for $\mathcal{G}(n, p)$ by a result from [19].

Proof. Condition on the event $\mathcal{V}$. Suppose that $x \in (\frac{1}{k+1}, \frac{1}{k}]$ for some positive integer $k \geq 2$. Then, by Lemma 4.1, the diameter of $\mathcal{G}(n, p)$ is $k+1$ w.h.p., and we also condition on this. Now, let $R \subseteq V$ have size $|R| = r$. For $i \in \{0, 1, \dots, k\}$, we say that a vertex v is *i-atypical* with respect to R if

$$|N_i^R(v)| \geq \max\left\{2(k+1)\frac{|N_i(v)|r}{n}, 1\right\}$$

and is *i-typical* with respect to R otherwise. If v is *i-typical* for all $i \in \{0, 1, \dots, k\}$ with respect to R , we simply say v is *typical* with respect to R . Our goal is to show that there are at least $(1 + o(1))n/2$ typical vertices.

Let A_i^R be the set of vertices which are *i-atypical* with respect to R , and let

$$P_i^R = \{(v, w) : v \in R, w \in A_i^R, d(v, w) \leq i\}.$$

On one hand, for any $i \in \{0, 1, \dots, k\}$, we have

$$|P_i^R| \geq 2(k+1)\frac{\min_{w \in A_i^R} |N_i(w)|r}{n}|A_i^R|,$$

since there are at least $2(k+1)\frac{\min_{w \in A_i^R} |N_i(w)|r}{n}$ “partners” in R for a fixed $w \in A_i^R$. On the other hand,

$$|P_i^R| \leq \max_{v \in R} |N_i(v)|r,$$

since there are at most $\max_{v \in R} |N_i(v)|$ partners in A_i^R for a fixed $v \in R$. Thus,

$$|A_i^R| \leq \frac{n}{2(k+1)} \cdot \frac{\max_{w \in A_i^R} |N_i(w)|}{\min_{v \in R} |N_i(v)|} = (1 + o(1)) \frac{n}{2(k+1)}$$

where the last equality holds from the fact that $|N_i(w)| = (1 + o(1))|N_i(v)|$ for all pairs v and w and all $i \in \{0, 1, \dots, k\}$ by Corollary 2.2 (i). It follows that the number of vertices which are i -atypical for some $i \in \{0, 1, \dots, k\}$ is at most $(1 + o(1))\frac{n}{2}$, i.e., there are at least $(1 + o(1))\frac{n}{2}$ typical vertices. Note that this holds for *any* subset R .

Now, since the diameter is $k+1$, for any $R \subseteq V$ and $v \in V$, the signature $\mathbf{m}_R(v)$ is fully determined by the coordinates $0, 1, \dots, k$. First suppose that $\frac{1}{k+1} < x < \frac{1}{k}$. Note that in this case we have $i^* = k$. Let $\epsilon = \frac{C}{\log n}$ with $C > 0$ a constant to be determined, and for a contradiction suppose that there exists a multiset resolving set R of size n^y with $y \leq y_1 - \epsilon$. By Corollary 2.2 (i), for any $v \in V$ and any $i \in \{0, 1, \dots, k\}$, we have $|N_i(v)| = (1 + o(1))d^i = \Theta(n^{ix})$. Thus, if v is typical with respect to R , then for any $i \in \{0, 1, \dots, k\}$, there are at most

$$\max \left\{ (1 + o(1))2(k+1)\frac{d^i r}{n}, 1 \right\} = O(\max\{n^{ix+y-1}, 1\}) = O(n^{\max\{ix+y-1, 0\}}) \quad (7)$$

possibilities for the i th coordinate of $\mathbf{m}_R(v)$. Therefore, the number of signatures which are allowed for typical vertices is at most

$$O\left(\prod_{i=0}^k n^{\max\{ix+y-1, 0\}}\right) = n^{f_x(y) + O(\log^{-1} n)} \leq n^{f_x(y_1) - \epsilon + O(\log^{-1} n)},$$

where in the final inequality we use that $f_x(y) \leq f_x(y_1 - \epsilon) \leq f_x(y_1) - \epsilon$ by Lemmas 1.1 and 1.2. (Note that $f_x(y_1 - \epsilon) > 0$ holds for n sufficiently large since f_x is continuous by Lemma 1.1 and $\epsilon = o(1)$, so we can indeed apply Lemma 1.2.) If C is chosen large enough, then the right-hand-side above can be made smaller than, say, $0.49n$. By our previous arguments, there are at least $(1 + o(1))\frac{n}{2}$ typical vertices with respect to R . But since there are only $0.49n$ possible signatures for these vertices, there must be a pair v, w of typical vertices with $\mathbf{m}_R(v) = \mathbf{m}_R(w)$, which gives us the desired contradiction.

The argument above extends easily to the case that $x = \frac{1}{k}$ for some positive integer $k \geq 2$. In this case we have $i^* = k-1$, so the bound (7) holds for any $i \in \{0, 1, \dots, k-1\}$ in this case. Trivially, we have that $|N_k(v)| = O(n)$ for any v , which means there are at most $O(n \cdot \frac{r}{n}) = O(r) = O(n^y)$ possibilities for the k th coordinate of any typical vertex's signature with respect to R . There are then at most

$$O\left(n^y \cdot \prod_{i=0}^{k-1} n^{\max\{ix+y-1, 0\}}\right) = n^{f_x(y) + O(\log^{-1} n)}$$

possible signatures for typical vertices, and the rest of the argument follows as before. $\square$

5 Conclusion

We have shown that for the regime $d = (n - 1)p = n^{x+O(\log^{-1} n)}$ for fixed $0 < x \leq 1/8$, the multiset metric dimension $\beta_{\text{ms}}(\mathcal{G}(n, p))$ of $\mathcal{G}(n, p)$ satisfies $\beta_{\text{ms}}(\mathcal{G}(n, p)) \leq n^{y_4+O(\log^{-1} n)}$ w.h.p., and for fixed $0 < x \leq \frac{1}{2}$, $\beta_{\text{ms}}(\mathcal{G}(n, p)) \geq n^{y_1+O(\log^{-1} n)}$ w.h.p., where y_4 and y_1 are constants in $(0, 1)$ which are explicitly computable from x . We chose the scaling $d = n^{x+O(\log^{-1} n)}$ primarily to make our results cleaner to state. However, the upper bound can be extended to the more general setting $d = \omega(\log n)$; indeed our argument depends only on $\mathcal{G}(n, p)$ satisfying the typical expansion profile guaranteed by Lemma 2.1, which holds w.h.p. for $d = \omega(\log n)$. In principle, the lower bound can also be extended to this regime, though the arguments there also depend on knowing the diameter of $\mathcal{G}(n, p)$ w.h.p.

We propose two directions for future research. The first is to determine whether $\beta_{\text{ms}}(\mathcal{G}(n, p))$ is finite w.h.p. for $d = n^x$, $\frac{1}{8} < x \leq \frac{1}{2}$. The second is to tighten the gap between the lower and upper bounds for $\beta_{\text{ms}}(\mathcal{G}(n, p))$ for $x \in (0, \frac{1}{2})$. New proof strategies are likely needed for both problems.

References

- [1] Ridho Alfarisi, Yuqing Lin, Joe Ryan, Dafik Dafik, and Ika Hesti Agustin. A note on multiset dimension and local multiset dimension of graphs. *Statistics, Optimization & Information Computing*, 8(4):890–901, 2020.
- [2] Béla Bollobás. *Random graphs*. Springer, 1998.
- [3] Béla Bollobás, Dieter Mitsche, and Paweł Prałat. Metric dimension for random graphs. *The Electronic Journal of Combinatorics*, 20(4):P52, 2013.
- [4] Novi H Bong and Yuqing Lin. Some properties of the multiset dimension of graphs. *Electron. J. Graph Theory Appl.*, 9(1):215–221, 2021.
- [5] Josep Díaz, Harrison Hartle, and Cristopher Moore. The metric dimension of sparse random graphs. *arXiv preprint arXiv:2504.21244*, 2025.
- [6] Andrzej Dudek, Sean English, Alan Frieze, Calum MacRury, and Paweł Prałat. Localization game for random graphs. *Discrete Applied Mathematics*, 309:202–214, 2022.
- [7] Andrzej Dudek, Alan Frieze, and Wesley Pegden. A note on the localization number of random graphs: diameter two case. *Discrete Applied Mathematics*, 254:107–112, 2019.
- [8] William Feller. *An introduction to probability theory and its applications*, volume 1. Wiley Series in Probability and Mathematical Statistics, 1957.
- [9] Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.

- [10] Reynaldo Gil-Pons, Yuniór Ramírez-Cruz, Rolando Trujillo-Rasua, and Ismael G Yero. Distance-based vertex identification in graphs: The outer multiset dimension. *Applied Mathematics and Computation*, 363:124612, 2019.
- [11] Anni Hakanen and Ismael G Yero. Complexity and equivalency of multiset dimension and id-colorings. *Fundamenta Informaticae*, 191(3-4):315–330, 2024.
- [12] F. Harary and R.A. Melter. The metric dimension of a graph. *Ars Combinatoria*, 2:191–195, 1976.
- [13] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs*. John Wiley & Sons, 2011.
- [14] Sandi Klavžar, Dorota Kuziak, and Ismael G Yero. Further contributions on the outer multiset dimension of graphs. *Results in Mathematics*, 78(2):50, 2023.
- [15] Lyuben Lichev, Dieter Mitsche, and Paweł Prałat. Localization game for random geometric graphs. *European Journal of Combinatorics*, 108:103616, 2023.
- [16] Dieter Mitsche and Juanjo Rué. On the limiting distribution of the metric dimension for random forests. *European Journal of Combinatorics*, 49:68–89, 2015.
- [17] Gergely Odor. The role of adaptivity in source identification with time queries. Technical report, EPFL, 2022.
- [18] Gergely Odor and Patrick Thiran. Sequential metric dimension for random graphs. *Journal of Applied Probability*, 58(4):909–951, 2021.
- [19] Rinovia Simanjuntak, Presli Siagian, and Tomas Vetrik. The multiset dimension of graphs. *arXiv preprint arXiv:1711.00225*, 2017.
- [20] P. Slater. Leaves of trees. *Congressus Numerantium*, 14:549–559, 1975.
- [21] Richard C Tillquist, Rafael M Frongillo, and Manuel E Lladser. Getting the lay of the land in discrete space: A survey of metric dimension and its applications. *SIAM Review*, 65(4):919–962, 2023.
- [22] Richard C Tillquist and Manuel E Lladser. Low-dimensional representation of genomic sequences. *Journal of mathematical biology*, 79(1):1–29, 2019.
- [23] Richard D Tillquist and Manuel E Lladser. Multilateration of random networks with community structure. *arXiv preprint arXiv:1911.01521*, 2019.